



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA MOLECULAR

**DESENVOLVIMENTO DE MARCADORES MICROSSATÉLITES E
ANÁLISE DE EXPRESSÃO DE GENES ENVOLVIDOS EM RESPOSTA
AO ESTRESSE BIÓTICO EM GENÓTIPOS DE *MUSA ACUMINATA***

FLÁVIA LEONEL EMEDIATO

Brasília – DF

2014

FLÁVIA LEONEL EMEDIATO

**DESENVOLVIMENTO DE MARCADORES MICROSSATÉLITES E
ANÁLISE DE EXPRESSÃO DE GENES ENVOLVIDOS EM RESPOSTA
AO ESTRESSE BIÓTICO EM GENÓTIPOS DE *MUSA ACUMINATA***

Trabalho apresentado à Universidade de
Brasília como requisito para defesa da tese de
doutorado em Biologia Molecular pelo
Programa de Pós Graduação em Biologia
Molecular

Orientador

Prof. Dr. Robert Neil Gerard Miller

**BRASÍLIA
DISTRITO FEDERAL - BRASIL
2014**

Dedico este trabalho ao meu filho Gabriel, por ser a grande fonte de inspiração e o motivo pelo qual me fez buscar e persistir neste sonho.

AGRADECIMENTOS

À Universidade de Brasília (UnB) pela oportunidade de realização do curso de Doutorado pelo Programa de Pós-Graduação em Biologia Molecular.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos.

Ao Professor Dr. Robert Miller, pela orientação, disponibilidade e atenção em todos os momentos que precisei, pela grande contribuição em minha formação profissional, pela amizade e confiança em mim depositada.

Aos pesquisadores Vânia Rennó, Ana Brasileiro, Orzenil Silva e Roberto Togawa pelos ensinamentos e auxílio nas análises.

Aos meus pais Sérgio e Regina e irmãos, Carol, Serginho e Paula, por serem meu suporte, pelo apoio incondicional, conselhos, torcida e sobretudo o amor que nos torna verdadeiramente uma família.

Ao meu marido Eduardo, pelo amor, incentivo, paciência e pela compreensão em tantos momentos de ausência os quais eu espero recuperar.

Ao meu amigo Marco Aurélio que me acompanhou nessa caminhada com tanta sabedoria, pelo convívio nos momentos de dificuldades e incertezas que enfrentamos juntos.

À todos os colegas do laboratório de Interação Planta-Praga pela amizade e troca de conhecimentos.

Finalmente, a DEUS, por me proporcionar mais este desafio, dando-me coragem e força constantes.

SUMÁRIO

RESUMO	9
ABSTRACT	11
REVISÃO DA LITERATURA	12
1. ASPECTOS BOTÂNICOS E CLASSIFICAÇÃO TAXONÔMICA DO GÊNERO <i>MUSA</i>	12
2. IMPORTÂNCIA SÓCIO-ECONÔMICA	13
3. ESTRESSES BIÓTICOS	14
3.1. MYCOSPHAERELLA FIJIENSIS, MYCOSPHAERELLA MUSICOLA E AS DOENÇAS DE SIGATOKA	16
4. RECURSOS GENÉTICOS E MELHORAMENTO EM <i>MUSA</i>	20
5. IMUNIDADE EM PLANTAS	24
5.1. OS RECEPTORES PRRS E A IMUNIDADE DISPARADA POR PAMPs – PTI	27
5.2. AS PROTEÍNAS R E A IMUNIDADE DISPARADA POR EFETORES – ETI	29
JUSTIFICATIVA	35
OBJETIVOS	36
1. OBJETIVOS GERAIS	36
2. OBJETIVOS ESPECÍFICOS	36
<u>CAPÍTULO 1: ANÁLISE DE DADOS DO TRANSCRITOMA DA INTERAÇÃO ENTRE <i>MUSA ACUMINATA</i> E <i>MYCOSPHAERELLA MUSICOLA</i> PARA GENES ENVOLVIDOS NO PROCESSO DE IMUNIDADE</u>	<u>37</u>
1. INTRODUÇÃO	37
2. MATERIAL E MÉTODOS	39
2.1. MATERIAL BIOLÓGICO	39
2.2. PREPARAÇÃO DO INÓCULO	39
2.3. MICROSCOPIA ELETRÔNICA DE VARREDURA	40
2.4. EXTRAÇÃO DE RNA	42
2.5. SEQUENCIAMENTO MASSAL E PROCESSAMENTO DE SEQUÊNCIAS	42
2.6. MINERAÇÃO DE DADOS 454 PARA GENES ENVOLVIDOS EM DEFESA E RESISTÊNCIA	42
3. RESULTADOS E DISCUSSÃO	44
3.1. MICROSCOPIA ELETRÔNICA DE VARREDURA	44
3.2. QUALIDADE DO MATERIAL VEGETAL	45
3.3. SEQUENCIAMENTO MASSAL E PROCESSAMENTO DE SEQUÊNCIAS	46
3.4. MINERAÇÃO DE DADOS DE PIROSEQUENCIAMENTO 454 PARA TRANSCRITOS DE GENES POTENCIALMENTE ENVOLVIDOS EM DEFESA E RESISTÊNCIA EM <i>M. ACUMINATA</i>	46
<u>CAPÍTULO 2: ANÁLISE DE EXPRESSÃO DE GENES POTENCIALMENTE ENVOLVIDOS EM ESTRESSES BIÓTICOS NA INTERAÇÃO <i>M. ACUMINATA</i> E <i>MYCOSPHAERELLA MUSICOLA</i></u>	<u>54</u>
1. INTRODUÇÃO	54
2. MATERIAL E MÉTODOS	55
2.1. MATERIAL BIOLÓGICO	55

2.2. DESENHO DE PRIMERS EM REGIÕES UTR DE GENES DE DEFESA DERIVADOS DE SEQUÊNCIAS DE TRANSCRITOMA (SEQUENCIAMENTO 454)	55
2.3. DESENHO DE PRIMERS ESPECÍFICOS DERIVADOS DE CONTIGS NBS-LRR	56
2.4. DESENHO DE PRIMERS EM REGIÕES UTR DE GENES DE RESISTÊNCIA	57
2.5. DESENHO DE PRIMERS PARA GENES REFERÊNCIA	60
2.6. CONFIRMAÇÃO DE AMPLIFICAÇÃO IN SÍLICO	60
2.7. PCR EM TEMPO REAL	60
3. RESULTADOS E DISCUSSÃO	62
3.1. QUANTIFICAÇÃO DE RNA TOTAL DE <i>M. ACUMINATA</i> CALCUTTA 4 E <i>M. ACUMINATA</i> CAVENDISH GRANDE NAINÉ INOCULADOS E NÃO INOCULADOS COM <i>M. MUSICOLA</i> E SÍNTESE DE CDNA	62
3.2. DESENVOLVIMENTO DE PRIMERS ESPECÍFICOS PARA ANÁLISE DE EXPRESSÃO DE GENES CANDIDATOS POR RT-qPCR	63
3.3. ANÁLISES DE ESPECIFICIDADE	67
3.4. ANÁLISES DE EXPRESSÃO	69
<u>CAPÍTULO 3: DESENVOLVIMENTO E CARACTERIZAÇÃO DE MARCADORES SSR PARA GENES ENVOLVIDOS EM RESPOSTA A ESTRESSES BIÓTICOS EM GENÓTIPOS DE <i>M. ACUMINATA</i></u>	
1. INTRODUÇÃO	77
2. MATERIAL E MÉTODOS	78
2.1. MATERIAL BIOLÓGICO	78
2.2. CARACTERIZAÇÃO E DESENHO DE PRIMERS PARA SSRs LOCALIZADOS EM GENES ANÁLOGOS DE RESISTÊNCIA (RGAs)	79
2.3. CARACTERIZAÇÃO E DESENHO DE PRIMERS PARA SSRs LOCALIZADOS EM GENES DE DEFESA	82
2.4. AMPLIFICAÇÃO DE MICROSSATÉLITES VIA PCR	82
2.5. ELETROFORESE E VISUALIZAÇÃO DOS LOCOS	83
3. RESULTADOS E DISCUSSÃO	84
3.1. DESENVOLVIMENTO DE MARCADORES SSR PARA GENES POTENCIALMENTE ENVOLVIDOS EM RESISTÊNCIA (RGAs)	85
3.3. CARACTERIZAÇÃO DE MARCADORES MICROSSATÉLITES PARA LOCOS POTENCIALMENTE ENVOLVIDOS EM RESPOSTA AOS ESTRESSES BIÓTICOS E ANÁLISE DE POLIMORFISMOS EM GENÓTIPOS DE <i>M. ACUMINATA</i>	99
<u>CONCLUSÃO</u>	107
<u>REFERÊNCIAS BIBLIOGRÁFICAS</u>	109
<u>ANEXOS</u>	129

LISTA DE TABELAS

TABELA 1: PRINCIPAIS PRODUTORES MUNDIAIS DE BANANA NO ANO DE 2011. FONTE: ADAPTADO DE FAOSTAT © FAO STATISTICS DIVISION 2013 01 OCTOBER 2013 _____	13
TABELA 2: RESUMO DOS DIFERENTES ESTÁGIOS DE LESÃO ASSOCIADOS ÀS DOENÇAS SIGATOKAS NEGRA E AMARELA _____	17
TABELA 3: REPRESENTAÇÃO DOS DIAS DE COLETA DE PLANTAS DE <i>M. ACUMINATA</i> CALCUTTA 4 E CAVENDISH GRANDE NAINÉ INOCULADAS E NÃO INOCULADAS (CONTROLE) COM O PATÓGENO <i>M. MUSICOLA</i> . _____	41
TABELA 4: ANÁLISE BLASTX E <i>CONTIGS</i> NBS-LRR DERIVADOS DO TRANSCRITOMA FOLIAR DA INTERAÇÃO <i>M. MUSICOLA</i> E <i>M. ACUMINATA</i> CALCUTTA 4 _____	48
TABELA 5: ANÁLISE BLASTX DE <i>CONTIGS</i> NBS-LRR DERIVADOS DO TRANSCRITOMA FOLIAR DA INTERAÇÃO <i>M. MUSICOLA</i> E <i>M. ACUMINATA</i> CAVENDISH GRANDE NAINÉ _____	48
TABELA 6: DADOS DE MAPEAMENTO PARA <i>CONTIGS UNIGENES</i> DE <i>M. ACUMINATA</i> CALCUTTA 4 ALINHADOS COM TODOS OS MODELOS DE GENES NB-ARC DE <i>M. ACUMINATA</i> SSP. <i>MALACCENSIS</i> VAR. PAHANG (DH PAHANG). _____	49
TABELA 7: DADOS DE MAPEAMENTO PARA <i>CONTIGS UNIGENES</i> DE <i>M. ACUMINATA</i> CAVENDISH GRANDE NAINÉ ALINHADOS COM TODOS OS MODELOS DE GENES NB-ARC DE <i>M. ACUMINATA</i> SSP. <i>MALACCENSIS</i> VAR. PAHANG (DH PAHANG). _____	51
TABELA 8: COMPARAÇÃO DE VALORES QUANTITATIVOS DE CONCENTRAÇÃO DE RNA TOTAL DE FOLHAS DE <i>M. ACUMINATA</i> DOS QUATRO <i>POOLS</i> BASEADOS EM ANÁLISES CONDUZIDAS COM ESPECTROFOTOMETRIA NANO DROP ND-1000 E QUBIT QUANT-IT™ RNA ASSAY. _____	62
TABELA 9: SEQUÊNCIAS DE <i>PRIMERS</i> PARA ANÁLISES DE EXPRESSÃO POR MEIO DE PCR EM TEMPO REAL DE GENES REFERÊNCIA E GENES POTENCIALMENTE ENVOLVIDOS EM PROCESSOS DE RESISTÊNCIA E DEFESA DE <i>M. ACUMINATA</i> . _____	64
TABELA 10: RESULTADOS DE EXPRESSÃO RELATIVA, EFICIÊNCIA DA REAÇÃO DE PCR EM TEMPO REAL, NÍVEL DE SIGNIFICÂNCIA E TIPO DE REGULAÇÃO DA EXPRESSÃO DOS GENES POTENCIALMENTE ENVOLVIDOS EM RESPOSTAS DE DEFESA E RESISTÊNCIA (RGAs). _____	71
TABELA 11: ACESSOS DE DIPLÓIDES (AA) DE <i>M. ACUMINATA</i> CONTRASTANTES EM RESISTÊNCIA À DOENÇAS CAUSADAS POR <i>MYCOSPHAERELLA</i> SELECIONADOS PARA USO EM VALIDAÇÃO DE MARCADORES MOLECULARES SSR. FONTE: ADAPTADO DE PASSOS <i>ET AL.</i> , 2012. _____	78
TABELA 12: AGRUPAMENTO DOS GENÓTIPOS EM <i>BULKS</i> CONTRASTANTES EM RESISTÊNCIA ÀS SIGATOKAS NEGRA E AMARELA. _____	79
TABELA 13: SUBGRUPO SELECIONADO DE 14 MARCADORES SSR, DERIVADOS DE GENES NB-ARC PRESENTES NAS BIBLIOTECAS BAC DE <i>M. ACUMINATA</i> CALCUTTA 4 E CAVENDISH GRANDE NAINÉ, VALIDADOS PARA AMPLIFICAÇÃO POR PCR E POLIMORFISMO NOS 22 GENÓTIPOS DIPLÓIDES DE <i>M. ACUMINATA</i> . _____	86
TABELA 14: INFORMAÇÃO DAS SEQUÊNCIAS DE <i>PRIMERS</i> SSR DERIVADOS DO GENOMA DE <i>M. ACUMINATA</i> DH PAHANG. DADOS REPRESENTAM TODOS OS LOCOS IDENTIFICADOS DISTANDO EM ATÉ 10.000PB DE GENES NB-ARC, ACOMPANHADOS DA INFORMAÇÃO DOS MOTIVOS DE REPETIÇÃO SSR E TEMPERATURA DE ANELAMENTO. _____	87
TABELA 15: INFORMAÇÕES DOS <i>PRIMERS</i> SSR DESENVOLVIDOS A PARTIR DE <i>UNIGENES</i> DERIVADOS DO TRANSCRITOMA DE <i>M. ACUMINATA</i> CALCUTTA 4 EM INTERAÇÃO COM <i>M. MUSICOLA</i> MAPEADOS NO GENOMA DE <i>M. ACUMINATA</i> DH PAHANG. _____	90
TABELA 16: INFORMAÇÕES DOS LOCOS SSR DERIVADOS DE TRANSCRITOS (SEQUENCIAMENTO 454) DE <i>M. ACUMINATA</i> CALCUTTA 4 PERTENCENTES À DIFERENTES FAMÍLIAS DE GENES R. _____	93
TABELA 17: INFORMAÇÕES DE <i>PRIMERS</i> SSR DERIVADOS DE <i>UNIGENES</i> DE TRANSCRITOS 454 DE <i>M. ACUMINATA</i> CALCUTTA 4 POTENCIALMENTE ENVOLVIDOS EM RESPOSTAS DE DEFESA. _____	96
TABELA 18: MARCADORES MICROSATÉLITES DERIVADOS DE <i>UNIGENES</i> POTENCIALMENTE ENVOLVIDOS EM RESPOSTAS A ESTRESSES BIÓTICOS. HE- HETEROZIGOSIDADE ESPERADA. HO - HETEROZIGOSIDADE OBSERVADA. PIC- CONTEÚDO DE INFORMAÇÃO POLIMÓRFICA _____	102

LISTA DE FIGURAS

FIGURA 1: O DESEMPENHO DA CULTURA DE BANANA NO BRASIL 1970 - 2010.	14
FIGURA 2: REPRESENTAÇÃO DOS CINCO ESTÁGIOS DA SIGATOKA AMARELA.	18
FIGURA 3: CICLO DE VIDA DE <i>M. MUSICOLA</i> , O PATÓGENO CAUSADOR DA SIGATOKA AMARELA.,	19
FIGURA 4: FORÇAS OPOSTAS DE SELEÇÃO SOBRE MOLÉCULAS GUARDA DE PROTEÍNAS R.	25
FIGURA 5: MODELO ZIG-ZAG DO SISTEMA IMUNE EM PLANTAS, PROPOSTO POR JONES E DANGL (2006).	26
FIGURA 6: PRINCIPAIS PRRs IDENTIFICADOS EM PLANTAS.	28
FIGURA 7: ESQUEMA DA LOCALIZAÇÃO DOS GENES R DE DIFERENTES CLASSES QUE CONFEREM RESISTÊNCIA RAÇA ESPECÍFICA E RAÇA NÃO ESPECÍFICA.	34
FIGURA 8: OBSERVAÇÃO DA INTERAÇÃO DE <i>M. MUSICOLA</i> COM A SUPERFÍCIE INTERIOR DAS FOLHAS DOS GENÓTIPOS DE <i>M. ACUMINATA</i> POR MEIO DE MICROSCOPIA ELETRÔNICA DE VARREDURA (MEV).	45
FIGURA 9: MODELO DE DESENHO DE <i>PRIMER</i> PARA O GENE ENVOLVIDO EM RESPOSTA DE DEFESA 'CALOSE SINTASE'.	56
FIGURA 10: ARQUIVO DE SAÍDA DO PROGRAMA FGGENESH MOSTRANDO A ESTRUTURA DE UM DETERMINADO GENE DE RESISTÊNCIA ENCONTRADO EM <i>M. ACUMINATA</i> DH PAHANG.	57
FIGURA 11: ARQUIVO DE SAÍDA DO PROGRAMA GENEWISE MOSTRANDO O ALINHAMENTO ENTRE PROTEÍNA (FGGENESH) E DNA (EMBOSS_001).	58
FIGURA 12: INTERFACE DO GNPANNOT MOSTRANDO SEQUÊNCIAS DERIVADAS DO TRANSCRITOMA (SEQUENCIAMENTO 454), <i>ASSEMBLIES</i> (1.1, 2.1, 3.1) MAPEADOS EM UM GENE POTENCIALMENTE ENVOLVIDO EM RESISTÊNCIA, NA SEQUÊNCIA DO BAC MAC054B03 DE <i>M. ACUMINATA</i> CAVENDISH GRANDE NAINÉ.	59
FIGURA 13: ANÁLISE ELETROFORÉTICA EM GEL DE AGAROSE 1% MOSTRANDO A AMPLIFICAÇÃO DOS QUATRO <i>POOLS</i> DE CDNA UTILIZANDO <i>PRIMERS</i> INTRÔNICO PARA DETECÇÃO DE CONTAMINAÇÃO POR DNA.	62
FIGURA 14. FLUXOGRAMA ILUSTRANDO O DESENVOLVIMENTO DOS PARES DE <i>PRIMERS</i> PARA GENES ENVOLVIDOS EM RESPOSTAS AOS ESTRESSES BIÓTICOS PARA ANÁLISES DE EXPRESSÃO VIA qPCR.	63
FIGURA 15: CURVAS APRESENTADAS NA CORRIDA DE PCR EM TEMPO REAL DO GENE REFERÊNCIA <i>ELONGATION FACTOR</i> .	67
FIGURA 16: CURVAS DE DISSOCIAÇÃO (A) E DE AMPLIFICAÇÃO (B) DO RGA 37 EM TRIPPLICATAS DE CAVENDISH GRANDE NAINÉ INOCULADO (VERMELHO) E CONTROLE NÃO INOCULADO (AMARELO).	68
FIGURA 17: NÍVEIS DE EXPRESSÃO RELATIVA DOS GENES POTENCIALMENTE ENVOLVIDOS EM RESISTÊNCIA (RGAs) E RESPOSTAS DE DEFESA DE <i>M. ACUMINATA</i> CALCUTTA 4 E CAVENDISH GRANDE NAINÉ SUBMETIDOS AO INÓCULO COM <i>M. MUSICOLA</i> , COMPARADOS AO CONTROLE NÃO INOCULADO.	70
FIGURA 18: ARQUIVO DE SAÍDA DO PROGRAMA WEBSAT MOSTRANDO DADOS DOS <i>PRIMERS FORWARD</i> E <i>REVERSE</i> DESENHADOS PARA O MICROSSATÉLITE (AGA) ₁₀ CONTIDO NA SEQUÊNCIA BAC MAC139M12, DENTRO DE UM GENE CODIFICADOR DA PROTEÍNA DA FAMÍLIA NB-ARC.	80
FIGURA 19: FLUXOGRAMA DEMONSTRANDO O DESENVOLVIMENTO DOS <i>PRIMERS</i> PARA SSRs EM LOCOS POTENCIALMENTE ENVOLVIDOS EM RESPOSTA AOS ESTRESSES BIÓTICOS.	84
FIGURA 20: DISTRIBUIÇÃO DE SSRs LOCALIZADOS PRÓXIMO OU DENTRO DE RGAs (FAMÍLIA NB-ARC) DE SEQUÊNCIAS DE CLONES BAC DE <i>M. ACUMINATA</i> CALCUTTA 4 E CAVENDISH GRANDE NAINÉ.	85
FIGURA 21: DISTRIBUIÇÃO DE LOCOS SSR DERIVADOS DE TRANSCRITOS NBS-LRR DE <i>M. ACUMINATA</i> CALCUTTA 4 MAPEADOS AO LONGO DO GENOMA DE <i>M. ACUMINATA</i> DH PAHANG.	89
FIGURA 22: REPRESENTAÇÃO GENÔMICA DE <i>M. ACUMINATA</i> DH PAHANG COM DESTAQUE PARA SSRs E GENES NB-ARC.	92
FIGURA 23: DISTRIBUIÇÃO DE SSRs DERIVADOS DE DIFERENTES FAMÍLIAS DE GENES R, IDENTIFICADOS EM TRANSCRITOS (SEQUENCIAMENTO 454) EM <i>M. ACUMINATA</i> CALCUTTA 4	92
FIGURA 24: ANÁLISE ELETROFORÉTICA EM GEL DE POLIACRILAMIDA 4% DE AMPLICONS DO LOCO POLIMÓRFICO 454_DEF_LOCUS120	99
FIGURA 25: ANÁLISE ELETROFORÉTICA EM GEL DE POLIACRILAMIDA 4% DE AMPLICONS POLIMÓRFICOS EM PARENTAIS DE <i>M. ACUMINATA</i> , LOCO 454_DEF_LOCUS120.	100

RESUMO

EMEDIATO, Flávia Leonel. DESENVOLVIMENTO DE MARCADORES MICROSSATÉLITES E ANÁLISE DE EXPRESSÃO DE GENES ENVOLVIDOS EM RESPOSTA AO ESTRESSE BIÓTICO EM GENÓTIPOS DE *MUSA ACUMINATA* 2014. 129 Páginas. Tese de Doutorado em Biologia Molecular – Universidade de Brasília, Brasília, DF.

Cultivares comerciais de bananeira (*Musa* spp) são cultivados em muitos países de ambientes tropicais e subtropicais. A reprodução assexuada da espécie resultou em uma base genética restrita, sem resistência a pragas e doenças, tornando as estratégias de melhoramento convencional muito limitadas devido a esterilidade de muitos cultivares comerciais. As estratégias de melhoramento não-convencionais, como a transformação e seleção assistida por marcadores, oferecem uma abordagem alternativa para a introgressão de genes de resistência em cultivares comerciais. Genes de resistência (R) em plantas codificam receptores de proteínas que reconhecem assinaturas moleculares de efetores de patógenos de plantas, desencadeando mecanismos de defesa como a imunidade disparada por efetores (ETI). Essa resistência pode ser perdida durante a co-evolução do hospedeiro e patógeno, favorecendo a evolução das novas raças patogênicas. Neste contexto, o desenvolvimento contínuo de cultivares de plantas resistentes é necessária, a fim de acompanhar a evolução de patógenos. O objetivo geral deste estudo foi identificar genes potencialmente envolvidos em respostas a estresses bióticos, que servirão como fonte de recursos para análises de expressão diferencial durante a interação de *Musa acuminata* com o patógeno *Mycosphaerella musicola* e o desenvolvimento de marcadores SSR para busca de polimorfismos em genótipos de *M. acuminata*. Dessa forma, foram identificados 14 *unigenes* expressando proteínas NBS-LRR em tecidos foliares de Calcutta 4 e 25 em Cavendish Grande Naine, ambos infectados e não-infectados. O mapeamento desses *unigenes* com modelos que contêm o domínio NB-ARC no genoma de referência de *M. acuminata* ssp. malaccensis var. Pahang (DH Pahang) identificou 38 *contigs unigenes* mapeando a 40 modelos de genes completos em Calcutta 4 e 43 *contigs unigenes* mapeando em 40 modelos de genes em Cavendish Grande Naine. Foram desenvolvidos 156 marcadores microsatélites para os genes alvo derivados de diferentes fontes: sequências de transcrito de folhas da interação *M. acuminata* e *M. musicola*, mapeadas em sequências de clones de BAC de *M. acuminata* Calcutta 4 contendo análogos a genes de resistência NB-ARC (RGAs); Sequências do genoma completo de *M. acuminata* DH Pahang contendo domínios conservados NB-ARC e outros genes R; e transcritos da interação entre *M. acuminata* e *M. musicola* minerados para genes de defesa. Destes locos SSR identificados, 21 (13,46%) foram polimórficos, com alelos por loco variando de 2 a 4, e o conteúdo de informação de polimorfismo variando de 0,31 a 0,67. As análises de expressão de 51 genes alvo demonstraram que 19 obtiveram especificidade na reação de PCR em tempo real e que a maioria dos genes envolvidos em processos de defesa não foram modulados, com exceção do gene da corismato sintase e NAC. O gene NPR1 foi expresso apenas no cultivar suscetível e os RGAs 37, 43, M09A31, M09A21, B03A51 e B03A11 tiveram um aumento de expressão quando inoculados com o patógeno. Os RGAs 43 e RGA 37 mostraram indução significativa em *M. acuminata* Calcutta 4 quando inoculada com o patógeno e os RGAs 9330, 7100, 6160, M09A31 e M09A21 apresentaram regulação negativa em Calcutta 4. Esta análise foi consistente com a tendência aparente de caracterizações anteriores de genes R individuais já que alguns genes R são induzidos por uma variedade de estímulos, alguns exibem respostas

muito específicas e outros parecem não ser responsivos ao patógeno. A caracterização contínua de genes envolvidos em respostas ao estresse biótico em *M. acuminata* durante a interação com *M. musicola*, bem como o desenvolvimento de marcadores moleculares, irá contribuir para o desenvolvimento de uma gestão eficaz da doença Sigatoka com base no melhoramento genético através da transformação de plantas ou seleção assistida por marcadores.

Palavras-chave: NBS-LRR RGAs, defesa, *Musa acuminata*, *Mycosphaerella musicola*, SSR, expressão gênica.

ABSTRACT

EMEDIATO, Flávia Leonel. DEVELOPMENT OF MICROSATELLITE MARKERS AND EXPRESSION ANALYSIS OF GENES INVOLVED IN BIOTIC STRESS RESPONSE IN *MUSA ACUMINATA* GENOTYPE. 2014. 129 Páginas. Thesis (Doctorate in Molecular Biology) – Universidade de Brasília, Brasília, DF, Brazil.

Commercial banana cultivars (*Musa* spp.) are grown across numerous countries in different tropical and subtropical environments. Conventional breeding strategies are limited due to sterility in many commercial cultivars. Asexually-driven evolution has resulted in a restricted genetic base, lacking resistance to pests and disease. Non-conventional breeding strategies, such as transformation and marker-assisted selection, offer an alternative approach for introgression of resistance genes in commercial cultivars. Resistance (R) genes in plants encode protein receptors that recognize molecular signatures of effectors of plant pathogens, triggering defense mechanisms of effector-triggered immunity (ETI). Such resistance can be lost during host and pathogen co-evolution, favouring the more rapid evolution of new pathogenic races. In this context, the continuous development of resistant plant cultivars is required in order to accompany pathogen evolution. The objective of this work was to identify genes potentially involved in responses to biotic stresses, which will serve as a resource for the design of specific primers for differential expression analysis during the interaction of *M. acuminata* with the pathogen *Mycosphaerella musicola* and development of SSR markers to search for polymorphisms in *M. acuminata* genotypes. From this, 14 expressed NBS-LRR R genes from Calcutta 4 and 25 in Cavendish Grande Naine. A total of 156 microsatellite markers were developed for genes potentially involved in resistance and defense responses, derived from different sources: 454-pyrosequencing derived leaf transcriptome sequences derived from the *M. acuminata* x *M. musicola* interaction, which mapped to BAC clone sequences from *M. acuminata* Calcutta 4 containing NB-ARC resistance gene analogs (RGAs); *M. acuminata* DH Pahang whole genome sequence regions containing conserved domains belonging to NB-ARC and other R genes; and 454-pyrosequencing transcripts from the *M. acuminata* x *M. musicola* interaction, enriched for defense genes. Among SSR loci evaluated, 21 (13.46%) were polymorphic, with alleles per locus ranging from 2 to 4, and polymorphism information content ranging from 0,31 to 0,67. The expression analysis of 51 target genes showed that 19 had specificity in real time PCR reaction and that most of the genes involved in defense processes were not modulated, with the exception of NAC and chorismate synthase gene. The NPR1 gene was expressed only in the susceptible cultivar and RGA 37, 43, M09A31, M09A21, B03A51 and B03A11 had increased expression when inoculated with the pathogen. The RGAs 43 and 37 showed significant induction in *M. acuminata* Calcutta 4 when inoculated with the pathogen and the RGAs 9330, 7100, 6160, M09A31 and M09A21 showed downregulation in Calcutta 4. This analysis was consistent with the apparent trend of previous individual genes R characterizations since some genes are induced by a variety of stimuli, exhibit some very specific responses and others seem to be responsive to the pathogen. The continued characterization of genes involved in response to biotic stress in *M. acuminata* during interaction with *M. musicola*, as well as the development of molecular markers, will contribute to the development of effective management of Sigatoka disease based on genetic improvement through plant transformation or marker assisted selection.

Keywords: NBS-LRR RGAs, defense, *Musa acuminata*, *Mycosphaerella musicola*, SSR, gene expression.

REVISÃO DA LITERATURA

1. Aspectos Botânicos e classificação taxonômica do gênero *Musa*

Apesar de comumente ser confundida com espécie arbórea devido ao seu grande porte, a bananeira é uma herbácea que pode alcançar até três metros de altura, sem as lignificações ou espessamentos secundários das hastes que acontecem nas árvores.

Taxonomicamente, a bananeira é classificada como membro da classe Liliopsida subclasse Zingiberidae, superordem Liliales, ordem Zingiberales, família *Musaceae* (Cronquist, 1960; Carvajal, 1991). A família monocotiledônea *Musaceae* é dividida em três subfamílias: Heliconoideae, Strelitzoideae e Musoideae.

Os gêneros *Ensete*, *Musa* e *Musella* pertencem à subfamília Musoideae, sendo que ao gênero *Musa* pertencem as bananeiras com frutos comestíveis. O gênero *Musa* tem sofrido grandes modificações taxonômicas ao longo do tempo. Em 1947, as espécies deste gênero com 11 pares de cromossomos foram agrupadas em duas seções denominadas *Eumusa* e *Rhodochlamys* enquanto que aquelas com 10 pares de cromossomos foram agrupadas nas seções *Australimusa* e *Callimusa* (Cheesman, 1947). Recentemente essas seções foram novamente revisadas e a seção *Rhodochlamys* foi incorporada a seção *Eumusa*, mudando a nomenclatura de *Eumusa* para *Musa*. A seção *Australimusa* foi incorporada pela seção *Callimusa*, mantendo a mesma nomenclatura da seção (Häkkinen e Markku, 2013).

Descrita por Stover e Simmonds (1987), a classificação dos genótipos da bananeira se deu com base na existência de dois cultivares selvagens denominados *M. acuminata* (A) e *Musa balbisiana* (B). Posteriormente, os estudos de Simmonds e Shepherd (1955) levaram à constatação de grupos genômicos diplóides AA e AB; triplóides AAA, AAB e ABB; tetraplóides AAAA, AAAB, AABB e ABBB, sendo esta classificação adotada em todo o mundo. Além dos grupos genômicos, foi estabelecido o uso do termo subgrupo para denominar um complexo de cultivares, originários de mutações de um único cultivar original, como nos casos dos grupos AAA, subgrupo Cavendish, e AAB, subgrupos Prata e Terra.

A maioria dos cultivares comerciais são triplóides provenientes do grupo genômico AAA, sendo neste caso considerados *M. acuminata*. Também podem ser AAB e ABB, possuindo uma composição genômica mista. Apenas um pequeno grupo de cultivares são diplóides (AA) ou tetraplóides (AAAB), porém essas possuem cultivo restrito e alta suscetibilidade a doenças.

2. Importância sócio-econômica

Amplamente distribuída por todo o globo, a banana possui o sudeste Asiático, que se estende da Índia a Papua Nova Guiné, como o centro de origem (Simmonds, 1962), sendo a Malásia e a Indonésia considerados os centros da diversidade (Daniells *et al.*, 2001). Atualmente nessa região ainda existe um grande número de bananas selvagens com sementes (Frison e Sharrock, 1998).

As bananas ranqueam com arroz, trigo e milho em termos de importância (Frison e Sharrock, 1998). Além de ser um importante produto comercial em todo o mundo, mais de 85% das bananas são cultivadas para o consumo local em regiões tropicais e subtropicais. O Brasil é um dos principais produtores de banana com o equivalente a quase 10% do total da produção global, ficando atrás apenas da Índia, China, Filipinas e Equador (Faostat, 2013) (Tabela 1).

Tabela 1: Principais produtores mundiais de banana no ano de 2011. Fonte: Adaptado de FAOSTAT | © FAO Statistics Division 2013 | 01 October 2013

Rank	Área	Produção (MT)
1	Índia	29666973
2	China, Mainland	10400000
3	Filipinas	9165043
4	Equador	7427776
5	Brasil	7329471
6	Indonésia	6132695
7	Tanzânia	3143835
8	Angola	2646073
9	Guatemala	2679934
10	México	2138687
11	Costa Rica	1937122
12	Burundi	1848727
13	Colômbia	2042925
14	Tailândia	1600000
15	Vietnã	1523428
16	Camarões	1394675
17	Quênia	1197988
18	Egito	1054243
19	Papua Nova Guiné	1000000
20	República Dominicana	829827

MT: measurement ton. 1MT= 1.133 m³

Países da África enfrentam uma necessidade cada vez maior de alcançar a sustentabilidade na produção agrícola, porém, a ameaça da fome acaba sendo um grande obstáculo. Existem regiões na Etiópia que sobreviveram e sobrevivem à fome devido a utilização do gênero *Ensete* como parte do sistema de subsistência, demonstrando a grande importância desse produto. Ele ajuda a alimentar cerca de 10 milhões de pessoas na Etiópia e Eritreia (Brandt *et al.*, 1997).

Do ponto de vista da fruticultura mundial, a área plantada de banana corresponde a 5.279.638 ha ficando atrás somente da área plantada de uva, com 7.060.245 ha (Faostat, 2013).

No Brasil, o estado de São Paulo, com 1.354.528 toneladas dessa fatia, é o maior produtor no país, seguido pelo estado da Bahia (1.239.650 de toneladas), Minas Gerais (654.566 de toneladas) e Santa Catarina (650.518 de toneladas) (IBGE, 2011). O desempenho da cultura de banana no Brasil tem aumentado muito nas últimas décadas, chegando a uma área colhida de 524.750 ha (Figura 1).

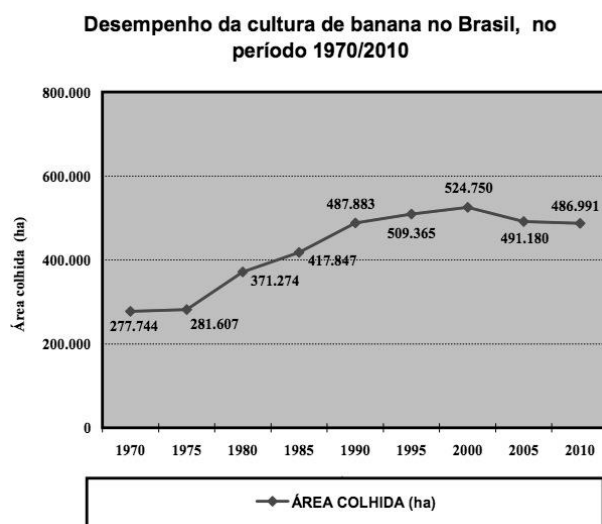


Figura 1: O desempenho da cultura de banana no Brasil 1970 - 2010. Fonte: Adaptado de IBGE - Produção Agrícola Municipal, 2011. Consultado em 08/11/2013.

3. Estresses Bióticos

Estresses bióticos e abióticos influenciam negativamente a sobrevivência, a produção de biomassa e o rendimento da cultura (Agarwal *et al.*, 2006). Os patógenos de plantas são uma ameaça real para a agricultura mundial, pois causam doenças que diminuem drasticamente as colheitas dependendo da gravidade da doença (Wani, 2010).

Estresses bióticos incluem fungos, bactérias, vírus, nematóides e insetos. Doenças de etiologia bacteriana também contribuem com sérias perdas no cultivo de banana. O agente causador da doença Moko em bananeiras, *Ralstonia solanacearum*, possui uma grande gama de hospedeiros, podendo infectar mais de 250 espécies de plantas (Elphinstone, 2005), entre elas o tomate, tabaco e batata (Norman *et al.*, 2009). O entendimento dos processos envolvendo este patógeno, como por exemplo, patogenicidade, número de hospedeiros, regulação de genes de virulência, imunidade, está bem avançado já que esse foi um dos primeiros genomas de bactérias de plantas sequenciados e anotados (Peeters *et al.*, 2013).

A murcha bacteriana, causada pela *X. campestris pv. Musacearum*, é considerada a maior ameaça à produção de banana na África Oriental e Central (Tripathi *et al.* 2009). A doença é muito destrutiva, infectando todas as variedades de banana, incluindo as de sobremesa, cozinhar e assar (Ssekiwoko *et al.* 2006). Estudos com transgenia têm resultados promissores no combate à murcha bacteriana (Namukwaya, *et al.*, 2012).

Mais de 70% de todas as doenças das principais culturas são causadas por fungos (Agrios, 2005). Na bananicultura ocorrem atualmente aproximadamente 50 patógenos fúngicos, porém os mais importantes são *Mycosphaerella fijiensis* causador da Sigatoka negra, *Mycosphaerella musicola* causador da Sigatoka amarela e *Fusarium oxysporum f. sp. cubense*, responsável pelo Mal do Panamá (Jones, 2000).

O Mal do Panamá, causado pelo fungo necrotrófico *F. oxysporum f. sp. cubense* (Vajna, 1985) é considerado uma das doenças mais devastadoras do mundo (Bai *et al.*, 2013). O fungo penetra as raízes de bananeiras por meio de pequenas aberturas na parede celular (Li *et al.*, 2011) e mata as células e tecidos hospedeiros para se nutrir de células mortas. Com a proliferação do fungo no tecido vascular, ocorre a interrupção da translocação de água deixando a planta com um aspecto murcho, entre outros sintomas, podendo levar a necrose e até a morte da planta (Beckman, 1990).

Antes de 1960, a cultivar 'Gros Michel' foi a principal variedade de banana doce. No entanto, um surto mundial da doença Mal do Panamá causada pela raça 1 do fungo patogênico *F. oxysporum f. sp. cubense* (Foc), levou à destruição mundial de bananas entre 1940 e 1960. O sub-grupo 'Cavendish' por ser mais resistente à Foc raça 1, foi desenvolvido para substituir Gros Michel, e hoje são as cultivares de bananeira mais amplamente distribuídas. Na década de 90, no entanto, cepas virulentas de Foc da raça 4, particularmente raça tropical 4 (Foc TR4), ao qual Cavendish é suscetível, evoluíram e espalharam-se rapidamente para áreas de produção de Cavendish, causando perdas substanciais em muitos países da Ásia e da Austrália (Li *et al.*, 2013). Em 2013, Foc TR4 foi relatado em

Moçambique, mostrando que o patógeno chegou no continente africano (Durroux, 2013). Como a banana cultivada é geralmente propagada assexuadamente usando ventosas de rizomas, o cultivo em larga escala de um único genótipo é particularmente vulnerável a patógenos. A falta de diversidade e variedades de banana que são resistentes a Foc TR4 e a dificuldade no processo de criação de novas variedades de banana têm levantado uma séria preocupação de que cultivares suscetíveis de banana estejam ameaçadas de desaparecer novamente (Li *et al.*, 2013).

3.1. *Mycosphaerella fijiensis*, *Mycosphaerella musicola* e as doenças de Sigatoka

A Sigatoka amarela, causada pelo fungo *M. musicola*, foi descrita pela primeira vez em Java por Zimmerman em 1902, porém, foi em Fiji, na região do Vale da Sigatoka em 1912 – 1923 que a doença tomou proporções de epidemia (Jones, 2000; Fullerton, 1994; Mourichon e Fullerton, 1990).

A Sigatoka negra, cujo agente causal é o fungo *M. fijiensis* Morelet (Mf), um fungo sexual, heterotálico, que possui *Pseudocercospora fijiensis* como a fase anamorfo (Churchill, 2011), está entre as doenças de plantas cultivadas de maior importância no mundo (Pennisi, 2010). Foi descrita pela primeira vez em Fiji em 1963, no entanto, a doença foi difundida pelo pacífico antes mesmo da sua descoberta oficial (Stover e Simmonds, 1987). Além das perdas de área foliar fotossintéticas e de perdas de produtividade, que podem alcançar cerca de 20–80% na ausência de fungicidas, esta doença provoca uma alteração de comportamento fisiológico do fruto na pré- e pós-colheita, como, por exemplo, modificação no conteúdo de hidratos de carbono (amido, açúcares solúveis, oligossacarídeos, amilose), conteúdo de compostos fenólicos, fitormônios e atividades enzimáticas (Saraiva *et al.*, 2013).

As Sigatokas afetam seriamente os cultivares de banana mais populares em todo o mundo, como é o caso das bananas de sobremesa (AAA e alguns genomas AAB) que pertencem ao subgrupo Cavendish Grande Naine e normalmente são cultivadas em monoculturas (Churchill, 2011). O controle dessas doenças é baseado no uso de fungicidas o que tem causado um círculo vicioso, pois, com o aumento de produção e área plantada, há uma maior necessidade do uso de químicos, gerando seleção de patógenos resistentes, chegando até a 66 aplicações de fungicidas por ano em alguns locais (Kema, 2006).

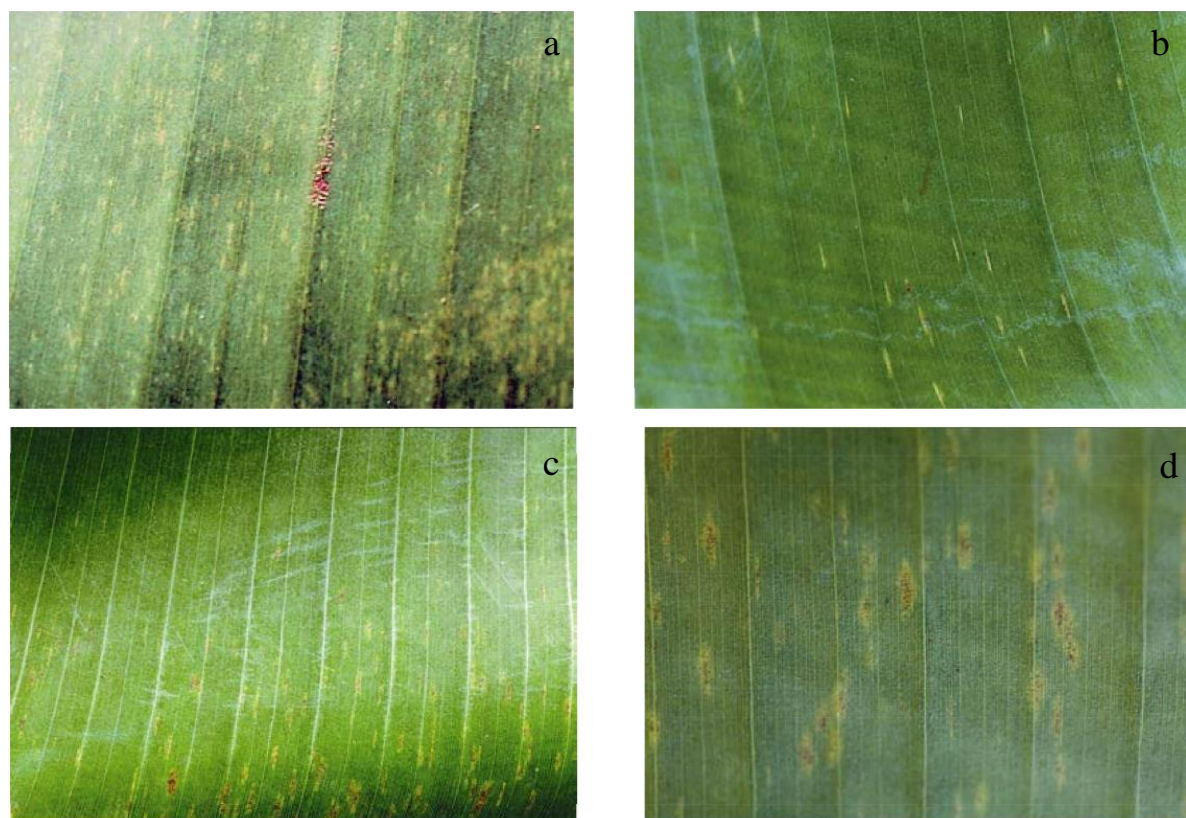
Apesar da semelhança sintomatológica entre a Sigatoka negra e a Sigatoka amarela, há várias diferenças que podem ser facilmente observadas. Enquanto que na Sigatoka amarela as folhas infectadas apresentam estrias de cor clara na face superior da folha, com

lesões de frequência bem mais baixas e bordos regulares de forma elíptica, na Sigatoka negra as estrias são marrons e visíveis na face inferior da folha, com frequência de lesões bem mais alta (Ferrari *et al.*, 2008). Para entender a sintomatologia da doença, os sintomas foram organizados em estágios que podem ser diferentes nas Sigatokas negra e amarela, como mostrado na Tabela 2.

Tabela 2: Resumo dos diferentes estágios de lesão associados às doenças Sigatokas negra e amarela

Estágio da lesão	Sigatoka amarela	Sigatoka negra
Estágio 1	Pequenos pontos amarelos em forma de listra de até 1 mm de comprimento	Pequenos pontos pigmentados de branco ou amarelo, similar a Sigatoka amarela nesta fase
Estágio 2	Estrias amareladas em vários mm de comprimento	Estrias de coloração marrom visíveis na parte inferior da folha, mais tarde visíveis na parte superior da folha como estria amarela inicialmente, progredindo para marrom e preto.
Estágio 3	Mancha marrom alongada com borda bem definida	Continuidade do estágio 2, com maior número de estrias
Estágio 4	mancha madura, com centro afundado marrom escuro, muitas vezes cercada por um halo amarelo. Conidióforos e conídios são produzidos nesta fase	Aparecimento da mancha parda na face abaxial e mancha negra na superfície da folha
Estágio 5	Desenvolvimento de uma mancha cinza com centro seco e um anelo preto periférico visível mesmo depois que a folha secou	Mancha elíptica totalmente preta, rodeada por um halo amarelo na superfície abaxial da folha
Estágio 6		Centro de lesão seco e cinza, rodeado por um halo bem definido amarelo brilhante

Os estágios podem ser visualizados *in vivo* na Figura 2.



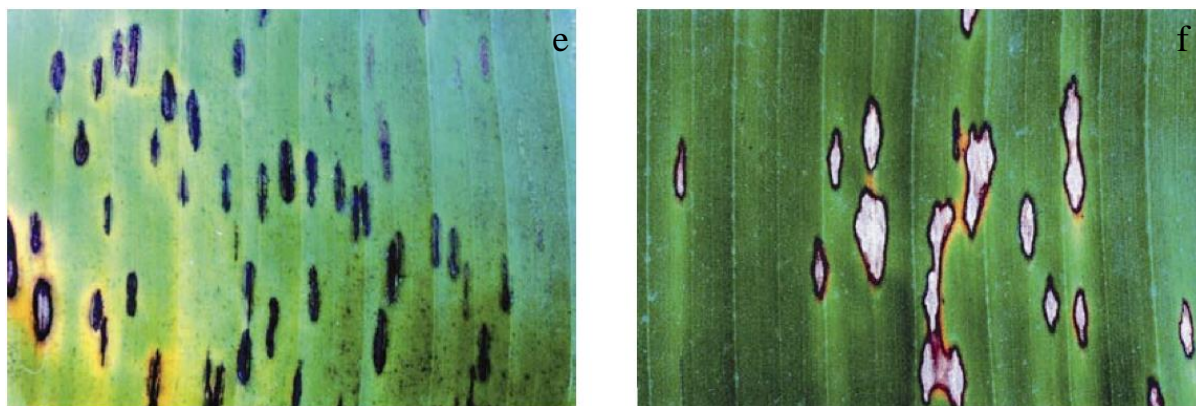


Figura 2: Representação dos cinco estágios da Sigatoka amarela. (a-f). (a) - Estágio 1 caracterizado por pontos amarelos; (b) – Lesões associadas ao estágio 2 inicial, caracterizado por estrias amareladas; (c) – Estágio 2 tardio caracterizado por mudança de cor nas estrias de amarelo para marrom ferrugem; (d) – Estágio 3 com alongamento e alargamento das estrias; (e) – Estágio 4, estrias se tornam manchas; (f) – Estágio 5 com centro seco e cinza e anel preto circundando a lesão. Fonte: adaptado de *Pest and Diseases Image Library*

As Sigatokas negra e amarela são propagadas por meio de dois tipos de esporos, conhecidos como conídios e ascósporos (Figura 3).

Os conídios ou esporos assexuais se formam nos ápice dos conidióforos e ocorrem a partir dos primeiros estádios da lesão na folha (face abaxial), estes se desprendem dos conidióforos por ação da água e/ou vento. A maioria das infecções de *M. fijiensis* e *M. musicola* começam com o depósito desses esporos sobre a folha de bananeiras suscetíveis. Os esporos germinam no prazo de 2-3 horas depois de serem depositados sobre a superfície da folha, em condições de umidade e temperatura ótima para a germinação de 27 °C para *M. fijiensis* e 25-29 °C para *M. musicola*. O tubo germinativo depois cresce durante vários dias (2-3 dias para *M. fijiensis* e 4-6 dias para *M. musicola*) antes de penetrar a folha através dos estômatos (Meredith 1970; Stover 1980). Uma vez no interior da folha, hifas começam a se desenvolver e crescer de maneira que emergem através de estômatos na estria que se desenvolveu, entrando novamente na folha por estômatos próximos.

Os ascósporos ou esporos sexuais se formam em grande número posteriormente em manchas mais evoluídas de coloração branco-acinzentado principalmente nas folhas mortas ou necrosadas. Esta é considerada a fase mais importante na reprodução da doença, devido à alta produção e disseminação desses esporos pelo vento a grandes distâncias.

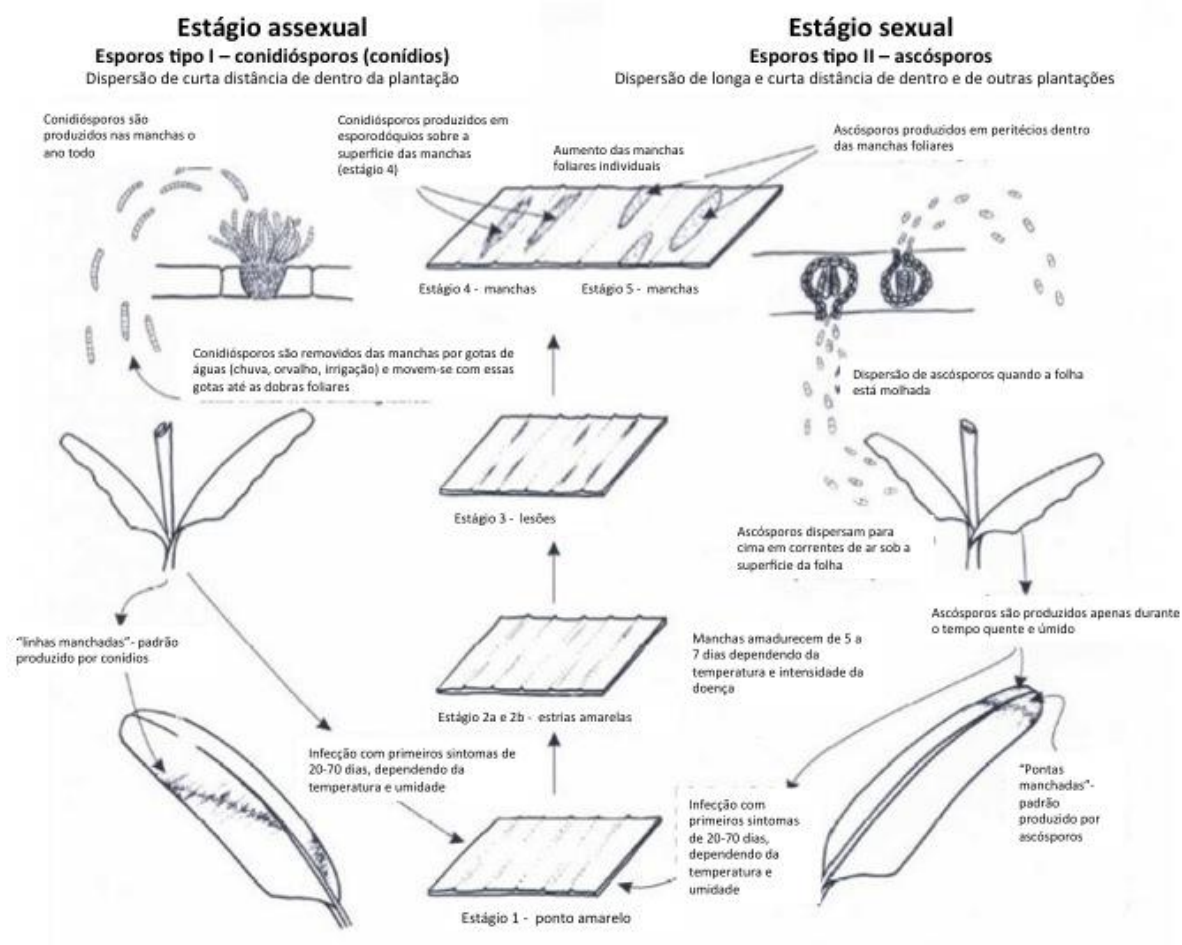


Figura 3: Ciclo de vida de *M. musicola*, o patógeno causador da Sigatoka amarela. Fonte: adaptado do Departamento de Indústrias Primárias e Pescas, Queensland,

É possível afirmar que a agressividade e a virulência da Sigatoka negra é muito superior a da Sigatoka amarela (Young e Conie, 2001). Isto ocorre devido, entre outras coisas, a diferenças em seus ciclos assexuais. O ciclo de vida da *M. fijiensis* é bem mais rápido que da *M. musicola*, o que proporciona um tempo mais curto para infecção, como é observado no aparecimento adiantado das necroses de Sigatoka negra, quando comparada à Sigatoka amarela (Stover, 1980).

Existe um período de tempo relativamente longo antes que quaisquer reações biotróficas incompatíveis sejam observadas nos cultivares suscetíveis. O patógeno coloniza o tecido foliar, crescendo intercelularmente sem a produção de haustórios, por quase um mês. Durante este período, a pouca evidência da presença do agente patogênico pode ser detectada externamente. Alterações citológicas são visíveis nas células do parênquima, após cerca de 28 dias, embora as células ainda parecem saudáveis. Externamente, no estágio 2 estrias

iniciais são visíveis (Beveraggi *et al.*, 1995; Sallé *et al.*, 1989). Após essa fase, iniciará uma estreita associação entre *Mycosphaerella* e *Musa* por meio do desenvolvimento de hifas e haustórios especializados dentro de células vivas da planta, de onde os nutrientes são absorvidos, caracterizando a fase biotrófica do fungo. Posteriormente inicia-se o processo de secreção de enzimas e toxinas matando a célula hospedeira e o fungo passa a se alimentar de nutrientes liberados do tecido morto (Koeck *et al.*, 2011).

4. Recursos Genéticos e Melhoramento em *Musa*

Os pesticidas são amplamente utilizados na produção agrícola para prevenir ou controlar as pragas, doenças e ervas daninhas, num esforço para reduzir ou eliminar as perdas de produção e de manter uma elevada qualidade do produto. Embora os pesticidas sejam desenvolvidos através de processos muito rigorosos de regulação para funcionar com razoável certeza, e um impacto mínimo sobre a saúde humana e o meio ambiente, sérias preocupações foram levantadas sobre os riscos à saúde decorrentes da exposição ocupacional e de resíduos em alimentos e água potável (Damalas e Eleftherohorinos, 2011). Além da saúde humana, pesticidas também tem causado efeitos tóxicos na fauna de diversos locais, como é o caso do jacaré (*Caiman crocodylus*). Alguns autores apontam que o uso de pesticidas em plantações de banana está impactando a espécie que possui alto nível trófico e habita uma das áreas selvagens mais importantes da Costa Rica (Grant *et al.*, 2013).

Levando em consideração a grande preocupação com o uso de agrotóxicos, uma alternativa vem sendo utilizada no controle de doenças, como é o caso da busca de variedades resistentes a estresses bióticos, seja mediante a seleção dentro dos recursos genéticos existentes, seja mediante a geração de novas variedades por hibridização com o intuito de diminuir o uso de pesticidas.

Para auxiliar os estudos e alavancar as pesquisas, um sistema de informações de germoplasma de *Musa* (MGIS - *Musa Germplasm Info. System* - <http://www.Musanet.org/>) foi desenvolvido com o objetivo de aumentar o conhecimento sobre a diversidade de *Musa*, ajudar na conservação e melhorar o uso de recursos genéticos de banana. O banco de dados contém informações de classificação botânica, descrições morfo-taxonômicas, características agronômicas, informações sobre resistência à doenças, tolerância à estresses, marcadores moleculares de 5188 acessos organizados em quase 100 coleções por quase todo o globo. Neste banco de dados pode ser consultado sobre a identidade, origem, características e

distribuição dos acessos individuais nas coleções, permitindo que as instituições participantes em todo o mundo possam compartilhar e comparar os seus dados. O banco de dados também é particularmente útil para vários usuários, ou seja, criadores, pesquisadores e comunidades de agricultores, para localizar fontes alternativas de germoplasma de banana e identificar os acessos mais adequados com características específicas de interesse.

Além do MGIS, outros recursos ainda estão disponíveis, como as bibliotecas BAC feitas de DNA de *M. acuminata* ‘Calcutta 4’ e *M. balbisiana* ‘Pisang Klutuk Wulung’ (Pifanelli *et al.*, 2008) com informações de sequências de DNA, com anotação e curadoria do *Musa genomics website*; Análises de 3 milhões de sequências de DNA de *Musa*, derivados de extremidades de BACs, mostrando muitas características da sua organização e estrutura genética (Cheung e Town, 2007).

Embora exista uma grande variedade de cultivares de bananas, o melhoramento genético em *Musa*, em comparação com outras grandes culturas alimentares, tem sido limitada. Cultivares têm evoluído desde diplóides, triplóides e tetraplóides selvagens asiáticos de espécies de *M. acuminata* (genoma A) e *M. balbisiana* (genoma B). Espécies selvagens são geralmente seminíferas, ao passo que hoje em dia a maioria dos cultivares comerciais são estéreis ou triplóides e diplóides, com frutas que se desenvolveram por meio de partenocarpia. Como esta cultura tem evoluído principalmente através de reprodução vegetativa assexuada, a variação genética é limitada, tornando-se uma cultura vulnerável a pragas e doenças.

As progênies obtidas a partir de cruzamentos entre genótipos de bananeira apresentam tamanho pequeno e são compostas por um número variável de ploidias, inclusive aneuploides. Esse fato dificulta o melhoramento e impõe barreiras ao desenvolvimento de novas cultivares. A produção de tetraplóides a partir de triplóides tem sido a metodologia básica aplicada desde o início dos trabalhos de melhoramento de bananeira. A técnica foi usada inicialmente para a produção de híbridos tetraplóides a partir de triplóides AAA (Gros Michel) polinizados com pólen A ($3X \times 2X = 4X$) e atualmente vem sendo amplamente empregada em cultivares AAB (Silva *et al.*, 2005).

As variedades triplóides apresentam fertilidade residual que permite a obtenção de sementes em cruzamentos com diplóides. A particularidade deste cruzamento está no fato de que os alelos da planta-mãe podem não segregar e, portanto, as características maternas são mantidas, resultando em híbridos tetraplóides (Silva *et al.*, 2013). De acordo com Menendez e Shepherd (1975), é possível acrescentar um conjunto de cromossomos de um diplóide para transferir características de resistência às doenças ao novo híbrido. Mesmo com diferentes possibilidades de segregação, apenas os híbridos tetraplóides (44 cromossomos) apresentam

características agronômicas que possibilitem sua adoção pelos agricultores, entre as quais sabor e aroma.

Os cruzamentos são possíveis mesmo envolvendo cultivares consideradas estéreis, assim, Aguilar Morán (2011) obteve sementes a partir da polinização de 20.000 cachos de cultivares do subgrupo Cavendish (Grande Naine e Williams) com pólen de 10 diplóides. Como resultado deste esforço, foram obtidas apenas 200 sementes com 40 embriões, e 20 híbridos tetraplóides foram recuperados, os quais foram cruzados com o diplóide SH3142, originando dois híbridos triplóides semelhantes às cultivares originais. De maneira geral, a maioria das sementes obtidas por meio dos cruzamentos entre diplóides e triplóides ou tetraplóides apresenta alguma anomalia e além disso, a taxa de germinação das sementes obtidas é em torno de 20% em condições de casa de vegetação. Anormalidades na meiose também são muito frequentes em diplóides e triplóides, em especial associadas com o pareamento cromossômico. Em tetraplóides, a formação de gametas $2n$ ou $4n$ é frequentemente relatada na literatura. Todas essas irregularidades dificultam a previsão da segregação observada nas progênes (Silva *et al.*, 2013).

Todas essas dificuldades tem levado ao desenvolvimento de novas técnicas de melhoramento para a criação de cultivares resistentes de forma a complementar e dar suporte às atividades convencionais de melhoramento.

A atenção e os esforços estão voltados para busca de genes de importância em cultivares selvagens, fontes de resistência, como por exemplo *M. acuminata* ‘Calcutta 4’, para que posteriormente possam ser transferidos à cultivares comerciais.

A partir dos recursos genéticos disponíveis, muitos trabalhos tem sido realizados, como a caracterização de componentes de DNA repetitivos do genoma de *Musa*, representando uma parte substancial dos 550 Mpb do DNA (Hřibová *et al.*, 2007 e Hřibová e Dolezel, 2011); Transcritoma durante a interação entre *M. acuminata* e *M. musicola*, caracterizando um grande número de *unigenes* de *M. acuminata* ‘Calcutta 4’ e ‘Cavendish Grande Naine’, aumentando o número de domínio público de *Musa* ESTs (Passos *et al.*, 2013); Genoma de *M. balbisiana*, representando um valioso recurso para apoiar o estudo do metabolismo em inter e intra-específicos de híbridos triplóides de *Musa* (Davey *et al.*, 2013); Genoma de *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang), servindo de referência para outros estudos (D'Hont *et al.*, 2012).

A utilização efetiva de marcadores tem grantido sucesso nos programas de melhoramento. Um dos principais usos de marcadores moleculares na pesquisa agrícola tem sido na construção de mapas de ligação para diversas espécies de culturas. Mapas de ligação

foram utilizados para a identificação de regiões cromossômicas que contêm genes que controlam características simples (controlada por um único gene) e características quantitativas usando análises de QTL (*Quantitative Trait Loci*). Os QTLs são regiões dentro genomas que contêm genes associados a uma característica quantitativa agronomicamente importante como produtividade, qualidade e algumas formas de resistência (Agarwal *et al.*, 2008; Collard *et al.*, 2005).

Outra aplicação importante dos marcadores moleculares se encontra na transferência de alelos de resistência de fontes exóticas e muitas vezes não adaptadas para cultivares elite. Esses marcadores, se estreitamente ligados aos alelos de resistência, podem ser usados na seleção assistida por marcadores (SAM), nas etapas do melhoramento (Alzate-Marin *et al.*, 2005). Nos últimos anos, marcadores têm se mostrado uma ferramenta poderosa para a estimativa da diversidade genética (variação na sequência de nucleotídeos, a estrutura genética, cromossomos e genomas inteiros) e das relações filogenéticas de espécies com base na conservação da sequência (Kalia *et al.*, 2011).

Em *Musa*, esses marcadores são amplamente utilizados para diferentes fins. Mukunthakumar e colaboradores (2013) utilizaram 43 acessos de *M. acuminata* para estudo de diversidade genética e diferenciação usando RAPD. Hrahsel e colaboradores *et al.* (2013) utilizaram marcadores RAPD para avaliar a estabilidade genômica *in vitro* de plântulas regeneradas de *M. acuminata* cv. Vaibalhla num processo de propagação *in vitro* tendo aplicação na produção em massa, conservação e estudos de transformação genética. Além desses, marcadores, AFLPs também foram utilizados (Loh *et al.* 2000; Wong *et al.* 2001a; Ude *et al.* 2002; Wang *et al.* 2007). Ambos estes marcadores têm um nível relativamente elevado de polimorfismo, porém são dominantes, e, no caso de RAPD, a reprodutibilidade pode ser uma limitação grave (Jones *et al.* 1997).

Marcadores co-dominantes também foram utilizados em *Musa*, como RFLPs (Gawel *et al.* 1992; Nwakanma *et al.* 2003; Ning *et al.* 2007) e SSRs (Kaemmer *et al.*, 1997; Grapin *et al.*, 1998; Lagoda *et al.*, 1998; Buhariwalla *et al.* 2005). Enquanto RFLPs possuem um bom desempenho em termos de reprodutibilidade, por outro lado tem um nível relativamente baixo de polimorfismo e são difíceis de utilizar. Pelo contrário, os marcadores microssatélites superam os RFLP e RAPD em todos os aspectos acima mencionados. Jesus e colaboradores (2013) determinaram a constituição genômica baseada em regiões polimórficas ITS (*Internal Transcribed Spacer*) em cerca de 220 acessos de *Musa*, investigando a estrutura populacional utilizando SSRs. Librelon e colaboradores (2013) analisou a diversidade genética de 20 clones de Prata-Anã Gorutuba, quatro clones de Prata-Rio e quatro clones de Prata-Catarina

por meio de marcadores SSR indicando a eficiência desses marcadores na discriminação e caracterização dos clones de banana. Davey *et al.* (2013) alinhou sequências do genoma B contra o genoma A de *Musa* e identificou a divergência entre os dois genomas por meio de SNPs. Além disso, uma abordagem robusta para genotipagem em *Musa* baseado em 19 locos microssatélites identificados com *primers* marcados com fluorescência por meio de separação por eletroforese capilar de alta resolução, tornou-se disponível recentemente (Hřibov *et al.*, 2013). Passos e colaboradores (2012) identificaram 624 locos microssatélites derivados de quatro bibliotecas contrastantes em resistência de *M. acuminata* Calcutta 4 e Cavendish Grande Naine e Passos e colaboradores (2013) identificaram 4068 locos microssatélites gênicos de Calcutta 4 e 4095 de Cavendish Grande Naine derivados de transcritos da interação *Musa* e *Mycosphaerella*.

5. Imunidade em plantas

Como em animais, as plantas estão continuamente expostas ao ataque de patógenos. Porém, tendo em vista a inexistência de resposta imune mediada por anticorpos, as plantas desenvolveram durante o processo de evolução, mecanismos diferenciados de defesa que, quando acionados (na maioria das vezes por fungos, bactérias e vírus) percebem a agressão, traduzindo essa percepção em uma resposta apropriada (Pieterse *et al.*, 2005; Shewry e Lucas, 1997; de Wit, 2007). Por outro lado, os fitopatógenos podem usar diversas estratégias de vida liberando moléculas efetoras (efetores de virulência) na célula da planta, garantindo uma maior aptidão do patógeno (Jones e Dangl, 2006).

As plantas, ao contrário dos mamíferos, não possuem células defensoras móveis e um sistema imunológico somático adaptativo (imunidade adquirida), plantas dependem exclusivamente da resposta inata. A ativação de respostas de defesa da planta é desencadeada pelo reconhecimento de organismos invasores pelos receptores imunes. Estes podem ser ligados a receptores de membrana que controlam o ambiente extracelular ou de receptores intracelulares que detectam a presença ou a atividade de efetores derivados do agente patogênico (Lapin e Van den Ackerveken, 2013).

No início do século 20, através da realização de experiências genéticas, Harold Flor mostrou que a imunidade da planta, bem como a capacidade do patógeno de causar doenças, são controladas por pares de genes correspondentes. O fator genético da planta foi referido como gene de resistência (R), enquanto que o fator genético do agente patogênico, foi

referido como gene de avirulência (AVR). Para Flor, a resistência gene-a-gene era conferida por uma interação direta entre o gene R e proteínas Avr (Thomma *et al.*, 2011).

No entanto, experimentos projetados para mostrar tais interações diretas receptor-ligante frequentemente geram resultados negativos o que levou a formulação da hipótese guarda, onde o reconhecimento ocorre indiretamente por meio da interação entre efetor e uma segunda proteína denominada *guarda*. Essa interação é percebida quimicamente pela proteína R (proteína guarda) ativando a resistência do hospedeiro (Głowacki *et al.*, 2011). Na falta de um gene funcional R, a seleção natural conduz um *guarda* para diminuir a sua afinidade de ligação com o efetor e, assim, evitar a detecção do efetor. No entanto, na presença de um gene funcional R, a seleção natural favorece *guardas* com maior afinidade de interação com o efetor, melhorando a percepção do patógeno. Estas duas pressões de seleção (Figura 4) conflitantes sobre a mesma superfície de interação efetoras resulta em situação evolutivamente instável dos *guardas* que poderia ser flexibilizada mediante a evolução de uma proteína do hospedeiro, denominado aqui "decoy", que é especializada na percepção dos efetores pela proteína R, mas que sozinha não tem nenhuma função no desenvolvimento da doença ou resistência (Hoorn e Kamoun, 2008).

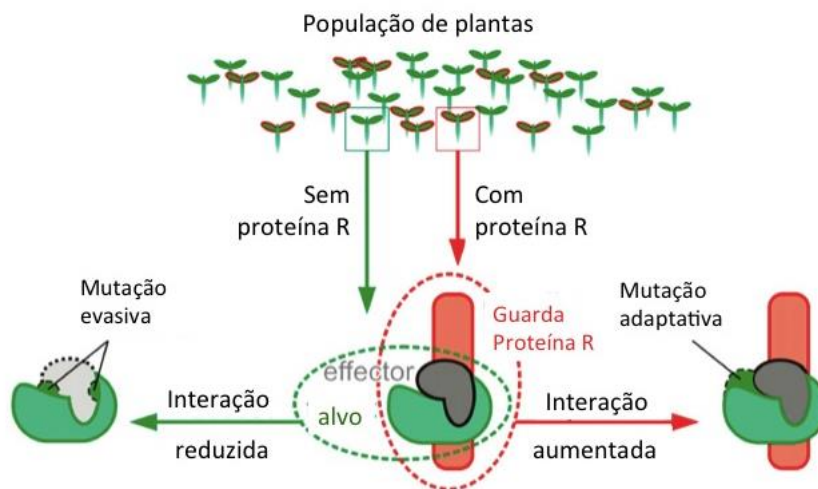


Figura 4: Forças opostas de seleção sobre moléculas guarda de proteínas R. Forças opostas de seleção são esperadas para operar em alvos efetores guardados em plantas com ou sem proteína R associada. Na ausência da proteína R (setas verdes) os alvos estarão sob pressão seletiva para reduzir a interação (esquerda). Na presença da proteína R (setas vermelhas), o alvo efetor guardado estará sob pressão seletiva para melhorar a interação com o efetor e melhorar a percepção do patógeno (direita). A duplicação de genes do alvo efetor ou a evolução independente do alvo reduziria as restrições evolutivas impostas ao alvo efetor guardado, permitindo que ele se especialize como um co-receptor (*decoy*), que regula a ativação da proteína R. Fonte: Adaptado de Hoorn e Kamoun, 2008

No modelo *decoy*, proteínas específicas que são similares as proteínas alvo do elicitor são geradas pela planta durante algumas interações planta-patógeno. Sua única função é de se

ligar ao efetor e atuar como um mediador nas interações com proteínas R (Hoorn e Kamoun, 2008).

Segundo Jones e Dangl (2006), a visão mais atual do sistema imunológico da planta pode ser representado por um modelo de quatro fases denominado “zig-zag” (Figura 5) em que na fase 1 PAMPs são reconhecidos por PRRs resultando na imunidade disparada por PAMPs (PTI); na fase 2 patógenos que obtiveram sucesso liberam efetores que interferem na PTI, desencadeando uma suscetibilidade disparada por efetores (ETS); na fase 3 o efetor é especificamente reconhecido por uma proteína R resultando em imunidade disparada por efetores (ETI).

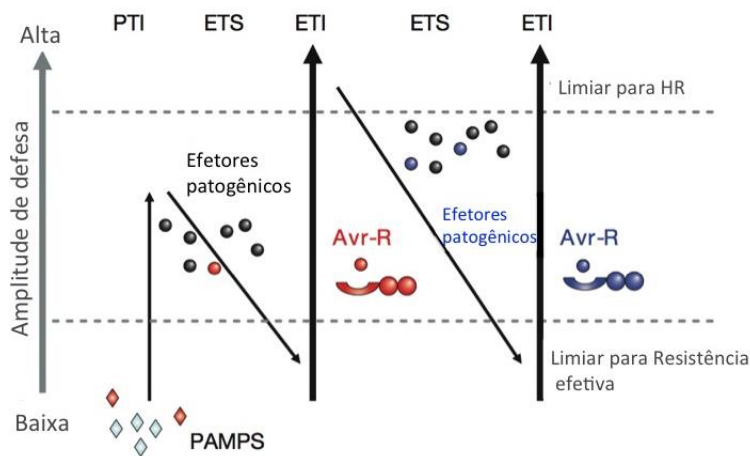


Figura 5: Modelo Zig-Zag do sistema imune em plantas, proposto por Jones e Dangl (2006). Neste esquema, as plantas detectam padrões moleculares associados a patógenos (PAMPs / mAmps, losangos vermelhos) via PRRs para desencadear a imunidade disparada por PAMPs (PTI). Patógenos de sucesso produzem efetores que interferem com a PTI, resultando em suscetibilidade desencadeada por efetores (ETS). Se um efetor (indicado em vermelho) é reconhecido por uma proteína da planta NB-LRR, a imunidade disparada por efetores (ETI) é ativada, sendo uma versão amplificada de PTI que permite a indução da morte celular e resposta de hipersensibilidade (HR). Isolados do patógeno podem evoluir, perdendo efetores (em vermelho) ou ganhar novos efetores (em azul), permitindo que o patógeno suprima ETI. Tal seleção, então, favorecem novos alelos NB-LRR a evoluir e reconhecer um dos efetores recém-adquiridas, resultando novamente em ETI.

Os mecanismos utilizados pelas plantas para reconhecer patógenos e a ativação de respostas imunes estão sendo elucidados nos últimos anos. As plantas possuem dois tipos distintos de receptores imunes. O primeiro tipo, os receptores de reconhecimento de padrões (PRRs), reconhece padrões moleculares conservados associados à patógenos (PAMPs) iniciando um processo de imunidade conhecido por PTI (*PAMP – triggered immunity*). O segundo tipo são as proteínas de resistência (R), que reconhecem efetores do patógeno e iniciam a imunidade conhecida com ETI (*Effector – triggered immunity*) (Qi *et al.*, 2011).

5.1. Os receptores PRRs e a imunidade disparada por PAMPs – PTI

A "primeira linha" de defesa das plantas pode ser acionada por meio do reconhecimento de PAMPs, resultando em ativação da defesa basal (ou não hospedeiro) da planta conhecida como PTI ou imunidade disparada por PAMPs (Deller *et al.*, 2011).

Diversas moléculas microbianas são conhecidas como PAMPs. Lipooligossacarídeos de bactérias gram-negativas, flagelina bacteriana, glucanas bacterianas, glicoproteínas de oomicetos, quitina da parede celular de fungos são apenas alguns exemplos (Zhang e Zhaou, 2010). PAMPs são detectados pelos receptores PRRs, normalmente receptores quinases RLK localizados na superfície celular, RLPs (*receptor-like proteins*) e GBP (*glucan-binding protein*) (Fritz-Laylin *et al.*, 2005).

Os receptores RLK formam uma das maiores famílias de genes em plantas e em geral são proteínas transmembranas com domínio N-terminal extracelular bem versátil e um domínio quinase intracelular no C-terminal. São encontrados distintos domínios extracelulares envolvidos em reconhecimento de patógeno (Figura 6), sendo eles:

1. LRR-RLK (*Leucine-rich repeat receptor-like protein kinases*): Suas características estruturais são responsáveis pela função PRR, onde, o motivo extracelular LRR muito variável, participa frequentemente na interação proteína-proteína. Além dos receptores PRR, o domínio LRR também é encontrado com grande frequência em proteínas R de plantas (Kobe e Deisenhofer, 1995). Nessa classe estão presentes os principais receptores PRRs, os FLS2 (Gómez-Gómez e Boller, 2000).

2. LecRKs (*lectin receptor protein kinases*): É a segunda maior classe de motivos extracelulares e seus domínios extracelulares contém o motivo lectina que podem ligar vários dissacarídeos e açúcares complexos. Estudos recentes demonstram que o receptor *LecRK-V.5* regula negativamente o fechamento dos estômatos em tomate na percepção de *P. syringae*. A perda da função *LecRK-V.5* promove um aumento da resistência a esse patógeno (Desclos-Theveniau, *et al.*, 2012).

3. LysM (*Lysin Motif Receptor-like Kinase*): Outro ligante de açúcar responsável por promover a ligação de peptídeoglicanos, componente da parede celular bacteriana. Também tem sido demonstrado essencial para a percepção do componente da parede celular de quitina de fungos, como por exemplo o receptor de *Arabidopsis* CERK1 (Petutschnig, *et al.*, 2010)

4. TLP-Ks (*Thaumatococcus-like protein Kinase*): resultam da fusão entre dois genes que codificam uma thaumatina e uma proteína-quinase. Eles têm sido propostos para atuar como receptores quinase (RLKs), onde o domínio thaumatina extracelular poderia reconhecer

patógenos e a quinase seria responsável pela sinalização citoplasmática. Esta hipótese foi reforçada pela demonstração de que a super expressão de TLP-K em *Arabidopsis thaliana* promove um atraso no aparecimento dos sintomas de doenças (Petre *et al.*, 2011), como é o caso de PR5K, uma proteína quinase com domínio extracelular semelhante a uma PR proteína e com base nos seus domínios estruturais, no entanto, pode ser um receptor de sinal para as moléculas patogênicas. O domínio extracelular PR5 pode reconhecer o mesmo ligando reconhecido pelas proteínas PR5 (Guo *et al.*, 2003).

O número de receptores de tipo RLP preditos a partir de sequências genômicas varia de acordo com as espécies de plantas estudadas. Uma vez que os receptores de tipo RLP falta um domínio catalítico citoplasmático, uma das perguntas intrigantes relativas a sinalização mediada pelo RLP é como a mensagem é transmitida a partir da matriz extracelular para o espaço intracelular. A explicação mais simples poderia ser semelhante ao proposto para CLV2 e TMM onde estes RLPs podem funcionar em combinação com receptores RLKs CLAVATA1 e ERECTA, respectivamente, transmitindo, assim, a mensagem (Tor, *et al.*, 2009). A *xylanase-inducing ethylene* (EIX) é um potente elicitor de resposta de defesas em tabaco (*Nicotiana tabacum*) e tomate (*Solanum lycopersicum*) (Bar *et al.*, 2010). Os receptores EIX (LeEix1 e LeEix2) pertencem a essa superclasse de proteínas ricas em leucina (Bar *et al.*, 2010). Estudos demonstram que o silenciamento de expressão dos genes LeEix impede a ligação do EIX à células de uma planta sensível a EIX e, portanto, inibem a resposta de hipersensibilidade (Ron e Avni, 2004).

Em legumes, uma proteína de β -glucano (GBP - *β -glucan-binding protein*) corresponde ao terceiro grupo de receptores PRR. Essa proteína é o sítio de ligação para uma heptaglucoside presente na parede do oomiceto *Phytophthora sojae* (Zipfel, 2008).

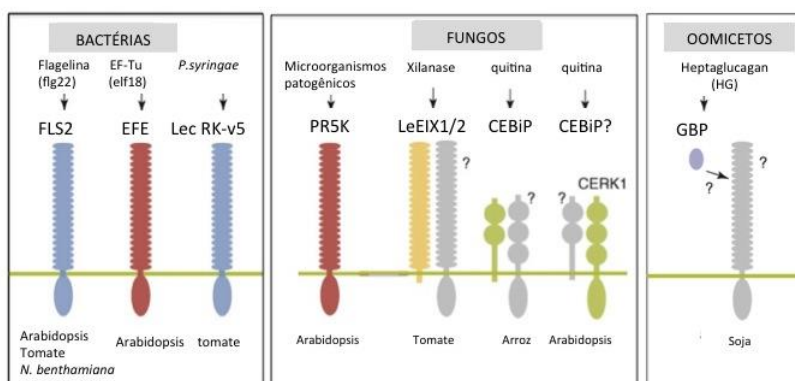


Figura 6: Principais PRRs identificados em plantas. Flagelina bacteriana (flg22) e EF-Tu (elf18) são reconhecidos pelos FLS2 e EFR (LRR-RLKs), respectivamente. Em tomate, xilanase é reconhecida pelo LPR LeEIX1 e LeEIX2. Embora ambos LeEIX1 e LeEIX2 pode ligar-se a EIX, apenas LeEIX2 é capaz de desencadear a sinalização. Como LeEIX1, CEBiP e GBP faltam domínios de sinalização óbvios, acredita-se que os mesmos interagem com proteínas transmembrana ainda desconhecidas. Fonte: Adaptado de Zipfel 2008.

Um exemplo clássico ocorre com a flagelina (flg22). Esta é a principal proteína estrutural de flagelos bacterianos e é essencial para a motilidade bacteriana, sendo também um fator de virulência necessário para a comprovada patogenicidade bacteriana (Ramos *et al.*, 2004). Essa flagelina é reconhecida por receptores PRR localizados na membrana plasmática da planta, conhecidos por FLS2 (LRR- RLK – *Leucine rich repeat/Receptor like kinase*). A percepção de flg22 por FLS2, ativa no citoplasma uma via de sinais MAP quinases e provavelmente os fatores de transcrição WRKY, provocando uma explosão oxidativa (ROS), alterações hormonais e reprogramação transcricional (Zipfel *et al.*, 2006 e Tao *et al.*, 2003).

A ativação de MAP quinases ocorre nos momentos iniciais da interação do PAMP com o receptor e é bastante transitória. Algumas quinases, como por exemplo MPK6/3 também são ativadas quando o patógeno carrega um efetor de avirulência (ETI), porém nesses casos essa ativação é mais prolongada quando comparada com a ativação realizada por um PAMP. O tempo de duração das atividades das quinases pode ser a chave para diferenciar as respostas em PTI e ETI (Tsuda e Katagiri, 2010).

A produção de ROS (*Reactive Oxygen Species*) é uma das primeiras respostas celulares após o reconhecimento do patógeno. ROS pode funcionar como moléculas de sinalização ou executor de patógenos. O reconhecimento de um PAMP desencadeia em PTI uma resposta rápida e transiente, pois existe uma relação sinérgica entre os setores de sinalização. A resposta PTI pode ser desencadeada por PAMPs benéficos ou patogênicos e essa baixa especificidade no reconhecimento do patógeno acaba sendo vantajoso para que as plantas não induzam fortes respostas imunológicas em estágios iniciais do reconhecimento (Tsuda e Katagiri, 2010).

5.2. As proteínas R e a imunidade disparada por efetores – ETI

Para neutralizar PTI e estabelecer uma infecção robusta em hospedeiros suscetíveis, os agentes patogênicos implantam proteínas efetoras (fatores de virulência) na célula hospedeira, bloqueando a cascata de MAPK por meio da segmentação dos receptores PRR RLK e BAK1, permitindo a esses patógenos superar PTI.

Por outro lado, esses efetores podem ser reconhecidos por específicos receptores de proteína de resistência (R) disparando ETI, uma versão mais forte e rápida de PTI, produzindo espécies reativas de oxigênio (ROS) que culmina, geralmente em uma reação de hipersensibilidade (HR - *Hypersensitive Response*) e morte celular programada (PCD -

Programmed Cell Death) normalmente local, não se estendendo para além da célula. Essa reação pode retardar o crescimento do patógeno em algumas interações, particularmente aquelas envolvendo parasitas haustoriais (Jones e Dangl, 2006). Obviamente a resposta envolvendo HR favorece o crescimento do patógeno, quando este possui um modo de vida necrotrófico, onde a sua estratégia de virulência depende da sua capacidade de matar células hospedeiras. Dessa forma, acredita-se que HR seja tipicamente ativa apenas contra patógenos biotróficos e hemibiotróficos (Glazebrook, 2005).

Após esses eventos de sinalização inicial, as plantas respondem pela ativação de um grande número de respostas de defesa integradas para repelir o invasor, como por exemplo, o fortalecimento da parede celular por meio da deposição de calose e lignina, produção de metabólitos secundários antimicrobianos, como fitoalexinas e acúmulo de proteínas PR (*pathogenesis-related*) como as quitinases e glucanases que degradam as paredes celulares de fungos e oomicetos (Pieterse *et al.*, 2009).

Entre outros fatores, a ativação das respostas de defesa envolve uma regulação do equilíbrio de fitormônios como o ácido salicílico (SA - *salicylic acid*) etileno (ET) e ácido jasmônico (JA - *jasmonic acid*) (Zvereva e Pooggin, 2012). Espécies transgênicas e mutantes com deficiência na biossíntese hormonal apresentaram uma grande alteração no nível de resistência a patógenos específicos, demonstrando a importância da regulação hormonal no desencadeamento de respostas de defesa. Estudos deixam evidente que patógenos biotróficos e hemibiotróficos são geralmente sensíveis à respostas de defesa que são reguladas por SA, enquanto patógenos com o modo de vida necrotrófico são comumente impedidos pelas defesas controladas por JA e ET (Glazebrook, 2005 e Deller *et al.*, 2011).

A resposta de defesa sistêmica em outros locais da planta é ativada pelos mecanismos de defesa elicitados no local da infecção para proteger tecidos saudáveis contra o patógeno. Esta resistência de longa duração e amplo espectro é conhecida como resistência sistêmica adquirida ou SAR (*Systemic acquired resistance*) e exige o acúmulo endógeno de SA, resultando na reprogramação transcricional de uma bateria de genes que codificam proteínas PR. O SA produzido no local da infecção se move célula a célula sob a forma de metil-SA através do plasmodesmata ou através do floema para o resto da planta (Muthamilarasan e Prasad, 2013). A concentração de SA regula a conversão da proteína NPR1 oligomérica para a forma monomérica levando à sua translocação nuclear e induzindo a expressão de proteínas PR. A homeostase dessa proteína é controlada pela ligação de SA a NPR3/NPR4 dependendo da concentração de SA. Em níveis baixos (alta suscetibilidade da planta) NPR1 está indisponível para induzir genes de defesa, uma vez que é direcionado, através de sua ligação

com NPR4 para a degradação no proteossoma. Com o aumento da concentração (resistência basal), SA se liga a NPR4 interrompendo a interação com NPR1 agora livre para desempenhar seu papel na ativação de genes de defesa (Moreau *et al.*, 2012).

5.2.1. Famílias de proteínas R

Inúmeros genes de resistência (R) de plantas já foram caracterizados e estão sendo utilizados de forma eficiente em programas de melhoramento (Figura 7). O uso destes genes para o desenvolvimento de variedades resistentes a doenças é uma alternativa conveniente à outras medidas, como pesticidas ou outros métodos de controle químico empregado para proteger as culturas de doenças. Diante da pressão seletiva exercida pelos patógenos, as plantas desenvolveram mecanismos de resistência pós invasão, muitas vezes controlados por genes dominantes de resistência, cujos produtos direta ou indiretamente detectam efetores patogênicos específicos e desencadeiam respostas de defesa eficazes (Chisholm *et al.*, 2006).

Até o momento, pelo menos oito classes de genes R são reconhecidas com base na organização de motivos de aminoácidos conservados e seus domínios de membrana (Gururani *et al.*, 2012). A classe mais numerosa é representada por membros de uma família gênica que codifica para proteínas contendo um sítio de ligação de nucleotídeo (NBS - *nucleotide-binding site*) e uma região rica em repetições de leucina (LRR - *leucine-rich repeats*). O domínio NBS está envolvido em sinalização e incluem motivos altamente conservados e estritamente ordenados, tais como P-loop, kinase-2 e Gly-Leu-Pro-Leu (GLPL) (Marone *et al.*, 2013). O domínio LRR, presente em muitas proteínas de funções diversas, está envolvido nas interações proteína-proteína. Este domínio interage diretamente com o seu efetor cognato, controlando o reconhecimento direto ou indireto de raças de patógenos específicos. Muitos LRRs podem interagir com outras proteínas transmembranares proporcionando também a capacidade de sinalização intercelular (Hammond-Kosack e Kanyuka, 2001). Análises mutacionais extensivas de algumas proteínas R identificaram vários resíduos essenciais em LRR, mas também em geral, um elevado grau de tolerância para substituições. Esta observação é consistente com o papel de LRR no reconhecimento, propriedade esta importante para a evolução de novas especificidades (Martin *et al.*, 2003).

As análises genômicas comparativas têm indicado que os genomas vegetais podem codificar várias centenas de genes NBS-LRR, e que existe uma grande diversidade no número e distribuição das subclasses desses genes. Até o presente momento, um grande número de sequências que codificam para NBS foram isolados a partir de várias espécies de

plantas por meio de análises de todo o genoma, encontrando cerca de 149 em *Arabidopsis thaliana* (Meyers *et al.*, 2003), 653 em *Oryza sativa* (Shang *et al.*, 2009), 459 em *Vitis vinifera* (Yang *et al.*, 2008), 319 em *Glycine max* (Kang *et al.*, 2012) e 117 em *M. acuminata* (D'Hont *et al.*, 2012). A distribuição cromossômica de genes NBS-LRR parece ser muito irregular na maioria das espécies estudadas, com alguns cromossomos caracterizados por conter muito mais genes NBS-LRR do que outros. Com batata, por exemplo, o maior número de genes de NBS-LRR são encontradas nos cromossomos 4 e 11, cerca de 15% dos genes mapeados, com o menor número de genes presentes no cromossomo 3 (1%) (Jupe *et al.*, 2012), enquanto que em banana o maior número concentra nos cromossomos 3 e 10, cerca de 17,1% dos genes NBS-LRR mapeados (D'Hont *et al.*, 2012). O fato de ser uma família gênica antiga, mecanismos como duplicação, cruzamento desigual, recombinação ectópica e conversão gênica têm sido propostos para ter contribuído para as estruturas de *cluster* de genes R e à evolução das especificidades de resistência (Marone *et al.*, 2013). Acredita-se que os receptores pertencentes a esta família foram desenvolvidos antes da separação do ramo das briófitas para plantas com sementes. Esta visão é corroborada pela constatação de que a família de genes PPC em musgos codificam proteínas TIR-NBS-LRR de outras plantas (Głowacki *et al.*, 2011).

As famílias de genes R podem ser classificadas de acordo com as suas características estruturais (Figura 7):

I - TIR-NBS-LRR (TNL) - Contêm um domínio de tipo Toll. São frequentes em dicotiledôneas e normalmente em maior número que os genes CNLs, sendo quase totalmente ausentes em monocotiledôneas. Os genes L6 e M que conferem resistência a *Melampsora lini* (Lawrence *et al.*, 1995) e N que confere resistência ao vírus mosaico (Lawrence *et al.*, 1995) são exemplos de genes R em tabaco. Os genes RPP1, RPP4 e RPP5 que conferem resistência a *Peronospora parasitica* (Botella *et al.*, 1998; van der Biezen *et al.*, 2002; Parker *et al.*, 1997) e RPS4 que confere resistência a *Pseudomonas syringae* (Gassmann *et al.*, 1999) são exemplos de genes TNL em *Arabidopsis*.

II - CC-NBS-LRR (CNL) - São proteínas que não codificam o domínio TIR e são caracterizadas por um domínio de filamento helicoidal. Vários exemplos desta classe são encontrados em *Arabidopsis*, tais como RPS2, RPS5 e RPM1 conferindo resistência a *Pseudomonas syringae* (Bent *et al.*, 1994; Warren *et al.*, 1998; Grant *et al.*, 1995) e os genes RPP8 e RPP13 que conferem resistência a *Peronospora parasitica* (McDowell *et al.*, 1998; Bittner-Eddy *et al.*, 2000). Em tomate, o gene I2 confere resistência a *F. oxysporum* (Ori *et al.*, 1997) e o gene Mi confere resistência a *Meloidogyne incógnita* (Milligan *et al.*, 1998) e

Macrosiphum euphorbiae (Rossi *et al.*, 1998). Os genes Bs2 de pimenta confere resistência a *X. campestris* (Tai *et al.*, 1999), Dm3 de alface confere resistência a *Bremia lactucae* (Meyers *et al.*, 1998), Gpa2 de batata confere resistência a *Globodera pallida* (van der Vossen, *et al.*, 2000), Hero de batata confere resistência a *Globodera rostochiensis* (Ernst *et al.*, 2002), R1 de batata confere resistência a *Phytophthora infestans* (Ballvora *et al.*, 2002), Mla de cevada confere resistência a *Blumeria graminis* (Zhou *et al.*, 2000), Pib e Pi-ta de arroz são responsáveis pela resistência a *Magnaporthe grisea* (Wang *et al.*, 1998; Orbach *et al.*, 2000) e Xa1 de arroz conferindo resistência a *Xanthomonas oryzae* (Yoshimura *et al.*, 1998) são exemplos de genes CNL já identificados.

III - LRR-TrD - É uma classe desprovida do domínio NBS consistindo em repetições de leucina extracelular eLRR ligada a um domínio transmembrana TrD. Os genes de tomate Cf-2 (Lozano-Torres *et al.*, 2012), Cf-4 (Thomas *et al.*, 1997), Cf-5 (Dixon *et al.*, 1998), Cf-6 (Grushetskaya *et al.*, 2007) e Cf-9 (Truong *et al.*, 2011) conferem resistência a *Cladosporium fulvum* e são os representantes já caracterizados nesta classe.

IV - LRR-TrD-kin - Esta classe é constituída de um domínio LRR extracelular ligado a um domínio transmembrana TrD e uma serina/treonina quinase como domínio intracelular. O principal representante desta classe é o gene Xa21 que confere resistência a *Xanthomonas oryzae* em arroz (Song *et al.*, 1997).

V - LRR-PEST-ECS - Esta classe possui um LRR extracelular juntamente com um domínio para degradação de proteínas PEST (Glu-Ser-Tre Pro) e motivos curtos ECS envolvidos na endocitose. As proteínas Ve1 contém um domínio CC, mas não possui o domínio PEST no C-terminal, enquanto que as proteínas Ve2 não possui o CC no N-terminal, mas possui o domínio PEST no C-terminal. Estas proteínas são encontradas em tomate e conferem resistência a *Verticillium albo-atrum* (Kawchuk *et al.*, 2001) e mais recentemente foram reconhecidos como receptores de PAMPs (Gururani *et al.*, 2012)

VI - TrD-CC - Constituído por um domínio de membrana TrD e um domínio CC intracelular. Seu principal representante é o gene RPW8 de Arabidopsis que confere resistência a *Erysiphe chicoracearum* (Xiao *et al.*, 2001).

VII – TIR-NBS-LRR-NLS-WRKY - Membro da classe I, porém com uma extensão C-terminal NLS (sinal de localização nuclear) e um domínio WRKY, região de aproximadamente 60 aminoácidos WRKYGQK bem conservados na extremidade N-terminal. Seu principal representante é gene de Arabidopsis RRS1-R que confere resistência a *Ralstonia solanacearum* (Deslandes *et al.*, 2002).

VIII – Kinase/Hm1 - Esta classe compreende um grupo de genes R enzimáticos que não possuem os domínios NBS-LRR. Como representantes encontram-se os genes Hm1 de milho que confere resistência ao fungo patogênico *Cochliobolus carbonum*. Ao contrário dos outros genes, Hm1 codifica uma enzima HC (toxina redutase) envolvida na detoxificação (Johal e Briggs, 1992). Outros exemplos notáveis são as proteínas Pto do tomate conferindo resistência a *P. syringae* por meio da codificação de uma serina/treonina quinase (Martin *et al.*, 1993) e Rpg1 de cevada conferindo resistência a ferrugem do caule codificando uma proteína quinase em *tandem* (quinase-quinase) (Brueggeman *et al.*, 2002).

Outros genes R com estruturas diferentes e que não estão inclusos nas oito classes acima também foram caracterizados. São eles: Xa27 de arroz conferindo resistência a *Xanthomonas oryzae* que consiste de 113 aminoácidos em dois domínios alfa-hélice (Gu *et al.*, 2005), RF01 de Arabidopsis conferindo resistência a *F. oxysporum* codificando uma quinase associada a parede celular, um domínio transmembrana (TrD) e uma quinase intracelular (Diener e Ausubel, 2005) e Pi-d2 de arroz conferindo resistência a *Magnaporthe grisea* codificando um receptor b-lectina ligado ao domínio TrD e uma quinase intracelular (Chen *et al.*, 2006).

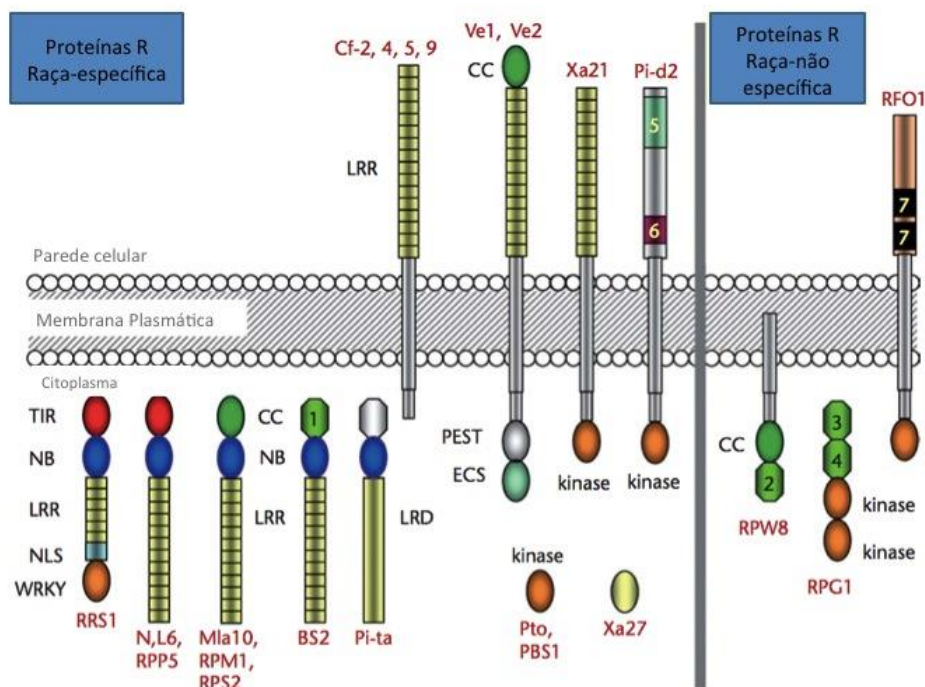


Figura 7: Esquema da localização dos genes R de diferentes classes que conferem resistência raça específica e raça não específica. Fonte: Adaptado de Hammond-Kosack e Kanyuka, 2001

JUSTIFICATIVA

O uso de fungicidas agrícolas em culturas de *M. acuminata* cv. Cavendish no controle da Sigatoka negra caminha em direção ao topo da lista de maiores insumos globais, com cerca de 500 milhões de dólares gastos anualmente. Em 2003, o custo anual de controle da doença em grandes plantações foi estimado em cerca de 1000 dólares por hectare. Mais de 30% dos custos totais de produção de banana são direcionados para o controle químico de *Mycosphaerella spp.* (Kema, 2006).

Considerando que a maioria dos estresses bióticos que afetam as bananeiras vêm sendo tratados com a aplicação de produtos químicos, encarecendo a produção e fomentando à adaptação dos organismos patogênicos, além da própria degradação ambiental que o uso de agrotóxicos enseja, o desenvolvimento de plantas resistentes às doenças seja uma medida de controle mais recomendável.

A busca de genes envolvidos em resposta aos estresses bióticos em *M. acuminata* durante a interação com *M. musicola*, bem como o desenvolvimento de marcadores moleculares, poderá oferecer uma potencial ferramenta para o desenvolvimento de uma estratégia eficaz de manejo da doença Sigatoka amarela com base em melhoramento genético da cultura por meio de transformação de plantas ou melhoramento convencional baseado em seleção assistida por marcadores, além de poder contribuir para a elucidação dos mecanismos de resistência e defesa em *Musa*.

OBJETIVOS

1. Objetivos Gerais

Identificar genes envolvidos em respostas a estresses bióticos derivados do transcrito foliar da interação *M. acuminata* e *M. musicola*.

Analisar a expressão de genes envolvidos em respostas aos estresses bióticos em genótipos de *M. acuminata* contrastantes em resistência, ‘Calcutta 4’ e ‘Cavendish Grande Naine’ submetidos à interação com *M. musicola*.

Desenvolver marcadores moleculares microssatélites presentes em genes potencialmente envolvidos em respostas de resistência e defesa em *M. acuminata* que possam ser utilizados em programas de melhoramento

2. Objetivos Específicos

- Identificar genes potencialmente envolvidos em respostas de resistência e defesa por meio de pirosequenciamento 454 do transcrito foliar da interação entre *M. musicola* e *M. acuminata* Calcutta 4 e Cavendish Grande Naine;

- Identificar e caracterizar o padrão de expressão *in silico* de genes-candidatos de defesa/resistência a partir de dados de pirosequenciamento 454 do transcrito foliar de *M. acuminata* Calcutta 4 (resistente) e Cavendish Grande Naine (suscetível) durante a interação com *M. musicola*;

- Analisar por RT-qPCR a expressão diferencial de genes-candidatos envolvidos em defesa/resistencia durante a interação *M. acuminata*- *M. musicola*

- Identificar em dados genômicos e transcritômicos de *M. acuminata* marcadores moleculares do tipo SSR que possam ser utilizados em programas de melhoramento;

- Validar os *primers* para locos SSR quanto a detecção de polimorfismos em diplóides de *M. acuminata* contrastando em resistência a doença Sigatoka

CAPÍTULO 1: ANÁLISE DE DADOS DO TRANSCRITOMA DA INTERAÇÃO ENTRE *MUSA ACUMINATA* E *MYCOSPHAERELLA MUSICOLA* PARA GENES ENVOLVIDOS NO PROCESSO DE IMUNIDADE

1. Introdução

Plantas evoluíram estratégias únicas de adaptação para lidar com estresses bióticos e abióticos devido à sua incapacidade de escapar da predação ou condições ambientais adversas. A resposta da planta a esses estresses é mediada por uma rede complexa de transdução de sinal que resulta em alterações na bioquímica, fisiologia e morfologia da planta. Essas respostas adaptativas estão associados a mudanças distintas na expressão gênica, mediada pela ação de fatores de transcrição. (Wang *et al.*, 2014). Portanto, são necessários extensos dados transcritômicos para descobrir genes relacionados com resistência à doenças. Esses dados podem servir também como uma boa fonte para o desenvolvimento de *primers* específicos para estes genes, para melhor caracterização de perfis de expressão gênica via PCR em tempo real durante a interação de banana / *M. musicola*.

A análise do transcriptoma é uma ferramenta poderosa para o estudo da função dos genes R na resistência às doenças. Recentemente, a disponibilidade de sequenciadores de última geração, como o GS- FLX 454 *Titanium* e Illumina-GA, tem permitido a identificação desses genes, bem como a avaliação dos níveis de expressão de genes preditos em uma determinada planta (Kang *et al.*, 2012).

Os objetivos deste capítulo foram gerar um recurso para o estudo de genes funcionais em *M. acuminata*, por meio da análise e identificação de genes expressos potencialmente envolvidos em processos de resistência e defesa. Foi realizado o sequenciamento *GS-FLX 454 Titanium* de transcritos derivados da interação entre *M. musicola*, patógeno causador da doença Sigatoka amarela em folhas de bananeiras e os genótipos contrastantes em resistência *M. acuminata* Calcutta 4 (resistente) e Cavendish Grande Naine (suscetível). Tais genótipos foram escolhidos com base em suas características relevantes. Calcutta 4 é um diplóide fértil selvagem amplamente utilizado em programas de melhoramento para a melhoria de cultivares comerciais. Como uma espécie doadora, considera-se uma fonte importante de resistência a fungos e nematóides patogênicos. Dada esta importância, tem sido adotado como um modelo para genômica comparativa com arroz (Cheung e Town, 2007; Lescot *et al.*, 2008), com aplicações de genômica funcional (Santos *et al.*, 2005) e descoberta de genes candidatos a resistência (Azhar e Heslop-Harrison, 2008; Miller *et al.*, 2008). Bananas do

subgrupo Cavendish, tais como Grande Naine, por contraste, são triplóides estéreis, que, apesar de representar mais de 40% da produção mundial, não têm a resistência ao estresse biótico, sendo que a aplicação regular de pesticidas é necessária para a produção comercial.

Os genes identificados neste transcrito foram mapeados no genoma de *M. acuminata* ssp. *malaccensis* var. Pahang para confirmação da montagem do sequenciamento *de novo* e foram utilizados como fonte de estudos de expressão e polimorfismos nos capítulos subsequentes.

2. Material e Métodos

2.1. Material Biológico

2.1.1. Material vegetal

Dois genótipos, Calcutta 4 e Cavendish Grande Naine foram selecionados levando em consideração a utilização comercial e resistência à pragas.

Calcutta 4 (AA) é um genótipo derivado de *M. acuminata* ssp. *Burmannica* que tem sido amplamente utilizado em programas de melhoramento por ser resistente às Sigatocas-negra e amarela, às diversas raças de *Fusarium* e nematóides (Fouré, 2000; Mobambo, 1994; Pegg, 1999; Forgain, 1996).

Cavendish Grande Naine (AAA) é o grupo de bananas mais cultivado mundialmente. Em 2010, estimava-se que 40% das bananas produzidas pertenciam a este genótipo e que 96% das bananas exportadas no comércio internacional eram bananas Cavendish (FruiTrop, 2012).

Um total de 36 plantas de seis meses de idade de *M. acuminata* Calcutta 4 e Cavendish Grande Naine (*Musa International Transit Centre* acessos ITC0249 e ITC0654) foram mantidas em estufa sob fotoperíodo de 12 h de luz e 12 h no escuro a 25 °C e 85% de umidade relativa. Plantas infectadas e controle não infectadas foram mantidas sob condições idênticas.

2.1.2. Fitopatógeno

A cepa de *M. musicola*, isolada de esporodóquios de lesões foliares de Cavendish Grande Naine pertencentes à Embrapa Mandioca e Frutas Tropicais, Brasil, foi utilizada para inóculo artificial da superfície abaxial da folha mais jovem de cada planta.

2.2. Preparação do inóculo

Os esporos de *M. musicola* foram anteriormente germinados em meio V8 por 14 dias. A contagem de esporos utilizados no inóculo totalizou 2×10^4 conidiósporos por mL, com adição de Tween 20 a 0.05%.

Trinta e seis mudas foram submetidas ao processo de inóculo, sendo 18 pertencentes ao genótipo Calcutta 4 (resistente) e 18 pertencentes ao genótipo Cavendish (suscetível). Dessas 18, nove mudas foram inoculadas com o patógeno na face abaxial das folhas 1 e 2 e nove plantas controle foram borrifadas apenas com solução de água destilada e tween com nas folhas 1 e 2. As folhas pulverizadas foram cobertas com sacos plásticos transparentes para garantir a alta umidade e três repetições independentes foram coletadas para cada amostra.

2.3. Microscopia Eletrônica de Varredura

Para a confirmar a presença do patógeno durante a infecção na superfície das folhas, foi realizada a técnica de Microscopia Eletrônica de Varredura (MEV)_ Zeiss DSM 962, gerando imagens tridimensionais de alta ampliação e resolução.

Um total de três repetições para cada amostra (três, seis e nove dias após o inóculo, período conhecido da entrada do patógeno) foram preparadas para a análise (Tabela 3). Para a preservação da estrutura biológica, as amostras foram primeiramente fixadas por meio da incubação durante 2,5 horas em solução fixadora de tampão cacodilato 0,05M a pH 6,8 contendo 2,5% de glutaraldeído e lavadas em tampão cacodilato 0,1M a pH 6,8. As amostras foram ainda pós-fixadas durante 1 hora em tampão tetróxido de ósmio 2%. A desidratação das amostras foi conduzida a 4 °C durante 20 minutos com concentrações crescentes de etanol (10, 20, 30, 50, 70, 80, 90, 95 e 100%). Após a secagem das amostras (Emitech K850, Kent, UK), as mesmas foram montadas em bases de cobre revestidas com 20 nm de partículas de ouro.

Tabela 3: Representação dos dias de coleta de plantas de *M. acuminata* Calcutta 4 e Cavendish Grande Naine inoculadas e não inoculadas (controle) com o patógeno *M. musicola*.

C4 não inoculada (controle)			C4 não inoculada (controle)			C4 não inoculada (controle)		
Tempo 1 3 dias	planta 1	folha 1	Tempo 2 6 dias	planta 1	folha 1	Tempo 3 9 dias	planta 1	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 2	folha 1		planta 2	folha 1		planta 2	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 3	folha 1		planta 3	folha 1		planta 3	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
CAV não inoculada (controle)			CAV não inoculada (controle)			CAV não inoculada (controle)		
Tempo 1 3 dias	planta 1	folha 1	Tempo 2 6 dias	planta 1	folha 1	Tempo 3 9 dias	planta 1	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 2	folha 1		planta 2	folha 1		planta 2	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 3	folha 1		planta 3	folha 1		planta 3	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
raíz			raíz			raíz		
C4 inoculada			C4 inoculada			C4 inoculada		
Tempo 1 3 dias	planta 1	folha 1	Tempo 2 6 dias	planta 1	folha 1	Tempo 3 9 dias	planta 1	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 2	folha 1		planta 2	folha 1		planta 2	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 3	folha 1		planta 3	folha 1		planta 3	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
CAV inoculada			CAV inoculada			CAV inoculada		
Tempo 1 3 dias	planta 1	folha 1	Tempo 2 6 dias	planta 1	folha 1	Tempo 3 9 dias	planta 1	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 2	folha 1		planta 2	folha 1		planta 2	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz
	planta 3	folha 1		planta 3	folha 1		planta 3	folha 1
		folha 2			folha 2			folha 2
		raíz			raíz			raíz

Legenda: C4: Calcutta 4; CAV: Cavendish

2.4. Extração de RNA

O RNA total foi extraído a partir de 1 g de amostras de folhas usando o reagente ConcertW RNA Plant Reagent (Invitrogen, Carlsbad, CA, USA) e INVISORB Spin Plant RNA Mini Kit (Invitex, Hayward, CA, USA), de acordo com as instruções do fabricante. Para a retirada de DNA das amostras, as mesmas foram tratadas com DNase (Ambion, Austin, TX, EUA) utilizando 1,5 unidades/ug de RNA total. A quantificação e integridade foi avaliada por eletroforese em gel de agarose 1% corado com brometo de etídio e espectrofotômetro Nanodrop ND-1000 (Thermo Scientific, Waltham, MA, EUA). Um total de 36 amostras de RNA foram extraídas a partir de amostras de plantas em triplicata para as folhas infectadas e as folhas controle não inoculadas, em três, seis e nove dias após o inóculo.

2.5. Sequenciamento massal e processamento de sequências

O RNA mensageiro isolado, foi preparado em bibliotecas cDNA consistindo de quatro *pools* derivados de Calcutta 4 inoculado, Calcutta 4 não inoculado, Cavendish inoculado e Cavendish não inoculado. O material foi mecanicamente fragmentado em sequências de 300 a 600 pb, ligados a adaptadores e submetidos a uma PCR emulsão e cada biblioteca foi sequenciada em um segmento de ¼ da placa, usando a plataforma baseada em pirosequenciamento 454 GS-FLX *and Titanium series chemistry*, realizada pela Eurofins MWG Operon (Ebersberg, Germany) a partir de aproximadamente 50 µg de cada *pool* de RNA total.

As leituras geradas a partir do sequenciamento foram pré-processadas anteriormente à análise para a remoção de adaptadores, leituras de baixa qualidade e pequenas leituras (< 50 pb) (Passos *et al.*, 2013), pois estas aparecem em grande número e repetidamente, levando a ambiguidades (Zerbino, 2008). Essas sequências foram montadas *de novo* devido a ausência de um genoma referência utilizando a estratégia de construção de grafos para geração de *K-mers*.

2.6. Mineração de dados 454 para genes envolvidos em defesa e resistência

Para as análises de dados derivados de sequenciamento de alto rendimento, o programa Neson versão 0.89 (Victorian Bioinformatics Consortium) foi utilizado para

estimar as frequências de DNA K-mers. (Hazelhurst, *et al.*, 2008) que foram montados pelo programa MIRA versão 3.1 (Chevreux *et al.*, 2007).

Os genes/proteínas envolvidos em processos de defesa e resistência a estresses bióticos foram anotados com base em regiões de similaridades encontradas em bancos de proteínas depositadas no NCBI por meio do algoritmo BLAST X (Altschul, 1997), com *e-value* de aceitação fixado em 10^{-5} . A análise funcional dessas proteínas foi feita a partir do programa InterProScan versão 4.8 (Zdobnov e Apweiler, 2001) identificando os domínios funcionais coincidentes com dados de famílias protéicas localizados em bancos de dados específicos. Para complementar com esta análise foi ainda utilizado o programa Blast2GO (Conesa, 2005) gerando não só anotações funcionais como também um significado biológico para os dados, como funções moleculares, processos biológicos e componentes celulares.

A fim de avaliar a precisão da montagem *de novo*, bem como a anotação, as sequências de *unigenes* foram mapeadas contra todos os modelos gênicos anotados no genoma completo de *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang) (<http://banana-genome.cirad.fr>) utilizando a ferramenta de mapeamento genômico e alinhamento denominada GMAP (Wu e Watanabe, 2005) com critérios de alinhamento composto por um mínimo de 90% da sequência de consulta mapeados no genoma e a percentagem de identidade de sequência em relação ao segmento mapeado superior a 95%.

3. Resultados e Discussão

O uso de tecidos foliares destacados em bioensaios com o patógeno *Mycosphaerella* tem sido relatado, porém os resultados obtidos são geralmente inconsistentes (por exemplo, [Arzanlou *et al.*, 2007; Donzelli e Churchill, 2007]), já que o desenvolvimento dos sintomas da doença nem sempre correlaciona-se com os sintomas normalmente observados em plantas intactas. Além disso, Arzanlou e colaboradores, (2007) sugeriram que o patógeno hemibiotrófico *Mycosphaerella* exige bananeiras saudáveis para o desenvolvimento da doença e Churchill (2011), sugeriu que o estado fisiológico de folhas destacadas não é comparável com o estado das folhas em plantas inteiras com um sistema radicular intacto. Expressão gênica em folhas do hospedeiro destacadas durante a interação com patógenos biotróficos também tem sido relatada em *Arabidopsis* (Liu *et al.*, 2007). Por estas razões, os bioensaios neste estudo foram conduzidos utilizando folhas jovens de plantas intactas com seis meses de idade, com as condições de temperatura e umidade ideal empregadas durante o experimento.

3.1. Microscopia Eletrônica de Varredura

Dada a caracterização ainda limitada da expressão gênica durante as interações entre banana e *Mycosphaerella* (Portal *et al.* 2011; Passos *et al.*, 2012), uma estratégia para o potencial enriquecimento de transcritos com genes envolvidos em respostas de defesa foi empregado. Dado que as respostas de defesa normalmente ocorrem mais cedo em interações incompatíveis, em relação às interações compatíveis, e que durante a fase inicial da infecção biotrófica de bananeira com *M. fijiensis*, tubos germinativos penetram em estômatos de três a seis dias após o inóculo (DAI) (Bévéraggi *et al.*, 1995), neste estudo, após uma investigação do patossistema com microscópio eletrônico de varredura, um curso temporal semelhante foi empregado para o preparo de bibliotecas de cDNA derivadas do patossistema *M. acuminata*-*M. musicola*.

De acordo com as imagens captadas pelo microscópio eletrônico de varredura, foi possível visualizar a presença do patógeno próximo aos estômatos foliares no intervalo de tempo proposto. Os resultados mostraram que os tubos germinativos e hifas eram visíveis aos três dias após o inóculo (DAI), com o crescimento de hifas sobre as células estomatais nos dias subsequentes em Cavendish Grande Naine em maior proporção que em Calcutta 4. (Figura 8).

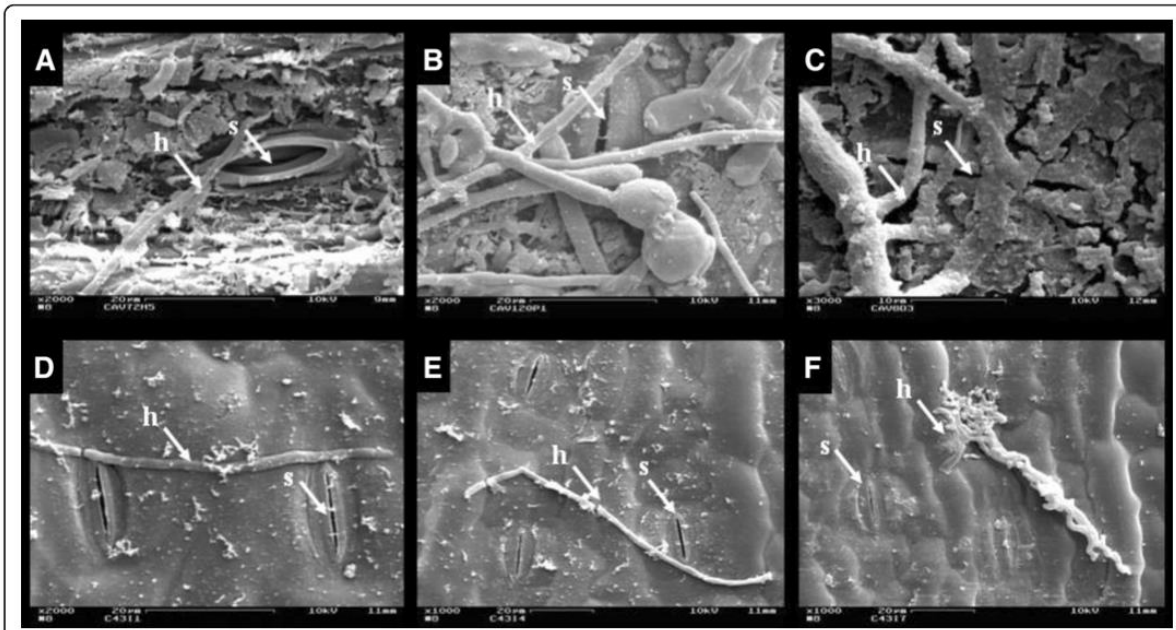


Figura 8: Observação da interação de *M. musicola* com a superfície interior das folhas dos genótipos de *M. acuminata* por meio de microscopia eletrônica de varredura (MEV). Tubos germinativos fúngicos e crescimento de hifas foi observado em 3, 6 e 9 DAI respectivamente, para Cavendish Grande Naine (A, B, C) e Calcutta 4 (D, E, F). Abreviações: DAI - dias após a inoculação; h - hifas; s- estômatos.

Constatação semelhante foi relatada por Liu *et al.*, 2002, que, investigando a interação *M. acuminata* – *M. fijiensis*, observou um aumento da penetração dos estômatos em Cavendish Grande Naine a partir de 5 DAI, com um total de 11% dos estômatos penetrados pelo patógeno até 21 DAI, e em contrapartida, para Calcutta 4, no mesmo período de tempo, apenas 0,95 % de estômatos foram infectados.

3.2. Qualidade do Material Vegetal

As amostras de RNA total foliar foram extraídas utilizando o reagente Concert™ demonstraram qualidade satisfatória. A concentração, bem como a qualidade das amostras foram avaliadas por meio da obtenção do espectro de absorvância utilizando um espectrofotômetro NanoDrop ND-1000 (NanoDrop Technologies). Os picos de absorvância apresentaram forma de sino, com razões 260/280nm e 260/230nm indicativos de RNA com integridade aceitável (Figura 10). Segundo Fleige e Pfaffl, 2006, uma relação OD 260/280 maior do que 1,8 é considerado um indicador de RNA de alta qualidade.

A precisão da técnica de PCR em tempo real, para avaliar a expressão de genes é reconhecida por ser influenciada pela quantidade e qualidade do RNA de partida. Pureza e integridade de RNA são elementos críticos para o sucesso global de análises baseadas em RNA.

3.3. Sequenciamento massal e processamento de sequências

A partir de quatro *pools* de RNA total, representando os transcritos de *M. acuminata* Calcutta 4 e Cavendish Grande Naine, na presença e ausência do patógeno *M. musicola*, bibliotecas de cDNA foram preparadas e pirosequenciadas (454) pela Eurofins MWG Operon (Ebersberg, Alemanha). Um total de 978.133 sequências (*reads*) foram geradas para os dois genótipos, totalizando mais de 466 Mb de dados. Este trabalho permitiu a caracterização global dos genes expressos nos dois genótipos de *M. acuminata* durante as reações compatíveis e incompatíveis com *M. musicola* e o processamento de sequências possibilitou a seleção de 846.762 sequências de alta qualidade (283Mb), organizadas em 24.259 agrupamentos ou *contigs* e 12.125 sequências não agrupadas ou *singetons* em Calcutta 4, e um padrão semelhante em Cavendish Grande Naine, com um total de 23.729 agrupamentos e 11.540 sequências não agrupadas (Passos *et al.*, 2013).

3.4. Mineração de dados de pirosequenciamento 454 para transcritos de genes potencialmente envolvidos em defesa e resistência em *M. acuminata*

A estratégia de mineração de genes candidatos envolvidos em respostas de defesa com base em termos GO relacionados com defesa e anotação Blast2GO, gerou *unigenes* potencialmente envolvidos na imunidade disparada por efetores (ETI) e na imunidade disparada por PAMPs (PTI) em ambos os genótipos (Passos *et al.*, 2013). Foi identificado um número significativo de *unigenes* potencialmente envolvidos em respostas de defesa em tecidos foliares de *M. acuminata* Calcutta 4 inoculados com o patógeno, incluindo dentre outros, componentes de síntese de glucano, RPM1, relatado na defesa de plantas envolvendo proteínas R, proteína mlo-like 1, conhecida por estar envolvida na defesa e no reforço da parede celular da planta e fatores de transcrição WRKY. Componentes envolvidos ROS também foram identificados em Calcutta 4 e Cavendish Grande Naine.

Com base nas anotações de genes expressos na interação *M. acuminata* e *M. musicola*, foram identificados por análise Blast de tecidos foliares infectados e não infectados, transcritos de 14 genes NBS-LRR em *M. acuminata* Calcutta 4 (Tabela 4) e 25 em *M. acuminata* Cavendish Grande Naine (Tabela 5). Apesar de não poder aferir a respeito da expressão diferencial de genes devido a falta de réplicas biológicas, é possível observar diferenças nos números de transcritos de *contigs* nos diferentes tratamentos (inoculado e controle). Em Calcutta 4, o *contig Musa_c4_small_rep_c6985*, cuja descrição Blast confere

com um gene NB-ARC, possui 31 cópias quando inoculado, em contrapartida, no controle existe apenas três cópias deste fragmento. Em Cavendish, o *contig Musa_cav_rep_c4136*, também NB-ARC, apresenta-se ausente no controle e 15 cópias quando o cultivar é inoculado.

Mapeamento dos *contigs* de *unigenes* aos modelos de genes que contêm o domínio NB-ARC no genoma de referência *M. acuminata* DH Pahang identificou 38 *contigs* de Calcutta 4 mapeando com 40 modelos de genes (Tabela 6) e 43 *contigs* derivados de Cavendish Grande Naine mapeando com 40 modelos de genes com o domínio NB-ARC em DH Pahang (Tabela 7). NBS-LRR é a classe mais abundante das famílias de genes R (Hammond-Kosack e Jones, 1997). Nas culturas de arroz, cerca de 653 NBS-LRR genes foram caracterizados (Shang *et al.*, 2009), 149 em *Arabidopsis thaliana* (Meyers *et al.*, 2003), 459 em *Vitis vinifera* (Yang *et al.*, 2008), 319 em *Glycine max* (Kang *et al.*, 2012) e 117 em *M. acuminata* (D'Hont *et al.*, 2012).

Análise Blast também previu inúmeros transcritos para outras classes de genes R conhecidas entre os dois genótipos como, por exemplo, a LRR extracelular ligada a um domínio transmembrana (LRR-TrD), LRR extracelular ligado a um domínio transmembrana TrD e uma serina/treonina quinase (LRR-TrD-kin), domínio de membrana TrD e um domínio CC intracelular (TrD-CC) e a classe Kinase.

Tabela 4: Análise BLASTX e contigs NBS-LRR derivados do transcrito foliar da interação *M. musicola* e *M. acuminata* Calcutta 4

Contigs	Número de reads C4NI	Número de reads C4I	Acesso Genbank	Descrição	E-value
<i>Musa_c4_small_rep_c6985</i>	3	31	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	7,94E-51
<i>Musa_c4_small_rep_c8909</i>	7	1	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	1,08E-21
<i>Musa_c4_small_c16788</i>	0	5	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	5,19E-16
<i>Musa_c4_small_c16314</i>	3	0	gi 166084741 emb CAP66367.1	NBS-LRR disease resistance protein [<i>M. balbisiana</i> x <i>Musa</i> <i>textilis</i>]	2,32E-93
<i>Musa_c4_small_c9165</i>	5	1	gi 301154126 emb CBW30231.1	Disease resistance protein (CC-NBS-LRR) [<i>M. balbisiana</i>]	5,92E-12
<i>Musa_c4_small_c15362</i>	0	3	gi 301154122 emb CBW30219.1	Disease resistance protein (CC-NBS-LRR) [<i>M. balbisiana</i>]	7,11E-13
<i>Musa_c4_small_c20045</i>	0	3	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	2,87E-14
<i>Musa_c4_small_c23692</i>	0	2	gi 217426417 gb ACK44404.1	NBS resistance protein [<i>Musa</i> ABB Group]	3,22E-45
<i>Musa_c4_small_c23853</i>	2	0	gi 224103271 ref XP_002334072.1	cc-nbs-lrr resistance protein [<i>Populus trichocarpa</i>]	7,63E-15
<i>Musa_c4_small_c23629</i>	3	1	gi 3411227 gb AAC31553.1	NBS-LRR type disease resistance protein O2 [<i>Avena sativa</i>]	1,23E-28
<i>Musa_c4_small_rep_c6968</i>	6	4	gi 224069146 ref XP_002302911.1	cc-nbs-lrr resistance protein [<i>Populus trichocarpa</i>]	2,28E-11
<i>Musa_c4_small_c18984</i>	2	1	gi 301154127 emb CBW30232.1	Disease resistance protein (CC-NBS-LRR) [<i>M. balbisiana</i>]	2,95E-06
<i>Musa_c4_small_c22826</i>	2	1	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	4,98E-57
<i>Musa_c4_small_c25375</i>	1	1	gi 166084569 emb CAP66281.1	NBS-LRR disease resistance protein [<i>M. acuminata</i>]	6,44E-71

C4: Calcutta 4; IN: inoculado; NI: Não inoculado (controle)

Tabela 5: Análise BLASTX de contigs NBS-LRR derivados do transcrito foliar da interação *M. musicola* e *M. acuminata* Cavendish Grande Naine

Contigs	Número de reads CAVNI	Número de reads CAVI	Acesso Genbank	Descrição	E-value
<i>Musa_cav_rep_c4136</i>	0	15	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	4,44E-023
<i>Musa_cav_c15974</i>	5	1	gi 4234953 gb AAD13036.1	NBS-LRR-like protein cD7 [<i>Phaseolus vulgaris</i>]	6,07E-006
<i>Musa_cav_c15914</i>	3	0	gi 166084569 emb CAP66281.1	NBS-LRR disease resistance protein [<i>M. acuminata</i>]	1,21E-055
<i>Musa_cav_c24470</i>	3	0	gi 301154130 emb CBW30237.1	Disease resistance protein (CC-NBS-LRR) [<i>M. balbisiana</i>]	8,93E-076
<i>Musa_cav_c14064</i>	4	1	gi 301154127 emb CBW30232.1	Disease resistance protein (CC-NBS-LRR) [<i>M. balbisiana</i>]	7,08E-082
<i>Musa_cav_c4853</i>	10	8	gi 224063557 ref XP_002301203.1	nbs-lrr resistance protein [<i>Populus trichocarpa</i>]	7,37E-116
<i>Musa_cav_c19018</i>	2	0	gi 26986180 emb CAD58967.1	Disease resistance protein NBS-LRR type [<i>M. acuminata</i>]	1,86E-037
<i>Musa_cav_c19720</i>	2	0	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	1,06E-008
<i>Musa_cav_c17152</i>	3	1	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	5,18E-047
<i>Musa_cav_rep_c9340</i>	0	5	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	6,03E-020

<i>Musa_cav_c13800</i>	0	4	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	2,71E-020
<i>Musa_cav_c13469</i>	1	7	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	1,41E-038
<i>Musa_cav_c16474</i>	0	3	gi 217337736 gb ACK43220.1	NBS1-like protein [Oryza sativa Indica Group]	2,90E-014
<i>Musa_cav_rep_c11257</i>	0	3	gi 41223413 gb AAR99708.1	NBS-LRR-like protein B [Oryza sativa Indica Group]	3,19E-013
<i>Musa_cav_c10767</i>	4	5	gi 301154130 emb CBW30237.1	Disease resistance protein (CC-NBS-LRR) [<i>M. balbisiana</i>]	3,60E-156
<i>Musa_cav_c13788</i>	3	4	gi 166084663 emb CAP66328.1	NBS-LRR disease resistance protein [<i>M. acuminata</i> AAA Group]	2,28E-075
<i>Musa_cav_c19806</i>	0	2	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	1,90E-013
<i>Musa_cav_c22011</i>	1	1	gi 166084784 emb CAP66390.1	truncated NBS-LRR disease resistance protein [<i>M. balbisiana</i>]	4,29E-021
<i>Musa_cav_c22092</i>	0	2	gi 166084784 emb CAP66390.1	truncated NBS-LRR disease resistance protein [<i>M. balbisiana</i>]	1,32E-023
<i>Musa_cav_rep_c15425</i>	0	2	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	3,46E-036
<i>Musa_cav_c15435</i>	1	2	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	1,50E-047
<i>Musa_cav_rep_c4136</i>	0	15	gi 193795177 gb ACF21695.1	NBS-type resistance protein RGC5 [<i>M. acuminata</i> subsp. malaccensis]	4,44E-023
<i>Musa_cav_c14416</i>	2	2	gi 224064832 ref XP_002301573.1	cc-nbs-lrr resistance protein [<i>Populus trichocarpa</i>]	1,20E-034
<i>Musa_cav_c17726</i>	0	2	gi 50080315 gb AAT69649.1	putative NBS-LRR type resistance protein [Oryza sativa Japonica Group]	1,32E-011
<i>Musa_cav_rep_c4917</i>	5	13	gi 224064832 ref XP_002301573.1	cc-nbs-lrr resistance protein [<i>Populus trichocarpa</i>]	4,92E-053

CAV: Cavendish Grande Naine; IN: inoculado; NI: Não inoculado (controle)

Tabela 6: Dados de mapeamento para contigs *Unigenes* de *M. acuminata* Calcutta 4 alinhados com todos os modelos de genes NB-ARC de *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang).

Modelos gênicos <i>M. acuminata</i> DH-Pahang	Anotação Gênica	Mapeamento com contigs <i>Unigenes</i> Calcutta 4
GSMUA_AchrUn_randomG28470_001	Putative disease resistance protein RGA4	<i>Musa_c4_clusters_pos_c560_1.path4#chrUn_random_138089599-138090468</i>
GSMUA_Achr9G11390_001	Putative Disease resistance protein RGA2	<i>Musa_c4_small_c10766.path1#chr9_7396812-7397100</i>
GSMUA_Achr6G32070_001	Putative Disease resistance protein RGA2	<i>Musa_c4_small_c11248.path1#chr6_31612273-31612735</i>
GSMUA_AchrUn_randomG25350_001	Putative NB-ARC domain containing protein%2C expressed	<i>Musa_c4_small_c11496.path1#chrUn_random_125202774-125203016</i> <i>Musa_c4_small_c11496.path2#chr3_23166837-23167079:Musa_c4_small_c21035.path1#chr3_23164050-23167079:Musa_c4_small_rep_c6985.path1#chr3_23164050-23167079:Musa_c4_small_rep_c8909.path1#chr3_23164050-23167079</i> <i>Musa_c4_small_c11588.path1#chr7_6157872-6159110:Musa_c4_small_c4633.path1#chr7_6159443-6159882</i>
GSMUA_Achr3G22250_001	Putative Disease resistance protein RGA2	<i>Musa_c4_small_c13356.path1#chr10_17441044-17441600</i>
GSMUA_Achr7G08250_001	Putative Probable disease resistance protein At4g33300	<i>Musa_c4_small_c14043.path1#chr9_1979883-1980151</i>
GSMUA_Achr10G07150_001	Putative disease resistance protein RGA3	<i>Musa_c4_small_c14244.path1#chr9_8303143-8303507</i>
GSMUA_Achr9G02800_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c14875.path1#chr5_8689085-8689397</i>
GSMUA_Achr9G12800_001	Putative Disease resistance protein RPM1	<i>Musa_c4_small_c15362.path1#chr9_28045686-28046018</i>
GSMUA_Achr5G11980_001	Putative Disease resistance protein RPS2	<i>Musa_c4_small_c16416.path1#chr9_1960080-1960410</i>
GSMUA_Achr9G23050_001	disease resistance RPP13-like protein 1%2C putative%2C expressed	
GSMUA_Achr9G02760_001	Putative disease resistance protein RGA3	

GSMUA_Achr6G21800_001	Putative disease resistance protein RGA3	<i>Musa_c4_small_c16788.path1#chr6_18426538-18426783</i>
GSMUA_Achr10G24910_001	Putative disease resistance RPP13-like protein 1	<i>Musa_c4_small_c17316.path1#chr10_29131049-29133579</i>
GSMUA_Achr4G26160_001	Putative disease resistance protein RGA3	<i>Musa_c4_small_c18165.path1#chr4_25546804-25547093</i>
GSMUA_Achr10G07050_001	NBS-LRR disease resistance protein%2C putative	<i>Musa_c4_small_c18312.path1#chr10_17253426-17253858</i>
GSMUA_Achr7G21400_001	disease resistance protein RGA2%2C putative%2C expressed	<i>Musa_c4_small_c18825.path1#chr7_24197305-24197708</i>
GSMUA_AchrUn_randomG27810_001	Putative Disease resistance protein RGA2	<i>Musa_c4_small_c18984.path1#chrUn_random_135436255-135436424</i>
GSMUA_Achr10G29330_001	Putative Disease resistance protein RGA2	<i>Musa_c4_small_c19292.path1#chr10_31890908-31890953</i>
GSMUA_AchrUn_randomG23630_001	Putative NBS-type resistance protein RGC5	<i>Musa_c4_small_c20045.path1#chrUn_random_115732179-115732421</i>
		<i>Musa_c4_small_c20045.path2#chr3_23171939-23172130:Musa_c4_small_c21035.path1#chr3_23170675-23173641:Musa_c4_small_rep_c6985.path1#chr3_23170675-23173641:Musa_c4_small_rep_c8909.path1#chr3_23170675-23173641</i>
GSMUA_Achr3G22260_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c20045.path3#chr3_23245017-23245255:Musa_c4_small_rep_c6985.path1#chr3_23242204-23242918</i>
GSMUA_Achr3G22310_001	Putative disease resistance RPP13-like protein 1	<i>Musa_c4_small_c20045.path5#chrUn_random_137169168-137169351</i>
GSMUA_AchrUn_randomG28220_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c21035.path1#chr3_23188166-23191165:Musa_c4_small_rep_c6985.path1#chr3_23188166-23191165:Musa_c4_small_rep_c8909.path1#chr3_23188166-23191165</i>
GSMUA_Achr3G22280_001	Putative NB-ARC domain containing protein	<i>Musa_c4_small_c21035.path1#chr3_23196686-23200666:Musa_c4_small_rep_c6985.path1#chr3_23196686-23200666:Musa_c4_small_rep_c8909.path1#chr3_23196686-23200666</i>
GSMUA_Achr3G22290_001	Putative disease resistance RPP13-like protein 1	<i>Musa_c4_small_c21635.path1#chr7_10298625-10298667</i>
GSMUA_Achr7G12770_001	Putative disease resistance protein RGA4	<i>Musa_c4_small_c22030.path1#chr10_29145724-29146081:Musa_c4_small_c24485.path1#chr10_29144970-29145128:Musa_c4_small_c25375.path1#chr10_29144450-29147763</i>
GSMUA_Achr10G24940_001	Putative Disease resistance protein RGA2	<i>Musa_c4_small_c22782.path1#chr9_11840905-11841427:Musa_c4_small_c23853.path1#chr9_11842919-11843223:Musa_c4_small_c4534.path1#chr9_11841615-11842320</i>
GSMUA_Achr9G17310_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c22826.path1#chr6_33189319-33189812</i>
GSMUA_Achr6G34690_001	Putative disease resistance protein RGA3	<i>Musa_c4_small_c23629.path1#chr4_27035193-27035797:Musa_c4_small_c9959.path1#chr4_27034238-27034651</i>
GSMUA_Achr4G28700_001	Putative Disease resistance protein RPS5	<i>Musa_c4_small_c24475.path1#chr6_18419197-18419300</i>
GSMUA_Achr6G21780_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c25174.path1#chrUn_random_137488847-137489866</i>
GSMUA_AchrUn_randomG28310_001	Putative disease resistance RPP13-like protein 1	<i>Musa_c4_small_c25375.path1#chr10_29137584-29138042</i>
GSMUA_Achr10G24920_001	Putative disease resistance protein RGA3	<i>Musa_c4_small_c4049.path1#chr9_7572070-7572160</i>
GSMUA_Achr9G11660_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c6893.path2#chr10_17271756-17272538</i>
GSMUA_Achr10G07060_001	Putative disease resistance protein RGA4	<i>Musa_c4_small_c6893.path2#chr10_17294401-17296785</i>
GSMUA_Achr10G07080_001	Putative disease resistance protein RGA1	<i>Musa_c4_small_c6893.path2#chr10_17309864-17310595</i>
GSMUA_Achr10G07090_001	Putative disease resistance protein RGA4	<i>Musa_c4_small_c6893.path2#chr10_17323106-17326579</i>
GSMUA_Achr10G07100_001	NBS-LRR disease resistance protein%2C putative	<i>Musa_c4_small_c8758.path1#chr9_32147509-32147878</i>
GSMUA_Achr9G28030_001	Putative Disease resistance protein RPM1	<i>Musa_c4_small_rep_c6985.path1#chr3_23233713-23238016:Musa_c4_small_rep_c8909.path1#chr3_23233713-23234703</i>
GSMUA_Achr3G22300_001	Putative disease resistance protein RGA1	

Tabela 7: Dados de mapeamento para contigs *Unigenes* de *M. acuminata* Cavendish Grande Naine alinhados com todos os modelos de genes NB-ARC de *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang).

Modelos gênicos <i>M. acuminata</i> DH-Pahang	Anotação Gênica	Mapeamento com contigs <i>Unigenes</i> Cavendish Grande Naine
GSMUA_Achr3G27790_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c10767.path1#chr3_27283945-27286600:Musa_cav_c14064.path4#chr3_27284501-27286600</i>
GSMUA_Achr3G27810_001	Putative disease resistance protein RGA1	<i>Musa_cav_c10767.path1#chr3_27296707-27297659:Musa_cav_c14064.path4#chr3_27296707-27299340</i>
GSMUA_AchrUn_randomG19290_001	Putative disease resistance protein RGA1	<i>Musa_cav_c10767.path2#chrUn_random_93055368-93055930:Musa_cav_c14064.path1#chrUn_random_93055275-93055768</i>
GSMUA_AchrUn_randomG19310_001	Putative disease resistance protein RGA1	<i>Musa_cav_c10767.path2#chrUn_random_93066947-93070042</i>
GSMUA_AchrUn_randomG19270_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c10767.path5#chrUn_random_93041129-93041223:Musa_cav_c24470.path1#chrUn_random_93049824-93050346</i>
GSMUA_Achr9G17060_001	Putative disease resistance protein RGA1	<i>Musa_cav_c11310.path1#chr9_11601589-11602017:Musa_cav_c14731.path1#chr9_11603298-11603743:Musa_cav_c18318.path1#chr9_11600969-11601322:Musa_cav_c18488.path1#chr9_11602143-11602558</i>
GSMUA_Achr9G28030_001	Putative Disease resistance protein RPM1	<i>Musa_cav_c12804.path1#chr9_32148260-32148678:Musa_cav_c5052.path1#chr9_32147509-32148631</i>
GSMUA_Achr3G22280_001	Putative NB-ARC domain containing protein	<i>Musa_cav_c13469.path2#chr3_23188166-23191165:Musa_cav_c13788.path1#chr3_23188166-23191165:Musa_cav_c17152.path1#chr3_23188166-23191165:Musa_cav_c19806.path2#chr3_23188166-23191165:Musa_cav_rep_c15425.path3#chr3_23188166-23189837:Musa_cav_rep_c4136.path2#chr3_23188166-23191165</i>
GSMUA_Achr3G22290_001	Putative disease resistance RPP13-like protein 1	<i>Musa_cav_c13469.path2#chr3_23196686-23197401:Musa_cav_c13788.path1#chr3_23196686-23200666:Musa_cav_c17152.path1#chr3_23196686-23200666:Musa_cav_c19806.path2#chr3_23196686-23200666:Musa_cav_c20011.path1#chr3_23200438-23200666:Musa_cav_rep_c4136.path2#chr3_23196686-23200666</i>
GSMUA_Achr3G22250_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c13788.path1#chr3_23165865-23167079:Musa_cav_c17152.path1#chr3_23164050-23167079:Musa_cav_c19720.path1#chr3_23164050-23167079:Musa_cav_c19806.path2#chr3_23164050-23167079:Musa_cav_rep_c4136.path1#chr3_23165389-23165654</i>
GSMUA_Achr3G22260_001	Putative disease resistance protein RGA1	<i>Musa_cav_c13788.path1#chr3_23170675-23173641:Musa_cav_c19720.path1#chr3_23170675-23173641:Musa_cav_c19806.path2#chr3_23170675-23170982:Musa_cav_c19806.path2#chr3_23170675-23173641:Musa_cav_rep_c15425.path3#chr3_23171806-23173641:Musa_cav_rep_c4136.path2#chr3_23171846-23173641</i>
GSMUA_Achr3G22300_001	Putative disease resistance protein RGA1	<i>Musa_cav_c13788.path1#chr3_23233713-23238016:Musa_cav_c19806.path2#chr3_23233713-23238016:Musa_cav_rep_c4136.path2#chr3_23233713-23237323</i>
GSMUA_Achr3G22310_001	Putative disease resistance RPP13-like protein 1	<i>Musa_cav_c13788.path1#chr3_23242204-23245785:Musa_cav_c19806.path2#chr3_23242204-23243934:Musa_cav_c20011.path1#chr3_23242204-23245785</i>
GSMUA_Achr3G22320_001	Putative disease resistance protein RGA1	<i>Musa_cav_c13788.path1#chr3_23289819-23291963:Musa_cav_c20011.path1#chr3_23289819-23292500</i>
GSMUA_AchrUn_randomG26110_001	Putative NB-ARC domain containing protein%2C expressed	<i>Musa_cav_c13788.path2#chrUn_random_128380841-128381512:Musa_cav_c20011.path3#chrUn_random_128381848-128382076</i>
GSMUA_AchrUn_randomG27050_001	NBS-LRR disease resistance protein%2C	<i>Musa_cav_c14064.path2#chrUn_random_132387329-132387822</i>

	putative%2C expressed	
GSMUA_AchrUn_randomG28310_001	Putative disease resistance RPP13-like protein 1	<i>Musa_cav_c14064.path3#chrUn_random_137489108-137489601</i> <i>Musa_cav_c14064.path4#chr3_27306203-27306898:Musa_cav_c24470.path2#chr3_27306498-27306884</i>
GSMUA_Achr3G27820_001	NBS-LRR class resistance protein (Fragment)	
GSMUA_Achr3G27830_001	Putative disease resistance protein RGA1	<i>Musa_cav_c14064.path4#chr3_27308495-27309312</i>
GSMUA_Achr9G28020_001	Putative Disease resistance protein RPM1	<i>Musa_cav_c14416.path1#chr9_32130387-32130991</i>
GSMUA_Achr7G21400_001	disease resistance protein RGA2%2C putative%2C expressed	<i>Musa_cav_c15359.path1#chr7_24190208-24190614</i>
GSMUA_AchrUn_randomG27820_001	Putative disease resistance RPP13-like protein 3	<i>Musa_cav_c15418.path4#chrUn_random_135467902-135469119</i>
GSMUA_Achr6G34690_001	Putative disease resistance protein RGA3	<i>Musa_cav_c15435.path1#chr6_33189421-33189752</i> <i>Musa_cav_c15914.path1#chr10_29133172-29133579:Musa_cav_c15974.path1#chr10_29131041-29133579</i>
GSMUA_Achr10G24910_001	Putative disease resistance RPP13-like protein 1	<i>Musa_cav_c15914.path1#chr10_29135802-29138042:Musa_cav_c15974.path1#chr10_29135802-29138042</i>
GSMUA_Achr10G24920_001	Putative disease resistance protein RGA3	<i>Musa_cav_c15914.path1#chr10_29144450-29147705:Musa_cav_c15974.path1#chr10_29144450-29145587:Musa_cav_c7673.path1#chr10_29145565-29146806</i>
GSMUA_Achr10G24940_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c16077.path1#chr4_27035140-27035607:Musa_cav_c19018.path1#chr4_27036391-27036770:Musa_cav_rep_c11306.path1#chr4_27034220-27034937</i>
GSMUA_Achr4G28700_001	Putative Disease resistance protein RPS5	
GSMUA_Achr7G08250_001	Putative Probable disease resistance protein At4g33300	<i>Musa_cav_c16960.path1#chr7_6156290-6157285:Musa_cav_c4853.path1#chr7_6158004-6159882</i> <i>Musa_cav_c17726.path1#chr10_31890253-31890641:Musa_cav_c7700.path1#chr10_31889674-31890053</i>
GSMUA_Achr10G29330_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c18642.path1#chr6_31612418-31612781</i>
GSMUA_Achr6G32070_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c19585.path2#chrUn_random_135436254-135438402:Musa_cav_rep_c23841.path1#chrUn_random_135436294-135436377</i>
GSMUA_AchrUn_randomG27810_001	Putative Disease resistance protein RGA2	<i>Musa_cav_c19720.path3#chrUn_random_137170302-137170612:Musa_cav_rep_c4136.path3#chrUn_random_137169185-137169453</i>
GSMUA_AchrUn_randomG28220_001	Putative disease resistance protein RGA1	
GSMUA_AchrUn_randomG25350_001	Putative NB-ARC domain containing protein%2C expressed	<i>Musa_cav_c20011.path2#chrUn_random_125202805-125203016</i>
GSMUA_Achr6G21780_001	Putative disease resistance protein RGA1	<i>Musa_cav_c20785.path1#chr6_18419267-18419300</i> <i>Musa_cav_c24501.path1#chr9_11839559-11839793:Musa_cav_c7213.path1#chr9_11840658-11841014</i>
GSMUA_Achr9G17310_001	Putative disease resistance protein RGA1	<i>Musa_cav_c4454.path1#chr9_1982164-1982516:Musa_cav_rep_c4227.path1#chr9_1982976-1983868:Musa_cav_rep_c6048.path1#chr9_1982934-1983614:Musa_cav_rep_c6048.path2#chr9_1982850-1982912</i>
GSMUA_Achr9G02800_001	Putative disease resistance protein RGA1	<i>Musa_cav_rep_c11257.path1#chr10_17294401-17294770</i>
GSMUA_Achr10G07080_001	Putative disease resistance protein RGA1	<i>Musa_cav_rep_c15425.path1#chrUn_random_115732143-115732536:Musa_cav_rep_c4136.path4#chrUn_random_115732240-115732508</i>
GSMUA_AchrUn_randomG23630_001	Putative NBS-type resistance protein RGC5	
GSMUA_Achr7G12770_001	Putative disease resistance protein RGA4	<i>Musa_cav_rep_c22529.path1#chr7_10298625-10298628</i>
GSMUA_Achr9G12800_001	Putative Disease resistance protein RPM1	<i>Musa_cav_rep_c4917.path1#chr9_8303143-8304138</i>

Embora ainda não esteja claro como as proteínas R funcionam na resposta de defesa da planta, é claro que elas agem em um passo inicial de vias de sinalização de defesa, quer como moléculas de reconhecimento primárias ou proteínas acessórias. Baixos níveis de expressão constitutiva de proteínas R são suficientes para a capacidade de reconhecer a infecção e induzir respostas de defesa. Existem, contudo, vários sinais de controle de transcrição da expressão do gene R e pelo menos um subconjunto de genes R é induzido acima dos seus baixos níveis de expressão constitutiva durante a elicitação de resistência basal, a expressão de RPS2, RPM1, e outros oito NBS-LRR e genes relacionados foi induzida pelo peptídeo flagelina bacteriana, flg22 (Tan *et al.*, 2007). Como resultado, existe um grande interesse da comunidade científica sobre os genes da família NB-ARC e estudos para a identificação dos mesmos podem ser vistos em diversas culturas, como em Curcuma, onde foram identificados 28 genes NBS-LRR derivados de transcrito (Joshi *et al.*, 2011), em Capim Buffel, onde a análise transcritômica derivada de Illumina RNAseq identificou, 325 genes diferencialmente expressos em dois cultivares, dentre eles genes pertencentes à família NB-ARC (Wachholtz *et al.*, 2013). O sequenciamento de 454 de transcrito de dois cultivares *Aegilops sharonensis*, um parente próximo de trigo gerou um grande número de sequências que foram mapeadas contra seis proteomas de gramíneas anteriormente caracterizados, identificando 139 sequências contendo domínios NB-ARC.

Os dados gerados neste trabalho contribuem para a elucidação dos mecanismos de resistência em *Musa*. Como sequências de RGAs foram caracterizadas através de padrões de transcrição durante as interações *Musa-Mycosphaerella*, dados de expressão temporal nesses genes candidatos potencialmente envolvidos em respostas de resistência ou de defesa em *Musa*, não só aumentam a nossa compreensão dos processos moleculares envolvidos na resistência a doenças em *Musa*, mas juntos com o desenvolvimento de marcadores RGA, oferecem um potencial para o desenvolvimento de uma gestão eficaz da doença com base em melhoramento genético da cultura, seja por meio de transformação de plantas ou de seleção assistida por marcadores.

CAPÍTULO 2: ANÁLISE DE EXPRESSÃO DE GENES POTENCIALMENTE ENVOLVIDOS EM ESTRESSES BIÓTICOS NA INTERAÇÃO *M. ACUMINATA* E *MYCOSPHAERELLA MUSICOLA*

1. Introdução

As plantas estão constantemente expostas a uma vasta gama de agentes patogênicos microbianos na natureza. Elas desenvolveram mecanismos de defesa eficazes para lidar com a infecção (Jones e Dangl, 2006), onde, na maior parte é regida por determinantes de especificidade chamados de genes de resistência (R). Este sistema imunitário simples, mas sofisticado envolve uma interação genética alelo-específico entre os produtos de genes do hospedeiro e do patógeno (Brugmans *et al.*, 2008). A partir do reconhecimento de efetores patogênicos por genes R, ocorre o desencadeamento de uma resistência sistêmica adquirida (SAR), resposta de defesa induzível que ocorre após a infecção local e fornece a resistência de longa duração e de largo espectro para a infecção secundária. O início da SAR requer aumentos de ácido salicílico (SA), que atua como molécula sinal que promove a ativação de genes PR que codificam proteínas relacionadas à patogênese (Song *et al.*, 2011).

Embora um número crescente de genes R estão sendo clonados, pouco se sabe sobre os mecanismos de regulação da expressão dos genes R em plantas. Análises de RNA tem detectado baixos níveis de transcritos para a maioria de genes R clonados de plantas (Tan *et al.*, 2007). Vários estudos têm demonstrado que os genes R são transcricionalmente regulados positivamente em resposta a agentes patogênicos e outros estímulos ambientais (Mohr *et al.*, 2010).

Os métodos para análise da expressão gênica têm avançado muito, desde microarranjos que permitem um número intermediário de amostras a serem analisadas, com conhecimento prévio dos genes, até o sequenciamento de transcritos e Análise Serial de Expressão Gênica (SAGE) que permitem uma avaliação quantitativa da expressão do gene, além da técnica de qPCR que é reconhecida pela sua alta sensibilidade e especificidade (Tan *et al.*, 2007).

Com base na importância dos genes envolvidos em processos de resistência e defesa, neste capítulo foi realizada a análise de expressão por meio de PCR em tempo real (qPCR) destes genes envolvidos na via de sinalização de respostas da planta, derivados do transcrito foliar dos genótipos contrastantes em resistência *M. acuminata* Calcutta 4 (resistente) e Cavendish Grande Naine (Suscetível) sob interação com *M. musicola*.

2. Material e Métodos

2.1. Material Biológico

2.1.1. Preparo de cDNA

Para a realização da transcrição reversa (RT) e posterior reação em cadeia de polimerase (PCR), foram desenvolvidos quatro *pools* de RNA total de *M. acuminata* Calcutta 4 e *M. acuminata* Cavendish Grande Naine inoculados e não inoculados com *M. musicola*, levando em consideração a suscetibilidade/resistência e o tratamento das amostras, inoculadas/não inoculadas.

Para evitar contaminação com DNA genômico, as amostras foram tratadas com Deoxyribonuclease I (DNase I Amplification Grade – Invitrogen), onde, para cada μg de RNA foi usada 1U da enzima. As amostras foram incubadas por 15 min a 37 °C e posteriormente precipitadas em 2,5 V de etanol 100% gelado e 0,1 V de acetato de sódio 3M pH 5,2 e incubadas *overnight* a -20 °C, seguindo o protocolo de acordo com o recomendado pelo fabricante.

Para obter-se o DNA complementar (cDNA) que serviu posteriormente como molde para amplificação pela reação de cadeia em polimerase, uma alíquota equivalente a 2 μg de RNA total foi destinado à síntese de cDNA. Para a mesma foi utilizada o kit SuperScript™ II Reverse Transcriptase – Invitrogen, com os seguintes reagentes: - Tampão 5X concentrado (Tris/HCl a 250mM, pH 8,3; KCl a 375mM; MgCl₂ a 15mM e ditiotretol a 100mM), mistura de desoxinucleotídeos (dATP, dCTP, dGTP e dTTP) com concentração de 10mM cada, iniciadores randômicos, 200 unidades de transcriptase reversa. Seguiu-se o protocolo experimental recomendado pelo fabricante da enzima.

2.2. Desenho de primers em regiões UTR de genes de defesa derivados de sequências de transcrito (sequenciamento 454)

Os genes foram selecionados de acordo com categorias relacionadas aos processos de defesa da planta. Por meio do mapeamento das sequências de transcritos derivadas de sequenciamento 454 com os dados genômicos de *M. acuminata* DH Pahang (<http://banana-genome.cirad.fr/>) foi possível identificar quais os *contigs* mapearam em genes

potencialmente envolvidos em respostas de defesa. Os genes candidatos selecionados, já conhecidos na literatura, participam de diferentes momentos da imunização da planta, como, por exemplo, morte celular programada (Cathepsin), choque oxidativo (Glutathione S-transferase), metabolismo de lipídios (Lipoxygenase A LOX1), modificação de parede celular (Callose synthase), fatores de transcrição (WRKY), via dos fenilpropanóis (Phenylalanine ammonia-lyase 3 PAL3), PR-proteínas (Thaumatococcus) e reguladores do processo de defesa (domínio NAC).

Utilizando o Gbrowse (Droc *et al.*, 2013) do genoma de *M. acuminata* DH Pahang, foi possível identificar o início e fim dos genes selecionados e com isso estipular a região UTR (200pb a mais do início e fim da sequência do gene). Os *primers* foram desenhados flanqueando a região de interesse (exemplo na Figura 9) utilizando o programa Primer3 plus (Untergasser *et al.*, 2007) abordando os seguintes parâmetros: tamanho do *primer*, de 19 pb a 22 pb; Tm do *primer* de 55°C a 62°C; porcentagem de GC do *primer*, de 45% a 55% e tamanho do produto de 150 pb a 200 pb.

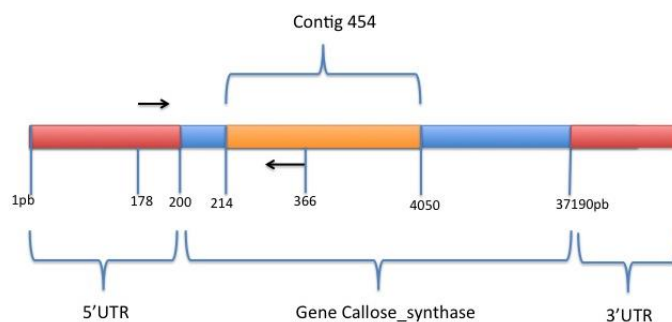


Figura 9: Modelo de desenho de *primer* para o gene envolvido em resposta de defesa 'Calose sintase'. O *primer forward* foi desenhado dentro da região 5'UTR (vermelho) e o *reverse* dentro do contig derivado do transcrito de C4 (amarelo).

2.3. Desenho de primers específicos derivados de contigs NBS-LRR

Uma estratégia computacional foi desenvolvida por Miller *et al.*, (2008) para descoberta dos domínios NBS e LRR presentes em genes R de plantas monocotiledôneas e, com *primers* degenerados, amplificou 174 RGAs em *M. acuminata* Calcutta 4 que foram montados em 52 *contigs*. Esses *contigs* foram utilizados neste trabalho para o desenho de *primers* específicos para RGAs por meio do programa Primer3Plus (Untergasser *et al.*, 2007) abordando os seguintes padrões: tamanho do *primer*, de 19 pb a 22 pb; Tm do *primer* de 55°C a 62°C; porcentagem de GC do *primer*, de 45% a 55%.

2.4. Desenho de primers em regiões UTR de genes de resistência

2.4.1. Primers para genes de resistência e defesa derivados de sequências de transcrito (sequenciamento 454) e mapeados no genoma de *M. acuminata* DH Pahang

Por meio do mapeamento dos *unigenes* de *M. acuminata* derivados de sequências de transcrito (sequenciamento 454) contra as sequências do genoma de *M. acuminata* DH Pahang foi possível identificar, com base na anotação genômica, os genes pertencentes às diversas classes de genes de resistência e defesa. Em posse da nomenclatura, foi realizada a extração da sequência, inclusive da região UTR e as informações completas dos genes de interesse, utilizando o programa Gramene (Monaco *et al.*, 2013), que é um programa aberto de recursos de dados para a análise de genoma comparativo em gramíneas. Em seguida a sequência UTR do gene de interesse foi submetida a um BLAST contra o banco de dados do genoma de *M. acuminata* DH Pahang (<http://banana-genome.cirad.fr/>) para confirmar a especificidade da sequência em todo o genoma. Como anteriormente, os *primers* foram desenhados utilizando o programa Primer3 plus (Untergasser *et al.*, 2007) abordando os seguintes padrões: tamanho do *primer*, de 19 pb a 22 pb; Tm do *primer* de 55°C a 62°C; porcentagem de GC do *primer*, de 45% a 55% e tamanho do produto de 150 pb a 200 pb.

Para o caso em que a sequência UTR não tenha sido anotada, a busca da região foi realizada utilizando o programa FGENESH (Solovyev *et al.*, 1994). O arquivo de saída (exemplo na Figura 10) disponibiliza além da sequência protéica equivalente à sequência gênica, demonstra também toda a estrutura do gene de interesse com base em um banco de dados de monocotiledôneas. Dessa forma tem-se disponível as posições de início e fim do gene.

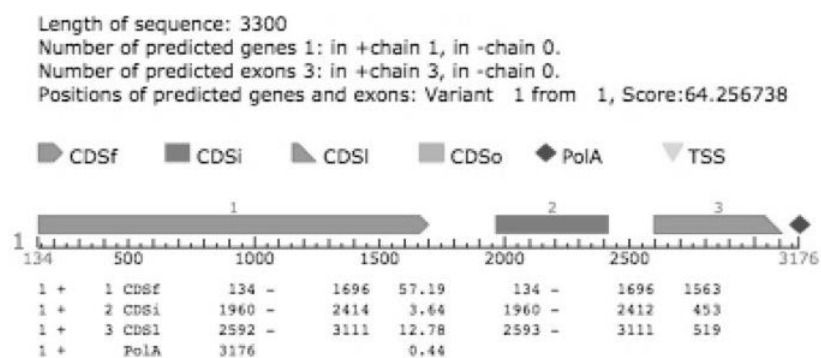


Figura 10: Arquivo de saída do programa FGENESH mostrando a estrutura de um determinado gene de resistência encontrado em *M. acuminata* DH Pahang. CDSf representa o códon de iniciação, CDSi, o exon interno e CDSl, o códon de terminação.

Para a confirmação dos dados obtidos pelo programa FGENESH (Solovyev *et al.*, 1994), a sequência protéica foi submetida ao programa Genewise (Birney *et al.*, 2004) onde é feita realizada a comparação entre sequência protéica e sequência genômica, permitindo a identificação das regiões intrônicas e exônicas. É possível observar pelo arquivo de saída do Genewise (Figura 11) que o alinhamento, entre proteína e gene, neste exemplo, começa somente na base 134, ou seja, as primeiras 133 bases que estavam presentes no gene, não foram traduzidas, identificando, dessa forma, a região 5'UTR. O mesmo procedimento foi realizado para a identificação da região 3'UTR.

```

genewise output
Score 2072.11 bits over entire alignment
Scores as bits over a synchronous coding model

Warning: The bits scores is not probablistically correct for single seqs
See WWW help for more info

FGENESH          1  MATDNVYVLFYLLAFLCIQPNVCDGALTSGCNPAERSALLEFKRGLKDP
                  MATDNVYVLFYLLAFLCIQPNVCDGALTSGCNPAERSALLEFKRGLKDP
                  MATDNVYVLFYLLAFLCIQPNVCDGALTSGCNPAERSALLEFKRGLKDP
EMBOSS_001      134 agagagtgcttttgtctaccagtgggcatgtacggaagccgtacgcagc
                  tccaatatttattottgtacatgagctccggaccaggottataggtaac
                  gtactcctcccagccactacccccatacgccctcaattatgcaacgatt

FGENESH          50  TNRLSSWVGEDCCKWEGVTCSNHTGHVVKLDLHNPHPFSDFGDEPYNNW
                  TNRLSSWVGEDCCKWEGVTCSNHTGHVVKLDLHNPHPFSDFGDEPYNNW
                  TNRLSSWVGEDCCKWEGVTCSNHTGHVVKLDLHNPHPFSDFGDEPYNNW
EMBOSS_001      281 aaactttggggttatgggataacagcggacgccacccttgtgggcta
                  cagtcctgtgaaggagagtcggaacgattatataacactcatgaacaaag
                  ccggctggtaccaggtggcctttgtccggccctgttcttttcgacccg

```

Figura 11: Arquivo de saída do programa GeneWise mostrando o alinhamento entre proteína (FGENESH) e DNA (EMBOSS_001).

Os *primers* foram desenhados para a região especificada utilizando o programa Primer3plus (Untergasser *et al.*, 2007), com os mesmos padrões para tamanho do *primer*, Tm do *primer*, porcentagem de GC e tamanho do produto como descrito anteriormente.

2.4.2. *Primers* para genes de resistência derivados de sequências de transcrito (sequenciamento 454) mapeados em sequências BAC de *M. acuminata* Calcutta 4 e Cavendish Grande Naine

Primeiramente foi feito um mapeamento dos dados brutos de sequências de transcrito (sequenciamento 454) com as sequências consensos de oito BACs de *M. acuminata* Calcutta 4 e Cavendish Grande Naine, previamente selecionados com sondas de NBS-LRR RGAs e sequenciados (Miller *et al.*, 2008). Isso possibilitou uma montagem guiada formando *contigs* na medida em que os dados brutos foram mapeando na posição do

BAC. Foi feita a predição de função dos genes mapeados por meio da anotação manual disponibilizada (<http://gnannot.Musagenomics.org/cgi-bin/gbrowse/Musa/>). Os genes candidatos potencialmente envolvidos em resposta aos estresses bióticos foram selecionados e as suas sequências nucleotídicas foram delimitadas para cerca de 200 pb amontante ao códon de iniciação e 200 pb ajusante ao códon de terminação, na tentativa de incluir as sequências UTR (Figura 12).

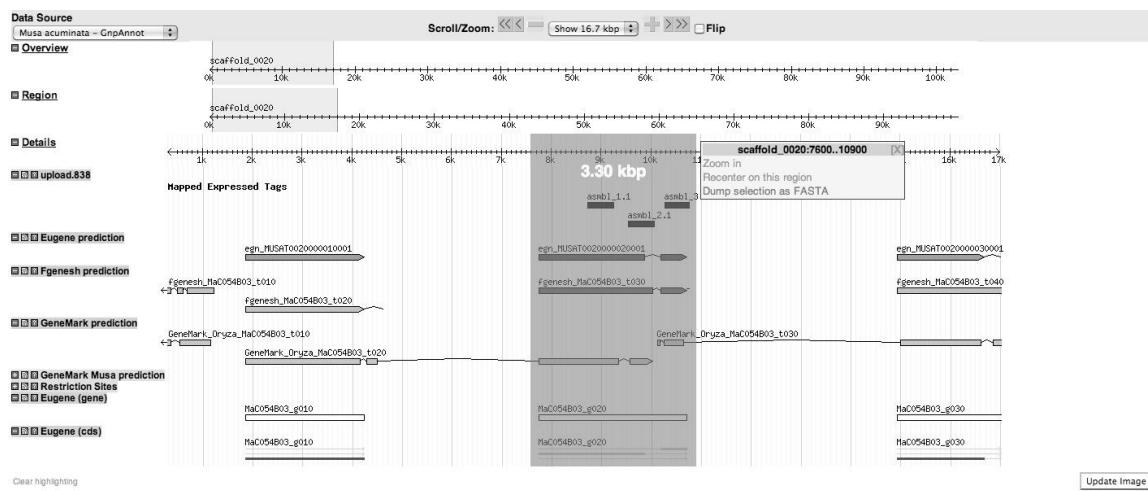


Figura 12: Interface do gnapnot mostrando sequências derivadas do transcrito (sequenciamento 454), assemblies (1.1, 2.1, 3.1) mapeados em um gene potencialmente envolvido em resistência, na sequência do BAC MaC054B03 de *M. acuminata* Cavendish Grande Naine.

A confirmação da região UTR foi dada por meio do programa FGENESH (Solovyev *et al.*, 1994) com a identificação da estrutura do gene mostrando exatamente em qual base o gene inicia e termina. Dessa forma foi possível visualizar que os 200 pb antes do gene e depois do gene não fazem parte de um outro gene ao lado. Os resultados foram também confirmados utilizando o programa GeneWise promovendo um alinhamento entre a sequência gênica e a sequência protéica, mostrando a região codante, excluindo as regiões UTR.

Os *primers* foram desenhados dentro da região UTR (*forward*, *reverse* ou ambos) pelo programa Primer3Plus (Untergasser *et al.*, 2007) obedecendo os parâmetros de tamanho do *primer*, de 19 pb a 22 pb; Tm do *primer* de 55°C a 62°C; porcentagem de GC do *primer*, de 45% a 55% e tamanho do produto de 150 pb a 200 pb.

2.5. Desenho de primers para genes referência

Primers potencialmente constitutivos foram desenhados por meio de uma busca inicial dos genes da actina (gi 151413782) e elongation factor (gi 66775523) e o gene 18S do rDNA (gi 2795850) de *M. acuminata* no banco de dados *National Center Biotechnology Information* (www.ncbi.nlm.nih.gov/). As sequências completas dos genes em questão foram submetidas ao BLAST (*Basic Local Alignment Search Tool*) versão 2.2 (Altschul, 1997) para confirmação de identidade e funcionalidade no genoma de *M. acuminata* DH Pahang. A partir disso, foram desenhados *primers* para esses genes com o uso do programa Primer3Plus (Untergasser *et al.*, 2007). Os *primers* desenhados anelam em exons diferentes com o objetivo de amplificar uma região não codante (intrônica) além de codante.

2.6. Confirmação de amplificação *in silico*

Todos os *primers* desenhados foram primeiramente testados via PCR eletrônico por meio do programa *primersearch* pertencente ao pacote EMBOSS_GUI (bioinfo.hku.hk/emboss/) para a confirmação de amplificação específica antes dos *primers* serem sintetizados.

Os testes foram realizados confrontando *primers* contra um banco de sequências brutas de transcritos de *M. acuminata* Cavendish Grande Naine e *M. acuminata* Calcutta 4 (sequenciamento 454); contra um banco de sequências de genes de interesse em transcritos (sequenciamento 454) mapeados em sequências de oito BACs de *M. acuminata* Calcutta 4 e Cavendish Grande Naine anotados e contra um banco de sequências anotadas dos mesmos oito BACs

2.7. PCR em tempo real

As amplificações por PCR em tempo real foram realizadas em triplicatas e utilizando 4 ng de cDNA por reação (diluição de 10^{-2}). As reações foram preparadas com reagentes padronizados para PCR em tempo real (Platinum® SYBR® Green qPCR Super Mix-UDG w/ROX kit) adicionado dos conjuntos de *primers*. As condições das reações utilizadas na RT-qPCR foram 50°C por 2 minutos, 95°C por 10 minutos, seguido de 40 ciclos compreendendo a etapa de desnaturação a 95°C por 15 segundos, anelamento e extensão do

primer a 60°C por 60 segundos, seguido de uma curva de dissociação gerada a partir da desnaturação do produto amplificado pelo aumento da temperatura no fim da reação para verificar a especificidade do *primer* (95°C por 15 segundos, 60°C por 60 segundos e 95°C por 15 segundos).

As reações de PCR foram realizadas pelo equipamento ABI 7300 *Real-Time PCR* System (Applied Biosystems, Foster City, CA, EUA) a cada ciclo de amplificação e, posteriormente, analisados pelo *Sequence Detection Programa* (SDS) v1.3 (Applied Biosystems). Todas as reações foram submetidas às mesmas condições de análise e normalizadas pelo sinal do corante de referência passiva ROX para correção de flutuações na leitura decorrentes de variações de volume e evaporação ao longo da reação. O resultado, expresso em valor de CT, se refere ao número de ciclos de PCR necessários para que o sinal fluorescente atinja o limiar de detecção. Os resultados individuais expressos em valores de CT foram a seguir transferidos para planilhas e agrupados de acordo com o tempo de inóculo (DAI) analisado e a condição infectado ou não-infectado para realização da análise estatística. Para tal, os dados foram primeiramente ajustados utilizando o programa Miner, que aborda um método objetivo utilizando cálculos baseados na cinética individual das reações, sem a necessidade da curva padrão para calcular a eficiência da reação de PCR (Zhao e Fernald, 2005). As amostras foram analisadas utilizando o programa REST© - *Relative expression programa tool* (Pfaffl *et al.*, 2002) comparando e gerando um nível de significância entre dois grupos, controle e inoculado em diferentes tempos de inóculo. A normalização da expressão gênica foi realizada por meio do gene referência *elongation fator* para *M. acuminata*.

3. Resultados e Discussão

3.1. Quantificação de RNA total de *M. acuminata* Calcutta 4 e *M. acuminata* Cavendish Grande Naine inoculados e não inoculados com *M. musicola* e síntese de cDNA

O RNA total isolado apresentou quantidade satisfatória em cada *pool* analisado (Tabela 8) e alta qualidade, com valores de $A_{260/280}$ variando de 1,9 a 2,0.

Tabela 8: Comparação de valores quantitativos de concentração de RNA total de folhas de *M. acuminata* dos quatro *pools* baseados em análises conduzidas com espectrofotometria NanoDrop ND-1000 e Qubit Quant-it™ RNA assay.

<i>Pools</i> de RNA total de folha de <i>M. acuminata</i>	Concentração de RNA total (ng/μl) (método Qubit Quant-it™ RNA assay)	Concentração de RNA total (ng/μl) (método Nanodrop ND-1000)
<i>Pool</i> 1: RNA total de folha de <i>M. acuminata</i> Calcutta 4 não inoculado	24,9	35,7
<i>Pool</i> 2: RNA total de folha de <i>M. acuminata</i> Calcutta 4 inoculado com <i>M. musicola</i>	27,6	38,1
<i>Pool</i> 3: RNA total de folha de <i>M. acuminata</i> Cavendish Grande Naine não inoculado	17,5	26,7
<i>Pool</i> 4: RNA total de folha de <i>M. acuminata</i> Cavendish Grande Naine inoculado com <i>M. musicola</i> .	29,8	43,4

Corroborando com a boa qualidade do RNA total extraído, as bandas no gel de agarose apresentaram-se íntegras e sem contaminação visível de DNA genômico (Figura 13).

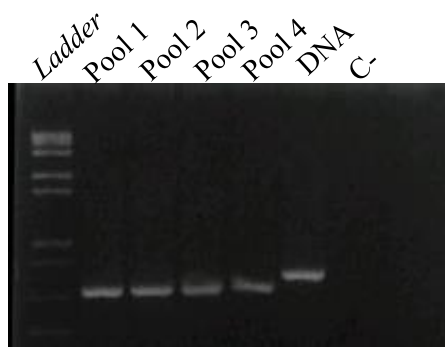


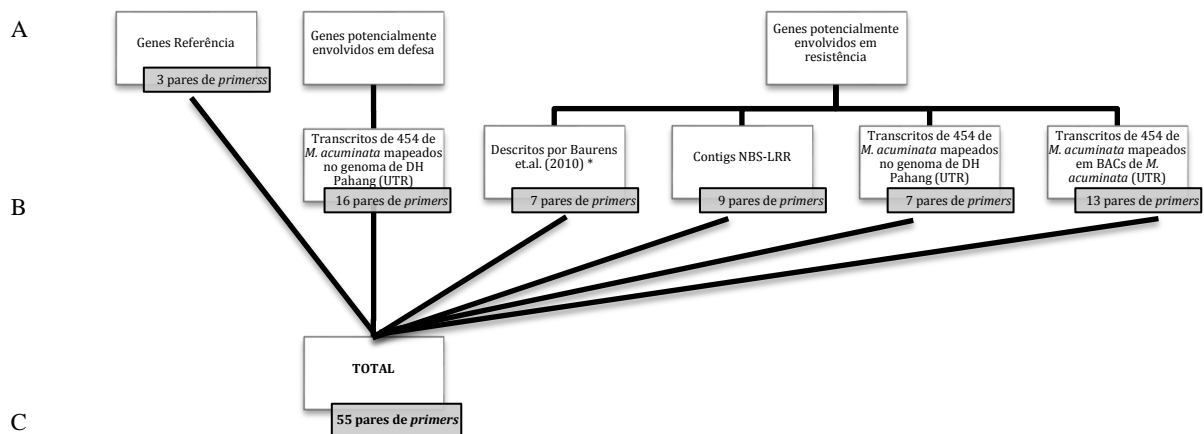
Figura 13: Análise eletroforética em gel de agarose 1% mostrando a amplificação dos quatro *pools* de cDNA utilizando *primers* intrônico para detecção de contaminação por DNA. Canaleta 1: Marcador molecular 1Kb Plus (Invitrogen); 2: amostra de RNA total de folha de *M. acuminata* Calcutta 4 não inoculado (*Pool* 1); 3: amostra de RNA total de folha de *M. acuminata* Calcutta 4 inoculado com *M. musicola* (*Pool* 2); 4: amostra de RNA total de folha de *M. acuminata* Cavendish Grande Naine não inoculado (*pool* 3); 5: RNA total de folha de *M. acuminata* Cavendish Grande Naine inoculado com *M. musicola* (*pool* 4).

Os *primers* específicos para o gene constitutivo *elongation factor* em *M. acuminata* (NCBI gi 66775523) estão localizados em regiões exônicas, que flanqueiam uma região intrônica. Neste caso, quando ocorre a amplificação no cDNA o fragmento gerado é menor

que quando amplificado a partir do DNA genômico, já que no cDNA existe apenas a região codante. Dessa forma, de acordo com a Figura 13, é possível notar a diferença nos tamanhos dos produtos gerados nos *pools* (de cDNA) e no DNA genômico, confirmando que o material está livre de DNA contaminante.

3.2. Desenvolvimento de primers específicos para análise de expressão de genes candidatos por RT-qPCR

Foram utilizados 55 pares de *primers* para análises de expressão por PCR em tempo real, incluindo os genes referência (Figura 14)



* *Primers* desenhados por Baurens et.al. (2010) para *M. balbisiana* a partir de RGA 08 de *M. acuminata*

Figura 14. Fluxograma ilustrando o desenvolvimento dos pares de *primers* para genes envolvidos em respostas aos estresses bióticos para análises de expressão via qPCR. A) Os pares de *primers* foram desenvolvidos a partir de genes potencialmente envolvidos em resistência e defesa da planta, além dos genes referência utilizados para a normalização da reação. B) Número de pares de *primers* desenvolvidos a partir de diferentes fontes. C) Número total de pares de *primers* utilizados em testes e análises de expressão por meio de PCR em tempo real.

Os pares de *primers* desenhados e utilizados estão listados na Tabela 9.

Tabela 9: Sequências de primers para análises de expressão por meio de PCR em tempo real de genes referência e genes potencialmente envolvidos em processos de resistência e defesa de *M. acuminata*.

Fonte	Primer	Primer FW	TM	Primer RV	TM	Produto	Condição da amplificação
<i>Contigs</i> NBS-LRR	RGA02*	GAATCGTTCGAGGGACAAAG	53.4	TCCAGACATCATCCAACACG	54.7	94	Específico
	RGA14	TCAACACCTGAGCACAACC	55.7	AACGTTCTTGTGGAAGACG	55.3	83	Inespecífico
	RGA30	AACATGGTCGCACAAGAGTG	55.7	TTCCTCACCTCCAGTTTG	54.7	147	Não expresso
	RGA31	GGTGCATCTCCATCAATC	54	GACAGATCGTTCGGTTTTGG	53.9	80	Não expresso
	RGA37*	AACATCATGGCAACCTCACC	55.6	CAAACCAAAGACCAGCACTC	54.2	71	Específico
	RGA40	ACCGCAAAACGAAGAGAGTG	55.6	ACAAAGCACCGCTTATAGGTG	56	81	Não expresso
	RGA43*	ACTTCGGTGGGATCCAAATC	54.8	AATGACTCCGCTGAAGTTGG	55.2	92	Específico
	RGA44*	TCAAGGACACAGGCATGAAG	55	CCACGCATTTCTGTGAGTC	54.2	107	Específico
RGA46	GGGGTTCACAGAACCAATG	54.1	AGATTTGCTCGATGGTCTC	54.5	77	Não expresso	
Baurens et al., 2010	MBPKW_RGA08L	GCAGCCAACCTATCGATCTGCG	60.4	ACGTCTTCTGGTTTTGGAGTGGCA	60.9	193	Inespecífico
	MBPKW_RGA08E	ACGTCTTCTGGTTTTGGAGTGGCA	60.9	CGCAACCAACCTCGATATCTGTGC	60.1	194	Inespecífico
	MBPKW_RGA08G	TCGGCAGCAGGATAGCAGTAGCA	62.3	GATGCCGCCAAAATGGCCCG	62.7	167	Inespecífico
	MBPKW_RGA08K	AGATGCAGACCTGTGAGTCTGTCAA	59.7	TGAGTGTTTTGCAGTATGCACGTTTC	58.3	248	Não expresso
	MBPKW_RGA08M	CATGCTCACGCTCGACCATC	57.8	TGGTAAATATCATGGAAGCCGTGAA	56.5	151	Não expresso
	MBPKW_RGA08O	ACGTCTTCTGGTTTTGGAGTGGCA	60.9	CGCAACCAACCTCTCGATCTGCAT	61.5	194	Inespecífico
	MBPKW_ACT	GAGAAGATACAGTGCCTGGA	50.7	ATTACCATCGAAATATTTAAAG	44.3	231	Inespecífico
Referência	EF*	AACCCCAAATATTCCAAGG	59.9	AGATTGGCACGAAAGGAATC	59.1	107	Específico
	18S	TTTATGAAAGACGAACCACTGC	59.3	GACGGTATCTGATCGTCTTCG	59.7	95	Específico
	ACTIN1	GAATGGTCAAGGCTGGTTTC	59.5	TCTTTTTGACCCATCCCAAC	59.8	109	Inespecífico
Transcritos de 454 de <i>M. acuminata</i> mapeados em BACs de <i>M. acuminata</i> (UTR)	454_Ma4140M09_A5.1_58502_61436_Glut_perox_EU_qPCR	CTACCTTGCCATCCCTCA	60,2	CAAGAGCGAAGGTTTCAGAGG	60,1	175	Inespecífico
	454_Ma4140M09_A3.1_43118_46111_FLS2_EU_qPCR*	AATCCCAATCCATGGCTACA	60,2	ATCTTTCAGGCCTCGTTTGA	59,8	164	Específico
	454_Ma4140M09_A2.1_34474_38284_LRR_EU_qPCR*	TTGGAGGAGTACCCGAAGAA	59,7	AGCGAGGCTAGAAGGAATTG	58,7	170	Específico
	454_Ma4140M09_A1.1_28855_31860_FSL2_EU_qPCR*	CCAGCAACCAACCTAGT	60	GATCTTTCAGGCCTCGTTTG	59,8	176	Específico
	454_Mac054B03_A9.1_52557_55372_Glut_perox_EU_qPCR	AAGAAGCTTTTGGGGCTTTC	59,8	TACACAAATGAACGCGGTGT	60	174	Inespecífico

	454_Mac054B03_A5.1_26128_28995_FLS2_EU_qPCR*	ATTCCATGGCTACGGACAAT	59,3	TTCAGGCCTCGTTTGAAGCTC	60,4	154	Específico	
	454_Mac054B03_A4.1_19218_22055_Glut_perox_EU_qPCR	ACATTGCTTGAAGGGATTGC	60,1	TGTGCTAATCCCAGGTTTCC	59,9	192	Inespecífico	
	454_Mac054B03_A1.1_7733_10711_FLS2_EU_qPCR*	CAGCAACCATAACCCCATTC	60,2	AGGATCTTTCAGGCCTCGTT	60,2	168	Específico	
	454_Mac139M12_A3.1_95651_98827_NB_ARC_EU_qPCR	TGGAGACGTGAGCATGAAAG	60	AGACTGGCCCATGATCAAAC	59,9	192	Inespecífico	
	454_Mac139M12_A2.1_37998_41141_NB_ARC_EU_qPCR	CGATCTCCCTACCACTCCAA	60,1	AGTCGATGATGATGCTGCTG	60	150	Inespecífico	
	454_Mac091O16_A1.1_24080_25619_RPI_NB_ARC_EU_qPCR	CGTTGGAGGTTTTGGATCTC	59,5	CGTGCTTGATGATTCCTTCA	59,8	150	Inespecífico	
	454_Mac063A04_A1.1_18741_21917_NB_ARC_EU_qPCR	GTGGAGACGTGAGCATGAAA	59,8	GGCCCATGATCAAACACTTT	59,8	188	Inespecífico	
	454_Ma4052E23_A4.1_54478_59826_ser/thr_Kin_EU_qPCR	CCTATTGAGCCGACGACGTA	61,2	GCTGCAATCCATGTCAGAA	59,8	190	Inespecífico	
Transcritos de 454 de <i>M. acuminata</i> mapeados no genoma de DH Pahang (UTR)	Chorismate_5UTR*	AGGCTTGTGGTAGGTCGTTG	60,2	ATCTCCTTCCCCGAGTAAA	59,9	193	Específico	
	Chalcone_5UTR	AGACTTCCTCGTCCTTCGTG	59,5	TGAAGGCGTGGGAGAAAAC	61,2	158	Pouco expresso	
	Bax_inhibitor	AATGGGCGATGTTTGTAAGG	59,8	CCGAATTGGATTACGTGCT	60	191	Pouco expresso	
	Thaumatococin	TGCTATGTTGATGAGCACGA	59	ACCCATCAATCAGATCACCA	58,7	150	Inespecífico	
	Npr1_5UTR*	CTGTAAACACCGACAAGCTCT	57,2	TCAGCTCTTCGATCTCATGG	59,1	177	Específico	
	Cell_death	GAAGACGAATGGGTCCTTCA	60	GGGAGTCAGCAAACGAAAA	60,2	160	Inespecífico	
	Peroxidase	GGTGGTGAGCAAGGAGAAGA	60,4	TTCCTGGTGGAAATGACTCC	59,9	183	Inespecífico	
	Ubiquitin	CAGTTGGTGC GACTTCCATA	59,7	AATACCTGGCATCCACATCTC	58,9	190	Pouco expresso	
	Cathepsin	CATCCGCTAAAGCCAGGATA	60,2	TTCTTCGGTTCCTGATCCAT	59,5	200	Inespecífico	
	Pal	GAGCATCACATGTGTGGTAGAA	60	TGCCCCTCGTTCATCTTCTC	59	170	Inespecífico	
	Nac*	TGGCAAGGGAGAGGAGTAAA	59,8	CCCTCAGCTTCTGGGTCATA	60,2	153	Específico	
	Lox_5UTR	GCCACGGCTTAAGGTTGTT	60,1	CAGAGTTTTCGAGCGTTTCC	60	192	Pouco expresso	
	isoflavona_reductase*	CCGATGATCAGGATTTTGCT	60	GGGGATCGTAATGTTGTGG	60,1	174	Específico	
	Endoglucanase	TTCGATATCTCCAGCGATCC	60,1	TTCCATTCCTCTCTCTCTT	60	155	Pouco expresso	
	defensin	AGAGCAAAAGCAGTGGAGGA	60,1	TGCAACATGAGCTGTTCAT	60,3	151	Pouco expresso	
	chitinase_5UTR	AAGCATTCTGCGTCACTCAA	58,1	CCAAGTGCTGAGAACAACCTTCTT	58,7	199	Pouco expresso	
	Transcritos de 454 de <i>M. acuminata</i> mapeados no genoma de DH Pahang (UTR)	GSMUA_Achr10G29330_001*	CAGGCCTACCCATTTGAAGCTG	60,5	TTTTCTTTGTGCAGGCTGTG	60	193	Específico
		GSMUA_Achr10G24910_001*	CTGGCAACACATTCATTTTCG	60,1	TAAGGGGAGGTTTGTATGAG	60,1	176	Específico
		GSMUA_Achr10G07100_001*	CGCATGTACGATGCAAATGT	60,6	AACCAAGCAGAGGCCTTTTA	59	165	Específico

GSMUA_Achr6G21780_001*	GAAGCACAGAGGGAGCAAGT	59,6	AGTTCTCGTCCCAATCATCG	60,1	180	Específico
GSMUA_Achr5G11980_001*	AGGTGAATGAATGCCCAAAG	59,9	ACAGAGGGAGGCAGAAGACA	60	184	Específico
GSMUA_Achr4G26160_001*	TCTTCCAAGGAATGGTCTGC	60,2	CGCATCTTGTCTGCATGGT	60,9	195	Específico
GSMUA_Achr3G22290_001	TATCAAGCTCCCGATGAAGG	60,2	CAACCTCGCTTTCGAAGAAG	60,1	187	Inespecífico

TM: Temperatura de *melting* (°C); *Primer FW* (*Primer forward*); *Primer RV* (*Primer reverso*); Produto (pb); * Pares de *primers* testados e analisados por PCR em tempo real.

3.3. Análises de especificidade

3.3.1. Genes Referência

O gene EF apresentou amplificação específica e foi utilizado como normalizador para as reações de PCR em tempo real. O uso dos demais genes como normalizadores foi inviabilizado devido à grande variação na amplificação (18S) e falta de especificidade (ACTN1 e MBPKW_ACT).

A Figura 15 apresenta os gráficos das análises de curva de dissociação geradas nas corridas dos genes de controle interno (*Elongation Factor*). Foi observada a formação de um pico representando *amplicons* únicos em todas as replicatas técnicas.

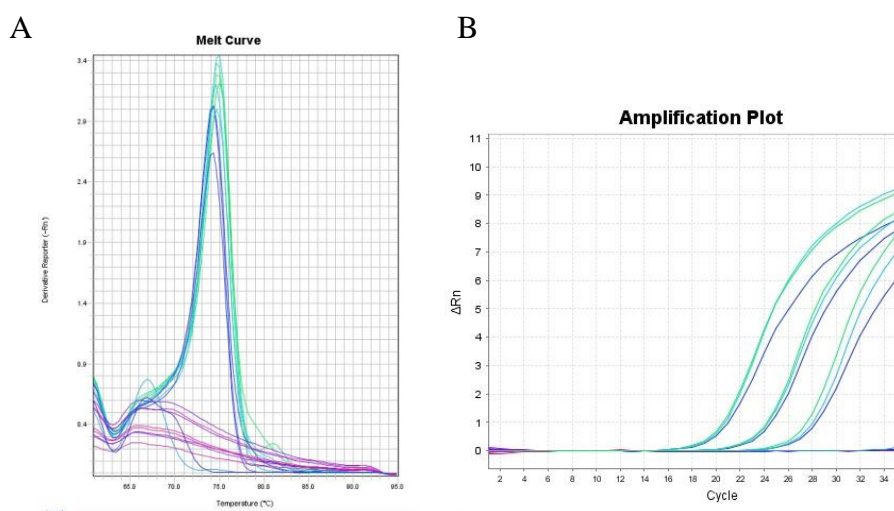


Figura 15: Curvas apresentadas na corrida de PCR em tempo real do gene referência *Elongation factor*. A) Curva de dissociação representando a temperatura em função da fluorescência com formação de pico único gerado pelo alcance da temperatura de *melting* da sequência e queda abrupta da fluorescência. B) Curva de amplificação do gene *Elongation Factor* com *pool* de amostras de *M. acuminata* Calcutta 4 controle nas concentrações de 10^{-1} , 10^{-2} e 10^{-3} .

3.3.2. Genes de defesa e RGAs

A eficiência de amplificação dos genes alvo com os pares de *primers* desenhados, foi avaliada inicialmente por meio de PCR convencional a partir de DNA genômico isolado de *M. acuminata* Calcutta 4 e posteriormente por PCR em tempo real o qual tiveram sua eficiência de amplificação determinada por meio do programa *Real-time PCR Miner* Version 3.0 (<http://www.miner.ewindup.info/>) e por meio de análises das curvas de *melting*, abordagem amplamente utilizada para proceder às análises de resultados obtidos na PCR em

tempo real (Figura 16). Durante a reação, dados da amplificação são gerados, levando em consideração as características de dissociação da dupla cadeia de DNA para proporcionar um ponto de fusão a que, pelo menos 50% do DNA é desnaturado (Faux *et al.*, 2014).

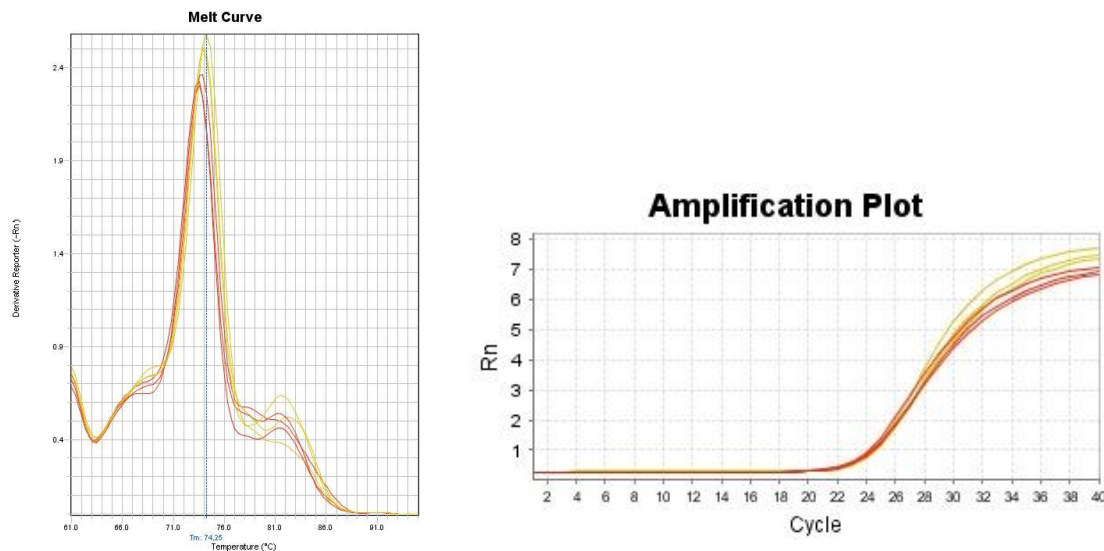


Figura 16: Curvas de dissociação (A) e de amplificação (B) do RGA 37 em triplicatas de Cavendish Grande Naine inoculado (vermelho) e controle não inoculado (amarelo). Em B é possível visualizar a qualidade das triplicatas por meio da padronização das curvas de cada réplica.

Segundo Filion (2012) um pico característico na curva de dissociação vai distingui-lo de artefatos de amplificação, que em geral fundem a temperaturas mais baixas em picos mais largos. Dessa forma, de todos os pares de *primers* avaliados, incluindo os genes referência, 21 geraram um pico único e bem definido na curva de dissociação, sem a presença de picos em temperaturas baixas, confirmando a especificidade dos *primers* escolhidos. Dos demais, seis pares de *primers* não mostraram expressão, esses foram derivados dos *contigs* NBS-LRR e de Baurens (2010), desenhados sem informação prévia de expressão em *M. acuminata*; 21 pares de *primers* não demonstraram especificidade esperada, sendo a maioria deles desenhada para regiões não UTR, e, sete pares de *primers* foram desenhados para genes com baixa expressão, ficando a geração da curva de dissociação prejudicada devido à baixa fluorescência nas análises desses genes.

Essa baixa especificidade se deve ao fato dos genes R NBS pertencerem a maior classe de famílias gênicas e a maior categoria de genes R de plantas, possuindo um domínio com diversos motivos altamente conservados, tais como P-loop, quinase e GLPL (Zheng *et al.*, 2012) corroborando com os resultados de baixa taxa de especificidade.

Os genes cujos *primers* foram anteriormente testados por Baurens (2010) não obtiveram amplificação específica em *M. acuminata*. Isso pode ser explicado já que foram desenhados para um ortólogo de NBS-LRR RGA08 (Miller *et al.*, 2008) em sequência de clone BAC de *M. balbisiana* "Pisang Klutuk Wulung" (PKW) e o fato de pertencer a outra espécie pode ter causado a inespecificidade dos *primers* em *M. acuminata*.

Após todas as análises de especificidade e eficiência, um total de 19 NBS-LRR RGAs e genes envolvidos em respostas de defesa foram selecionados e submetidos às análises de expressão por PCR em tempo real, além do gene referência *elongation factor* (EF).

3.4. Análises de expressão

Após os pares de *primers* terem sido testados e a eficiência da amplificação mensurada, procedeu-se à investigação do padrão de expressão dos genes candidatos em *pools* de *M. acuminata* Calcutta 4 e *M. acuminata* Cavendish Grande Naine, inoculados e não inoculados com *M. musicola*, por meio de reações de PCR em tempo real (Figura 17)

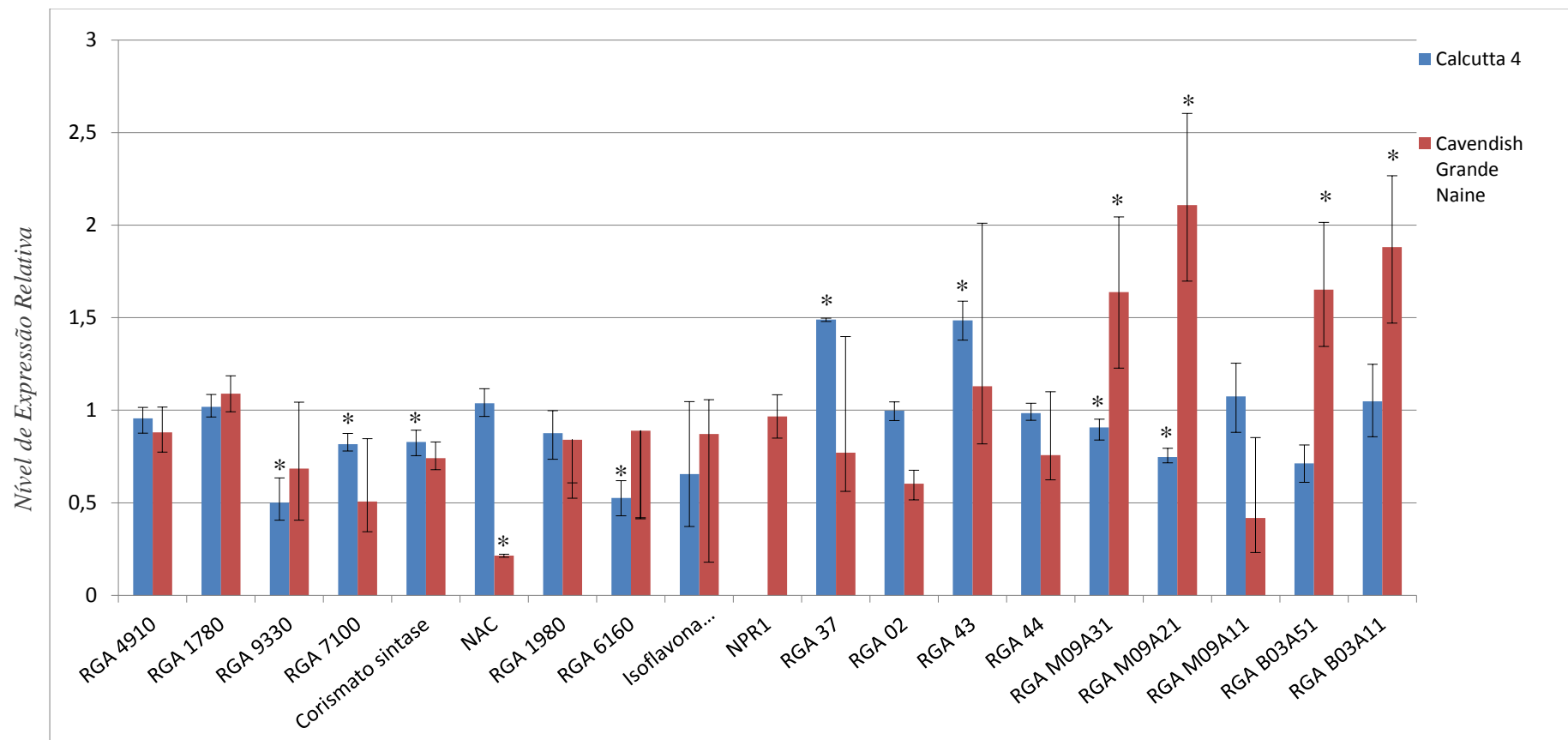


Figura 17: Níveis de expressão relativa dos genes potencialmente envolvidos em resistência (RGAs) e respostas de defesa de *M. acuminata* Calcutta 4 e Cavendish Grande Naine submetidos ao inóculo com *M. musicola*, comparados ao controle não inoculado. Valores acima de 1,0 representam regulação positiva (aumento de expressão em relação com o controle). * Modulação estatisticamente significante

As reações de qPCR dos 19 genes de alvo foram realizadas conforme descrito anteriormente. A Tabela 10 mostra os valores de expressão gênica relativa e eficiência da reação de PCR em tempo real dos grupos resistente e suscetível inoculados com o patógeno em relação aos grupos controle.

Tabela 10: Resultados de expressão relativa, eficiência da reação de PCR em tempo real, nível de significância e tipo de regulação da expressão dos genes potencialmente envolvidos em respostas de defesa e resistência (RGAs).

Gene	Calcutta 4				Cavendish Grande Naine				
	Expressão	Eficiência da reação	P(H1)	Regulação	Expressão	Eficiência da reação	P(H1)	Regulação	
RGA 4910		0,956	0,8554	0,505	0,881	0,8554	0,258		
RGA 1780		1,018	0,8479	0,83	1,09	0,8479	0,224		
RGA 9330		0,501	0,8803	0	Negativa	0,684	0,8803	0,258	
RGA 7100		0,817	0,86	0	Negativa	0,506	0,86	0,183	
Corismato sintase		0,828	0,55	0,034	Negativa	0,741	0,55	0,086	
NAC		1,037	0,847	0,501		0,214	0,847	0	Negativa
RGA 1980		0,876	0,84	0,198		0,84	0,84	0,065	
RGA 6160		0,526	0,89	0,035	Negativa	0,89	0,89	0,921	
Isoflavona redutase		0,655	0,872	0,221		0,872	0,872	0,25	
NPR1	-	-	-	-		0,967	0,89	0,554	
RGA 37		1,49	0,8371	0,046	Positiva	0,771	0,8371	0,305	
RGA 02		0,998	0,8933	0,808		0,603	0,8933	0,068	
RGA 43		1,485	0,8636	0,046	Positiva	1,129	0,8636	0,893	
RGA 44		0,984	0,8764	0,744		0,758	0,8764	0,337	
RGA M09A31		0,907	1	0	Negativa	1,638	1	0,032	Positiva
RGA M09A21		0,747	1	0,035	Negativa	2,109	1	0	Positiva
RGA M09A11		1,074	1	0,651		0,418	1	0,142	
RGA B03A51		0,713	1	0,072		1,651	1	0	Positiva
RGA B03A11		1,048	1	0,573		1,881	1	0,032	Positiva

H1 é aceito quando $p < 0,05$. H1 (grupo de amostras inoculadas são diferentes de amostras do grupo controle). Regulação negativa é entendida aqui por *down regulated* e Regulação positiva é entendida por *up regulated*. Eficiência de reação aceita de 0,8 a 1,1.

A maioria dos genes envolvidos em processos de defesa não foram modulados. Os mesmos se mantiveram constantes em níveis de expressão no genótipo resistente e suscetível, mesmo depois de serem submetidos ao inóculo com o patógeno. Em geral, quando o agente patogênico ataca o hospedeiro, o passo inicial seria o reconhecimento de efetores de estirpes específicas de agentes patogênicos com a atuação de proteínas de resistência, que conduzem à

ativação de redes de sinalização de defesa em estágios mais tardios (Swarupa et al., 2013). Dessa forma, seria importante testar esses mesmos genes em estágios posteriores a 3, 6 e 9 dias após o inóculo para verificação de modulação de expressão dos mesmos.

A corismato sintase participa da via do ácido chiquímico, via esta que constitui um excelente alvo para o desenho de novos agentes antimicrobianos. A produção de compostos fenólicos através da indução das enzimas da via fenilpropanóide chiquimato e do metabolismo de flavonóides é uma resposta geral do estresse (Weaver e Herrmann, 1997). Apesar da corismato sintase participar da via que produz fitoalexinas, compostos estes sintetizados e acumulados em plantas depois da exposição a microorganismos (Hayashi et al., 2008), o gene codificando a mesma apresentou uma pequena regulação negativa no genótipo resistente Calcutta 4 após o inóculo com o patógeno. É possível que este patógeno não seja o indutor específico deste gene, como ocorre em tomate. Existem dois genes da corismato sintase em tomate, designados LeCS1, e LeCS2 e apesar de trabalhos mostrarem a indução do gene LeCS1 após a infecção com patógenos, para LeCS2 a indução não ocorreu pelo elicitor fúngico *Phytophthora infestans* em culturas de células e plantas intactas de tomate (Macheroux et al., 1999). Em estudos de expressão da corismato sintase em *A. thaliana* inoculadas com *Pseudomonas syringae*, Scheideler e colaboradores (2002) identificaram regulação positiva desta enzima apenas nos estágios iniciais da doença (7 horas após o inóculo), observando uma queda nos níveis de transcritos nos estágios subsequentes (24 horas após o inóculo).

A proteína NPR1 foi expressa somente no genótipo suscetível Cavendish Grande Naine, sendo que a mesma não foi modulada após o inóculo com o patógeno. Nas plantas, SAR é estabelecida como resultado da regulação promovida por NPR1 na expressão de proteínas PR (relacionadas à patogênese). É possível que a produção de SAR nem tenha chegado a ocorrer no genótipo resistente devido ao desencadeamento da imunidade disparada por PAMPs, onde os PAMPs são reconhecidos por receptores PRRs logo no início do processo, bloqueando a entrada do patógeno, mantendo os níveis de AS normais e sem a necessidade da atividade de NPR1 para produção de proteínas PR. Além disso, sabe-se que é necessária a ligação de um complexo de proteínas WRKY induzidas por AS ao promotor de NPR1 para que o mesmo seja expresso (Shah, 2003). O processo de regulação ainda é desconhecido e muitos fatores estão envolvidos. Segundo Shekhawat e Ganapathi, (2013), plantas que superexpressam OsWRKY71 (homólogo próximo de MusaWRKY71) mostraram expressão aumentada de NPR1 e PR1 em *O. sativa*, enquanto que em *M. acuminata* de nove genes NPR1 testados, nenhum obteve expressão diferencial em linhagens transgênicas com

genes WRKY superexpressos. Assim, parece que, mesmo entre espécies monocotiledôneas relacionadas, como arroz e banana, a regulação das vias de sinalização de defesa que envolvem genes WRKY e NPR1 pode ser bastante divergente.

O gene NAC demonstrou regulação negativa no genótipo suscetível Cavendish Grande Naine, sendo que no genótipo resistente não houve modulação significativa. O genoma de *M. acuminata* DH Pahang possui 167 genes NAC potencialmente funcionais (Cenci et al., 2014), com regiões altamente conservadas. Segundo Shan e colaboradores (2012), o alinhamento de seis proteínas NACs de *M. acuminata* mostrou que as mesmas partilhavam um N-terminal conservado, denominado o domínio NAC e um C terminal variável, passível de regulação transcricional, que é capaz de ativar ou reprimir a transcrição de múltiplas genes-alvo. Dessa forma, a estratégia de desenho de *primers* para regiões UTR foi altamente satisfatória, garantindo alta especificidade nas reações de PCR em tempo real.

Os genes NAC codificam para um grupo de fator de transcrição (TFs) planta-específicos e têm sido implicados na regulação de diversos processos, incluindo a sinalização mediada por hormônios, como por exemplo ácido salicílico, defesa e tolerância ao stress, como na resposta a estresses bióticos e abióticos, tais como infecções por fungos, seca, frio e salinidade elevada. A superexpressão de OsNAC6 e OsNAC52 em plantas transgênicas de *O. sativa* conferiu um aumento a tolerância ao estresse de desidratação (Lin et al., 2013; Shan et al, 2012). Sabe-se que essa família gênica está relacionada a resistência e corroborando aos resultados de repressão de NAC no genótipo suscetível, Nakashima e colaboradores (2007) e Lin e colaboradores (2007) mostraram que genes OsNAC6 e OsNAC19 foram induzidos em arroz frente ao desafio com o fungo *Magnaporthe grisea*, e a superexpressão de OsNAC6 levou a um aumento da resistência à brusone. Em batata, os genes foram induzidos em folhas após o inóculo com *Phytophthora infestans* (Collinge e Boller, 2001) e os genes BnNAC1-1, BnNAC5-1 e BnNAC5-7 foram induzidos em colza (*Brassica napus*) durante a colonização do Besouro saltador (*Exartematopus coccineus*) e a infecção pelo fungo *Sclerotinia sclerotiorum* (Hegedus et al., 2003).

Os RGAs 37, 43, M09A31, M09A21, B03A51 e B03A11 tiveram um aumento de expressão quando inoculados com o patógeno. Os RGAs 43 e RGA 37 mostraram indução significativa em *M. acuminata* Calcutta 4 quando inoculada com o patógeno (Figura 29). Indução de RGAs também foi encontrada por Wang e colaboradores (2006) durante a interação entre cultivares de batata (Russet Burbank, RB e Kennebec, KB) e *Phytophthora infestans*, e as diferenças de níveis de transcrição foram correlacionados com a reação de suscetibilidade/resistência da planta hospedeira em relação ao patógeno; Gao e colaboradores

(2006) constataram que alguns RGAs em algodão foram expressos apenas após a planta ser desafiada com o patógeno; Gao e colaboradores (2010) mostraram que os níveis de transcrição de quatro RGAs NBS presentes em tabaco foram regulados positivamente após o inóculo com *P. parasítica* var. *nicotianae* e Mohr e colaboradores (2010) identificaram em *Arabidopsis* cinco genes NBS-LRR sendo induzidos por *Alternaria brassicicola*, 14 por *Agrobacterium tumefaciens*, seis por *Botrytis cinerea*, nove por *Erysiphe chicoracearum*, 16 por *Heterodera schachtii*, seis por *Phytophthora infestans*, 28 por *Pseudomonas syringae*.

Os pares de *primers* específicos para os RGAs 37 e 43 foram derivados de ampliações de DNA genômico de *M. acuminata* Calcutta 4 (Miller *et al.*, 2008) utilizando *primers* degenerados direcionados para as sequências dos motivos proteicos conservados “P-loop” e “GLPL” em genes R da família NBS-LRR. O motivo P-loop no NBS é necessário para a ligação de nucleotídeos, e mutações neste motivo é a causa da perda de função da maioria das proteínas NBS-LRR (Marone *et al.*, 2013). Esses mesmos RGAs foram avaliados como marcadores moleculares RFLP, com múltiplos locos polimórficos observados nos parentais Borneo e Pisang Lilin, utilizados no desenvolvimento de um mapa genético de *M. acuminata* spp. *microcarpa* (Hippolyte *et al.*, 2010). O RGA37 foi ainda utilizado para selecionar clones BAC contendo NBS-LRR RGAs em *M. acuminata* Calcutta 4 (AA), *M. balbisiana* Pisang Klutuk Wulung (PKW) (BB) e o triplóide comercial *M. acuminata* Cavendish Grande Naine (AAA), identificando um total de 43 clones positivos contendo ortólogos de RGA37. Estes clones foram re-hibridizados com a sonda de RGA37, para verificar coordenadas positivas de clones BAC e para fornecer dados sobre o número de cópias de sequências NBS-LRR nos três genomas de *Musa*, encontrando três clones positivos contendo ortólogos em *M. balbisiana* Pisang Klutuk Wulung, 25 em *M. acuminata* Cavendish Grande Naine e nove em *M. acuminata* Calcutta 4 (Miller *et al.*, 2008).

Os RGAs 9330, 7100, 6160, M09A31 e M09A21 apresentaram regulação negativa em Calcutta 4. Da mesma forma, Kang e colaboradores (2012) mostraram dois genes NBS-LRR Glyma06g40740 e Glyma06g40950 como sendo regulados negativamente em soja após o inóculo com *Xanthomonas axonopodis* pv. *glycines*.

É possível que a evolução tenha levado os genes R para serem expressos em níveis muito baixos (basais), preferencialmente abaixo dos limites para efeitos colaterais deletérios. A regulação positiva induzida por patógeno poderia elevar transitoriamente a abundância de proteínas R acima do limiar necessário para uma resposta de resistência rápida e robusta. A importância da regulação positiva pode mudar ao longo da vida evolutiva de qualquer gene, onde a seleção natural pode favorecer a regulação positiva no caso de genes R que

desencadeiam resistência por meio de ligação fraca ao ligando, ou, essa regulação pode se tornar desnecessária quando o gene R evolui para sua forma otimizada (forte ligação ao ligando). Por outro lado, a seleção natural pode agir de maneira a reduzir a regulação positiva ou até mesmo a expressão basal se a proteína R evolui para atividade de fundo deletério. Isto pode explicar por que alguns genes R são regulados em diferentes graus durante a indução de defesa, enquanto outros são não responsivos aos patógenos (Mohr et al., 2010).

Em suma, esta análise é consistente com a tendência aparente de caracterizações anteriores de genes R individuais: alguns genes R são induzidos por uma variedade de estímulos, alguns exibem respostas muito específicas e outros parecem não ser responsivos ao patógeno. Conclusões semelhantes foram encontradas por Mohr e colaboradores (2010) onde dos 124 genes NBS-LRR de *Arabidopsis*, 49 não foram induzidos. Em um exame global de genes NB-LRR de *Arabidopsis* usando dados de microarranjo, Tan e colaboradores (2007) relataram que a expressão de dois genes NBS-LRR (At1g65400 e At1g56510) foram regulados negativamente, enquanto que dois outros genes (At1g72940 e At1g72920) foram induzidos após inóculo com *Pseudomonas syringae*.

Os demais RGAs não apresentaram diferenças significativas nos níveis de expressão já que os valores de significância foram superiores a $P=0,05$ (Tabela 20), demonstrando que as diferenças visíveis na figura 1 são decorrentes de flutuações biológicas normais. Backiyarani *et al.* (2013) também encontrou expressão constitutiva de RGAs em cultivares de *Musa* resistentes (Karthobiumtham) e suscetíveis (Nendran) ao nematóide *Pratylenchus coffeae*. Expressão constitutiva também foi encontrada por Peraza-Echeverria *et al.* (2008) afirmando que pelo menos três dos cinco RGAs de bananeira foram expressos constitutivamente em tecidos de raízes e folhas de cultivares resistentes e suscetíveis. A mesma tendência é encontrada em outras espécies, como é o caso de trigo, onde Gong e colaboradores (2013) também mostrou não haver diferenças na expressão entre amostras de RGAs NBS-LRR. Estes resultados são corroborados por Tan *et al.* (2007) que analisou os padrões de expressão de aproximadamente 170 NBS-LRR em *Arabidopsis* utilizando várias abordagens analíticas, dentre elas microarranjo e PCR, detectando que em geral os padrões de expressão desses genes ocorre normalmente em baixos níveis e não é induzida, ou seja, não há uma mudança significativa nesses níveis por meio de tratamentos com sinais de defesa.

Embora ainda não se saiba exatamente como as proteínas R agem durante a resposta de defesa, está claro que elas agem em um passo inicial das vias de sinalização de defesa, e mesmo em baixos níveis e expressão constitutiva possuem a capacidade para reconhecer o

patógeno e induzir respostas de defesa. A falta de um sistema circulatório em plantas pode ser compensada por uma capacidade para expressar cada uma das proteínas R constitutivamente em todas as células que poderiam potencialmente ser atacadas (Van der Biezen e Jones, 1998).

Em muitos casos, um gene R único pode proporcionar completa resistência a um ou mais estirpes de patógeno em particular, quando transferido para uma planta previamente susceptível das mesmas espécies. Da mesma forma, a coordenação da regulação positiva de genes R pode ser importante se vários genes são necessários para a resistência. No loco Rpp2 de *Arabidopsis*, dois genes R localizados em tandem funcionam de forma interdependente, e ambos são necessários para iniciar a resistência contra *Hyaloperonospora arabidopsidis*. Por esta razão, os genes de R têm sido utilizados em programas de melhoramento de resistência durante décadas (McDowell e Woffenden, 2003; Eckardt, 2007).

CAPÍTULO 3: DESENVOLVIMENTO E CARACTERIZAÇÃO DE MARCADORES SSR PARA GENES ENVOLVIDOS EM RESPOSTA A ESTRESSES BIÓTICOS EM GENÓTIPOS DE *M. ACUMINATA*

1. Introdução

Sequências derivadas de transcritomas de plantas representam um recurso valioso para a genética de plantas e programas de melhoramento. SSRs podem ser detectados em sequências de genes e os polimorfismos podem ser detectados utilizando *primers* complementares às sequências conservadas que flanqueiam essas repetições (Holland *et al.*, 2001). Esses marcadores moleculares são ferramentas úteis para ensaios de variação genética e fornecem um meio eficiente para ligar variação fenotípica e genotípica (Kalia *et al.*, 2011).

A engenharia genética e da biotecnologia tem grande potencial para o melhoramento de plantas, uma vez que diminui o tempo necessário para produzir variedades de culturas com caracteres desejáveis. Com a utilização de técnicas de biologia molecular é possível acelerar a transferência de genes desejáveis entre variedades e introduzir novos genes de espécies selvagens relacionadas. Caracteres poligênicos que antes eram muito difíceis de analisar por métodos tradicionais de melhoramento de plantas, agora é facilmente analisados utilizando marcadores moleculares (Mohan *et al.*, 1997).

Os microssatélites podem utilizados para diferentes fins, como por exemplo genotipagem, evolução e taxonomia, saturação de mapas genéticos, seleção assistida por marcadores (SAM) co-localizados com QTLs de resistência. Até o momento muitos marcadores foram desenvolvidos a partir de *M. acuminata* e *M. balbisiana*, contudo, em comparação com outras espécies cultivadas, o número total disponível para as análises genéticas continua a ser limitada, uma vez que os alelos podem estar ausentes ou possuírem características monomórficas, quando testados nos diferentes cultivares (Miller *et al.*, 2010).

Neste capítulo foi realizado o desenvolvimento de marcadores microssatélites dentro ou próximos a genes potencialmente envolvidos em respostas de resistência e defesa em sequências BAC de *M. acuminata* Calcutta 4 e Cavendish Grande Naine, no genoma de *M. acuminata* DH Pahang e em transcritos de *M. acuminata* Calcutta 4 (sequenciamento 454) mapeados no genoma de *M. acuminata* DH Pahang. Esses marcadores foram validados em 20 genótipos diplóides parentais (AA) de *M. acuminata* para análises de polimorfismos.

2. Material e Métodos

2.1. Material Biológico

2.1.1. Material Genético

Para a caracterização dos locos SSR (amplificação e análises de polimorfismos) foram usados 20 genótipos diplóides parentais (AA) de *M. acuminata* (Tabela 8) produzidos pelo programa de melhoramento da EMBRAPA Mandioca e Fruticultura contrastando em resistência as doenças Sigatokas negra e amarela, além de dois indivíduos (03115) derivados do cruzamento de *M. acuminata* Calcutta 4 e Pisang Berlin (progênie). Este conjunto de genótipos compreende uma amostra representativa dos genótipos utilizados em programas de melhoramento pois são utilizados em cruzamentos de pais selecionados para características desejadas e que apresentam gametas férteis para obtenção de melhores híbridos diplóides (Dantas *et al.*, 1997).

Tabela 11: Acessos de diplóides (AA) de *M. acuminata* contrastantes em resistência à doenças causadas por *Mycosphaerella* selecionados para uso em validação de marcadores moleculares SSR. Fonte: adaptado de Passos *et al.*, 2012.

Acessos <i>M. acuminata</i>	Suscetível a Sigatoka negra	Suscetível a Sigatoka amarela
Calcutta 4	Resistente	Resistente
Lidi	Resistente	Resistente
0323-03	Resistente	Resistente
SH32-63	Resistente	Suscetível
1304-06	Resistente	Resistente
0116-01	Resistente	Resistente
Burmanica	Resistente	Resistente
Microcarpa	Resistente	Resistente
1741-01	Nd	Resistente
9179-03	Nd	Resistente
1318-01	Nd	Resistente
4279-06	Nd	Resistente
Pisang Berlin	Parcialmente resistente	Suscetível
Niyarma Yik	Suscetível	Suscetível
Raja Uter	Nd	Suscetível
Tjau Lagada	Suscetível	Suscetível
F2P2	Nd	Suscetível
Khai Nai On	Suscetível	Suscetível
Sowmuk	Resistente	Suscetível
Jaribuaya	Resistente	Suscetível

Nd: não definido

O DNA genômico foi extraído dos 20 genótipos e agrupado em quatro *bulks*, de acordo com a resistência, como apresentado na Tabela 12. A estratégia de agrupar o DNA dos genótipos em *bulks* objetivou uma rápida avaliação de possíveis polimorfismos dos locos SSR em *M. acuminata*.

Tabela 12: Agrupamento dos genótipos em *bulks* contrastantes em resistência às Sigatocas negra e amarela.

Resistente à Sigatoca negra <i>Bulk</i> 1	Suscetível à Sigatoca negra <i>Bulk</i> 2	Resistente à Sigatoca amarela <i>Bulk</i> 3	Suscetível à Sigatoca amarela <i>Bulk</i> 4
Calcuttá 4	Pisang Berlin	Calcuttá 4	Raja Uter
Lidi	Niyarma Yik	Burmannica	Tjau Lagada
323-03		Microcarpa	F2P2
SH32-63		Lidi	Khai Nai On
1304-06		0323-03	Pisang Berlin
0116-01		1304-06	Niyarma Yik
		1741-01	Sowmuk
		9179-03	Jaribuaya
		0116-01	SH32-63
		318-01	
		4279-06	

2.2. Caracterização e desenho de primers para SSRs localizados em Genes Análogos de Resistência (RGAs)

2.2.1. Identificação de marcadores moleculares SSR *in silico* em sequências BAC

Sequências BAC contendo genes análogos de resistência (RGAs) (Pfam id: NB-ARC) derivados de *M. acuminata* Calcutta 4 e *M. acuminata* Cavendish Grande Naine foram identificados anteriormente (Miller *et al.*, 2008) e usados como arquivo de entrada para a busca de microssatélites. Para a predição e desenvolvimento dos marcadores moleculares SSR foi utilizado o programa WebSat (Martins *et al.*, 2009), responsável pela geração de todas as posições dos microssatélites (Figura 18). Este arquivo serviu como entrada para o programa Primer3 (Rozen e Skaletsky, 2000) no qual executou o desenho de pares de *primers* flanqueando a região determinada. O programa SSRIT - *Simple Sequence Repeat Identification Tool* (Temnykh *et al.*, 2001) foi ainda utilizado para corroborar e confirmar a presença dos microssatélites.

Os *primers* desenhados foram organizados em oito arquivos diferentes de acordo com a sequência BAC de origem e foram testados usando o programa Primer_Match (<http://edwardslab.bmcg.georgetown.edu>), um conjunto de ferramentas projetadas para

encontrar e contar correspondências exatas ou próximas entre as sequências dos *primers* gerados e grandes bancos de sequências, neste caso, as sequências BAC. O programa gera scores de zero a seis de acordo com os alinhamentos, onde o score zero corresponde a nenhuma inserção entre os pares de *primers* e a sequência do BAC, ou seja, de alinhamento de 100% das bases e o score máximo de seis, gerado a partir de três inserções derivadas do *primer forward* e três derivadas do *primer reverse*. O arquivo de saída .bed foi criado e para cada BAC foi gerado um arquivo gff com a anotação predita.

Para fazer a intersecção entre o arquivo “.bed” e o arquivo “.gff”, ou seja, saber exatamente a localização e a anotação onde cada par de *primer* está inserido, foi utilizado o pacote de programas BEDtools (Quinlan e Hall, 2010), desenvolvendo um *pipeline* capaz de ler ambos arquivos e executar as intersecções. Com base nas anotações, os *primers* localizados dentro de genes com Pfam ID NB-ARC foram selecionados.

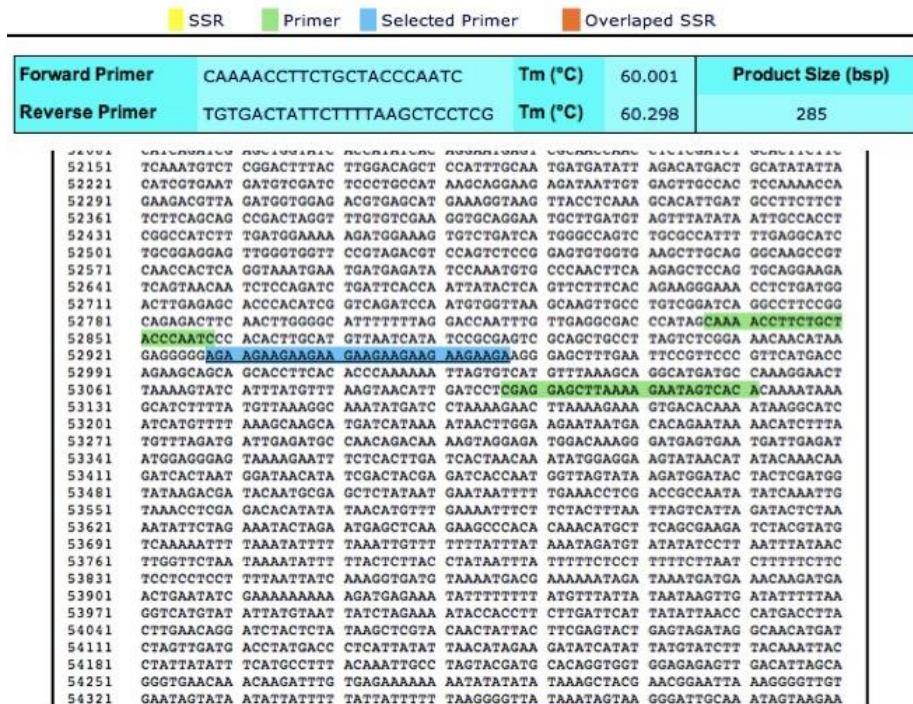


Figura 18: Arquivo de saída do programa *websat* mostrando dados dos *primers forward* e *reverse* desenhados para o microsatélite (AGA)₁₀ contido na sequência BAC MaC139M12, dentro de um gene codificador da proteína da família NB-ARC. São mostrados em azul o SSR encontrado dentro da sequência alvo e em verde a posição dos pares de *primers* flanqueando o microsatélite. É mostrado a Tm em °C, as sequências 5'-3' dos *primers* e o tamanho esperado do fragmento amplificado.

2.2.2. Identificação de marcadores moleculares SSR *in silico* no genoma de *M. acuminata* DH Pahang

Genes potencialmente envolvidos em resistência foram identificados no genoma de *M. acuminata* DH Pahang com base em dados de domínios conservados de Pfam ID NB-ARC por meio do InterPro Domain Search (<http://banana-genome.cirad.fr/advanced>). As coordenadas dos 102 locos encontrados foram utilizadas para a delimitação na busca de SSRs dentro ou próximos desses genes de interesse. Com base nessa informação, foi estipulado um valor de 10000 pb antes e depois do gene para a busca dos microssatélites. Os programas WebSat (Martins *et al.*, 2009) e SSRIT (Temnykh *et al.*, 2001) foram utilizados para a identificação dos SSRs. Os arquivos gerados serviram como entrada para o programa Primer3 (Rozen e Skaletsky, 2000) no qual executou o desenho de pares de *primers* flanqueando a região determinada, obedecendo os parâmetros de tamanho do *primer*, de 18 pb a 22 pb; Tm do *primer* de 57°C a 68°C; porcentagem de GC do *primer*, de 40% a 80% e tamanho do produto de 100 pb a 400 pb.

2.2.3. Identificação de marcadores moleculares SSR *in silico* derivados de transcritos de *M. acuminata* Calcutta 4 (sequenciamento 454) mapeados no genoma de *M. acuminata* DH Pahang

A busca computacional de microssatélites no transcrito de *M. acuminata* Calcutta 4 gerado por pirosequenciamento se deu por meio do programa Mreps, versão 2.5 (Kolpakov, *et al.*, 2003) detectando a presença de pelo menos duas unidades de repetição, com no mínimo dez repetições por unidade. Os *primers* foram desenhados conforme demonstrado em trabalho anterior (Passos *et al.*, 2013).

Para verificar se os SSRs identificados são encontrados também no genoma de *M. acuminata* DH Pahang, foi utilizado o programa PrimerMatch (<http://edwardslab.bmc.b.gorgetown.edu>). Para ampliar a análise, foi deixada uma janela de 500000 pb a jusante/a montante ao SSR e para visualizar os genes circundantes aos SSRs utilizou-se a ferramenta BEDtools (Quinlan e Hall, 2010), comando *intersect* para fazer a intersecção entre as sequências dos SSRs e a anotação do genoma, selecionando apenas aqueles pertencentes à família de genes de resistência com Pfam ID NB-ARC.

2.2.4. Identificação de marcadores moleculares SSR *in silico*, derivados de transcritos de *M. acuminata* Calcutta 4 (sequenciamento 454) presentes em genes de resistência de diversas famílias.

Os microssatélites foram identificados e desenhados anteriormente (Passos *et al.*, 2013) e aqueles localizados em genes potencialmente envolvidos em resistência, pertencentes às diversas famílias de Genes R foram selecionados com base em similaridades BLASTX (Altschul *et al.*, 1997) e classificação KOG - *Clusters of Eukaryotic Orthologous Proteins from Complete Eukaryotic Genomes* (Tatusov *et al.*, 2003). Os mesmos foram classificados quanto aos motivos repetidos.

2.3. Caracterização e desenho de primers para SSRs localizados em Genes de Defesa

Os microssatélites foram identificados anteriormente (Passos *et al.*, 2013) e aqueles localizados em genes potencialmente envolvidos em respostas de defesa, foram selecionados com base em similaridades BLASTX (Altschul, *et al.*, 1997) e classificação KOG (Tatusov *et al.*, 2003).

2.4. Amplificação de microssatélites via PCR

Para a caracterização dos SSRs, foi avaliado um total de 156 locos. Estes foram amplificados utilizando-se *primers* específicos obtidos a partir da sequência correspondente à região flanqueadora de cada loco. Os SSRs foram amplificados através de reações de PCR nos 4 *bulks* que foram posteriormente individualizados em casos de polimorfismos.

As reações de amplificação foram conduzidas usando o termociclador Gene Amp® PCR System 9700 (Applied Biosystems) com volume final de 13 µL, contendo 3 ng de DNA genômico, tampão PCR 1X (100 mM Tris-HCl, pH 8,4 e 500 mM KCl), 2,5 mM de dNTP, 2,5 mM de de BSA (Albumina de Soro Bovino); 1,5 mM de MgCl₂, 0,9 µM de cada *primer* (*forward e reverse*) e 1,25 U Taq DNA polimerase. As amplificações foram realizadas nas seguintes condições: 94° C por 5 minutos; seguido de 30 ciclos de 94° C por 1 minuto; temperatura de anelamento específica para cada par de *primer* por 1 minuto (52-60°C), 72°C por 1 minuto, e extensão final a 72°C por 7 minutos.

2.5. Eletroforese e Visualização dos locos

Inicialmente os produtos amplificados foram submetidos a eletroforese em gel de agarose 2% corado com brometo de etídio (50ng/L) por 20 minutos a 90 Volts apenas para confirmação da amplificação eficiente do loco com base no tamanho esperado do fragmento, comparado com o marcador de peso molecular 1Kb ladder (Invitrogen). Os resultados foram visualizados sob luz ultravioleta e as imagens foram capturadas por meio do Gel logic 200 imaging system (KODAK®, EUA). Os ajustes e otimizações foram realizados de acordo com esse resultado, aumentando ou diminuindo a temperatura de anelamento (variação de 50°C – 58°C), bem como a concentração de MgCl₂. Os locos que obtiveram êxito nesta primeira etapa foram posteriormente analisados em gel denaturante de poliacrilamida 5% utilizando ureia 7M (Creste *et al.*, 2001; Promega, 2006) e visualizados com coloração de nitrato de prata. A coloração seguiu três passos: fixação do gel com uma solução de ácido acético glacial (Merck) 10% por 10 minutos; coloração (oxidação) do gel por uma solução de AgNO₃ acrescido de formaldeído 37 % (2g de AgNO₃ em 2 L de água milli-Q) por 3 minutos; revelação do gel com uma solução de Na₂CO₃ gelado (60 g de Na₂CO₃ em 2 L de água milli-Q com 3 mL de formaldeído 37% e 400 µL de Na₂O₃S₂ 10 mg/mL) por aproximadamente 10 minutos ou até que as bandas pudessem ser visualizadas. Os tamanhos dos locos foram comparados aos marcadores de peso molecular *ladder* 10 pb e 50 pb.

Aqueles locos que apresentaram polimorfismos na análise dos *bulks* foram novamente submetidos a PCR e visualizados em gel de nitrato de prata com os DNAs dos *bulks* individualizados para a identificação dos alelos em cada genótipo utilizando as progênes planta 1 e planta 2 como controle.

Após a revelação, os géis foram expostos a temperatura ambiente para secar por aproximadamente 24h, e depois foram escaneados e scoreados no computador. Os resultados foram analisados no programa *PowerMarker* v3.25 (Liu e Muse, 2005) visando identificar o número de locos polimórficos, média de alelos por loco, heterozigosidade esperada e observada e o PIC (*Polymorphic Information Content*).

3. Resultados e Discussão

Um total de 195 pares de *primers* SSR foram desenvolvidos para genes potencialmente envolvidos em respostas de resistência e defesa, conforme o resumo na Figura 19.

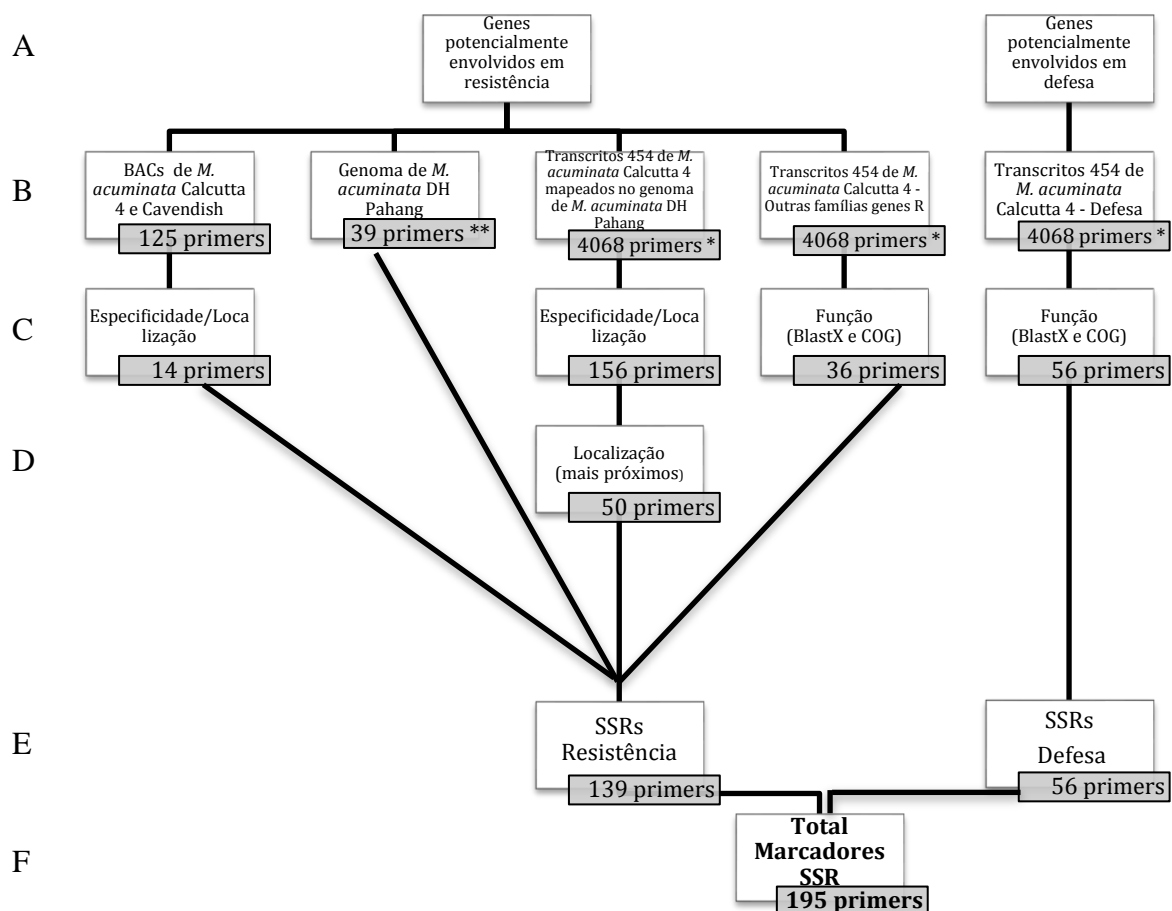


Figura 19: Fluxograma demonstrando o desenvolvimento dos *primers* para SSRs em locos potencialmente envolvidos em resposta aos estresses bióticos. A) Pares de *primers* para SSRs foram desenvolvidos a partir de genes potencialmente envolvidos em respostas de resistência e defesa. B) Número de *primers* SSR desenvolvidos a partir de diferentes fontes. C) Mineração de *primers* por meio de análises de especificidade e localização no caso dos BACs.; em transcritos 454 de *M. acuminata* Calcutta 4 mapeados no genoma de *M. acuminata* DH Pahang foi realizado PCR eletrônico contra o genoma de DH Pahang e apenas os que anelaram próximos a regiões de genes NB-ARC foram selecionados; resultados BlastX e COG foram utilizados para mineração *primers* derivados de transcritos (454) de outras famílias de genes R e genes de defesa. D) Dos 156 *primers* identificados, apenas os 50 mais próximos à genes NB-ARC foram utilizados. E) Número total de pares de *primers* desenvolvidos para SSR potencialmente envolvidos em respostas de resistência e defesa. F) Número total de pares de *primers* SSR. **Primers* desenvolvidos em trabalho recente (Passos *et al.*, 2013). Os mesmos foram minerados conforme objetivo de buscar genes de resistência próximos a NB-ARC no genoma de *M. acuminata* DH Pahang; buscar genes R pertencentes à diferentes famílias e genes de defesa. ***Primers* desenvolvidos que serão utilizados apenas em trabalhos futuros.

3.1. Desenvolvimento de marcadores SSR para genes potencialmente envolvidos em resistência (RGAs)

3.1.1. Desenvolvimento de marcadores moleculares SSR *in silico* a partir de sequências BAC de *M. acuminata* Calcutta 4 e Cavendish Grande Naine

Foram desenvolvidos 125 pares de *primers* SSR flanqueando ou dentro de regiões gênicas NBS-LRR (Anexo 3) de maneira que os SSRs aparecem distribuídos nas 8 sequências de clones BAC de Calcutta 4 e Cavendish (Figura 20). O número extenso de RGAs dentro dos BACs vem corroborar com Miller e colaboradores, 2008 que identificaram 20 clones BAC de C4 e 43 clones BAC de Cavendish contendo genes ortólogos e parálogos de dois RGAs pertencentes à família NB-ARC. Sabendo que os genes R estão frequentemente localizados em *clusters* ou famílias gênicas por todo o genoma com numerosas sequências homólogas, a estratégia adotada para busca de SSRs em RGAs presentes em BACs obteve êxito.

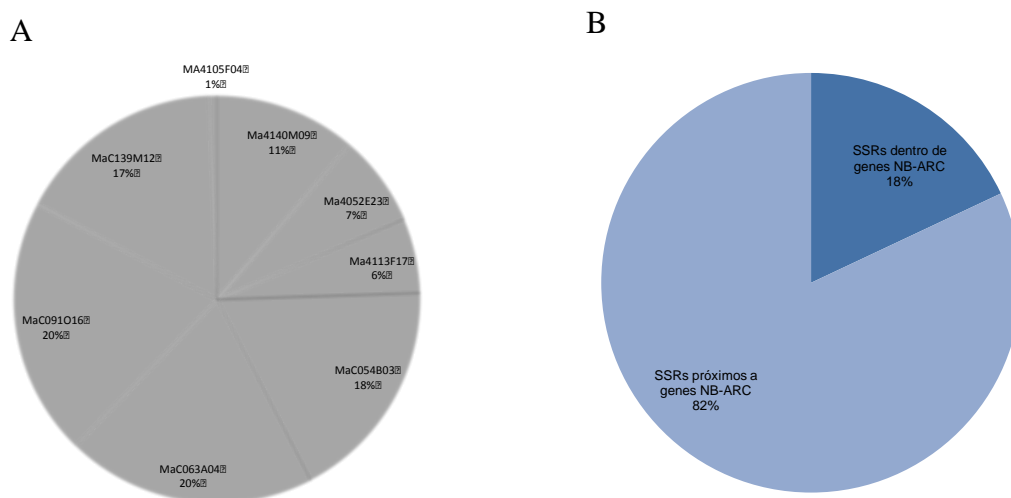


Figura 20: Distribuição de SSRs localizados próximo ou dentro de RGAs (família NB-ARC) de sequências de clones BAC de *M. acuminata* Calcutta 4 e Cavendish Grande Naine. A – porcentagem de SSRs presentes em cada sequência de clone BAC. B- Número de SSRs dentro ou próximos a RGAs (família NB-ARC)

Apesar da maioria dos marcadores microssatélites serem encontrados próximos a RGAs, os *primers* selecionados para a avaliação de polimorfismo foram aqueles em que o SSR está localizado dentro do gene de interesse, eliminando assim a possibilidade de não segregação do marcador microssatélite com o gene da família NB-ARC.

O PCR eletrônico entre os *primers* e as sequências de clones BAC possibilitou o conhecimento do comportamento dos *primers in silico*, garantindo uma seleção mais acurada dos *primers* em questão. Com base nisso, apenas os *primers* que apresentaram alta especificidade, ou seja, apenas aqueles que anelaram em somente um loco no BAC foram utilizados. Outro critério foi a atribuição de *scores* para o número de *mismatches*, onde somente os *primers* que obtiveram um pareamento de 100% de suas bases com o DNA alvo (*score* 1) foram selecionados. E por fim, os SSRs que estavam dentro dos genes de interesse foram selecionados com base na posição do loco e na anotação do BAC. Dessa forma, um subgrupo de 14 marcadores SSR foi selecionado para validação em relação o polimorfismo em genótipos de *M. acuminata* (Tabela 13).

Tabela 13: Subgrupo selecionado de 14 marcadores SSR, derivados de genes NB-ARC presentes nas bibliotecas BAC de *M. acuminata* Calcutta 4 e Cavendish Grande Naine, validados para amplificação por PCR e polimorfismo nos 22 genótipos diplóides de *M. acuminata*.

Loco	Primer FW	TM	Primer RV	TM	Anotação de sequencia BAC
MaC139M12_loc110	CAAAACCTTCTGC TACCCAATC	60.0	TGTGACTATTCTTT AAGTCCTCG	60.2	Putative disease resistant protein
MaC139M12_loc112	AACTTGGGACAAT TATGGAGGA	59.7	CCAGTTTGGAGTTTC TTTTCGT	59.6	Disease resistance protein RGA
MaC139M12_loc115	AACCTCTGATGGA CTTGAGAACA	60.1	CCCAGTTTGGAGTTT CTAGGTG	60.0	NB-ARC domain containing protein
Ma4052E23_loc2	ACCAGAAGACGTT AGATGGTGG	60.4	GCTGATGCTAGTGCT GATGCTA	60.7	RGA1;RGA3;Disease resistance protein
Ma4052E23_loc3	AACCTCTGATGGA CTTGAGAACA	60.1	CCCAGTTTGGAGTTT CTAGGTG	60.0	RGA1; RGA3;Disease resistance protein
MaC063A04_loc47	AACCTCTGATGGA CTTGAGAACA	60.1	CCCAGTTTGGAGTTT CTAGGTG	60.0	NB-ARC domain
MaC054B03_loc35	TATAGGATCGACG TTTTGTTGC	59.1	GACAGAAGCAAATG GAGTGCTA	59.5	RGA1;Putative disease resistance protein
Ma4052E23_loc5	TCAAGAGGTATGC TACATTCAGGA	60.1	GAAAGATGACCGAG GCAATTA	59.1	Putative disease resistance protein;
MaC091O16_loc79	ATGGACTTGAGAG CACACACAT	59.6	CTGCTACTGGAAAT GATCGTGA	60.2	Putative disease resistance protein RGA3;Leucine-rich repeat
MaC091O16_loc82	TCAACCTCCTCTTT AGCCATGT	60.1	GCAGTATGACTTCC ACCTACC	60.1	Putative disease resistance protein; RGA3
MaC091O16_loc92	TGACGCCTTCTTCT TCTTCTTC	60.1	TCATTTACCTGAGTG GTTGACG	60.0	Disease resistance protein; RGA2
MaC091O16_loc94	GGTATGCTACATT CAGGAAACGA	60.3	CCTGCACCTTCGAGA CAAAT	60.2	Putative disease resistance protein; RGA3
MaC063A04_loc62	AGACTCGGAAACA ACAGAGGAG	59.9	CAACCTATGGTCAC ACGAAGAA	60.0	Leucine-rich repeat RGA4; NB-ARC
MaC063A04_loc75	AACTTGGGACAAT TATGGAGGA	59.7	CCAGTTTGGAGTTTC TTTTCGT	59.6	Leucine-rich repeat RGA4; NB-ARC

TM: Temperatura de melting (°C); Primer FW (Primer forward) ; Primer RV (Primer reverso).

3.1.2. Desenvolvimento de marcadores moleculares SSR *in silico* a partir de sequências do genoma de *M. acuminata* DH Pahang

Um total de 39 pares de *primers* foram desenhados com base na distância entre um microsatélite e um gene com Pfam ID de NB-ARC. Para restringir a análise, foi montado um *ranking* com os SSRs distando até no máximo 10.000 pb dos genes de interesse (Tabela 14).

Tabela 14: Informação das sequências de primers SSR derivados do genoma de *M. acuminata* DH Pahang. Dados representam todos os locos identificados distando em até 10.000pb de genes NB-ARC, acompanhados da informação dos motivos de repetição SSR e temperatura de anelamento.

Primer	SSR	Primer FW	TM	Primer RV	TM	Anotação genoma (NB-ARC)	Distância SSR/gene NB-ARC (pb)
PAHANG_RES_PAHANG_LOCUS1	(AG)9	GAATTTGATTATGTCCGCCACT	60.2	TAGATTCCGTTTCAGTCTCATGC	59.3	GSMUA_Achr6P19710_001	26
PAHANG_RES_PAHANG_LOCUS2	(CT)6	TTAGCCATGTCCTTCAACGTA	59.6	TGACTTGCACCTACCTTTGTCT	58.9	GSMUA_AchrUn_randomP19300_001	41
PAHANG_RES_PAHANG_LOCUS3	(CT)9	GTGAAGGAATCATCAAGCACAA	60.1	GCAAGAATAAGAATGGGTTTGG	59.8	GSMUA_AchrUn_randomP19270_001	42
PAHANG_RES_PAHANG_LOCUS4	(TC)21	GATGGCTGTGCTGAGTAGATTG	59.9	AACAAGTTAAAGGCCGTCTGAA	60.1	GSMUA_Achr4P26160_001	44
PAHANG_RES_PAHANG_LOCUS5	(TC)6	ACCGTATCAGGTCGAATCCTAA	59.8	GGAGCAAGTCGTCCCAGAT	60.2	GSMUA_AchrUn_randomP27050_001	112
PAHANG_RES_PAHANG_LOCUS6	(CT)7	ACATCAGGTAACATCAAACCCC	59.9	CTCGTCTCAGAAGGTGTTTCAGT	59.6	GSMUA_AchrUn_randomP19270_001	221
PAHANG_RES_PAHANG_LOCUS7	(TCA)12	CCCAAAGAGTTTTATTATCCTCCC	60.3	CAAGTTTGTATTGGAGATGGAGC	60.0	GSMUA_Achr9P02765_001	269
PAHANG_RES_PAHANG_LOCUS8	(CTT)10	GCAAAGATGACTGTGAGTTGC	59.9	CAAACCTAGTCGATGATGATGC	59.5	GSMUA_AchrUn_randomP22870_001	347
PAHANG_RES_PAHANG_LOCUS9	(AG)11	AGCTTCTTCTCATTGTCCTTCG	60.0	CTATTGGGATTGCTTGACCTTC	59.9	GSMUA_Achr6P32070_001	353
PAHANG_RES_PAHANG_LOCUS10	(TC)14	CTGTGTTGGTTTTTCCTTGAG	59.9	TCAACGGTGGCTAAACTCTTTT	60.1	GSMUA_Achr7P20220_001	477
PAHANG_RES_PAHANG_LOCUS11	(TC)31	ACGAACCGATGGGCTTAAATA	60.6	ACAATAGCGCAGTGACAAGAAA	59.9	GSMUA_Achr7P20230_001	498
PAHANG_RES_PAHANG_LOCUS12	(AAT)7	TGAAAATCGAGGAAGAGGAGTA	60.2	GGAAAATTACGGAGGGATGTTT	60.4	GSMUA_Achr10P07050_001	832
PAHANG_RES_PAHANG_LOCUS13	(GA)14	AAAGACGAAAAGAGAAGCAATCG	60.0	GAAAGAGAATAGATCGGGAGCA	59.8	GSMUA_Achr7P08250_001	970
PAHANG_RES_PAHANG_LOCUS14	(AG)10	CACACAGAGAGAGAGAGAGAGAGA	60.1	AACATCAAGGTAGGGAGGGG	60.1	GSMUA_Achr9P19800_001	1108
PAHANG_RES_PAHANG_LOCUS15	(AG)22	TGCTGCAAAATGGTCAGAGTT	59.7	GCATTCAAATTCACACAGAAGC	59.7	GSMUA_Achr9P19800_001	1436
PAHANG_RES_PAHANG_LOCUS16	(CT)8	ATTCTTTGAATGAGCCTGTGG	60.4	TTCTCTCTCCCTCTCCCTCT	60.0	GSMUA_Achr3P10930_001	1550
PAHANG_RES_PAHANG_LOCUS17	(TAT)10	AATCCCAACTCCAAATCACATC	60.0	AAGGAGGTTTCAGATTCCAACA	59.9	GSMUA_Achr7P08250_001	1649
PAHANG_RES_PAHANG_LOCUS18	(AG)17	GATGGGTGATGCTGTTGTCTAA	60.0	GGGTTTGTGAAAGTCTCTTTGG	60.0	GSMUA_Achr7P08250_001	2330
PAHANG_RES_PAHANG_LOCUS19	(CT)7	CAAGACAACAACACAAACACACAC	60.4	TGCTACAGCTAACACCATCGTT	59.8	GSMUA_Achr7P20230_001	2370
PAHANG_RES_PAHANG_LOCUS20	(AC)8	GCAACGATGGTGTAGCTGTAG	59.8	GGTCCACTTAGCATCAGAGCA	60.4	GSMUA_Achr7P20230_001	2523
PAHANG_RES_PAHANG_LOCUS21	(CTT)7	AAAGAATCGAGAAGGTGTTGGA	60.1	CTTAGACTCGGAAACAACGGAG	60.2	GSMUA_Achr3P27820_001	2530
PAHANG_RES_PAHANG_LOCUS22	(AG)11	TGACGACTGCTACTCCTGTCC	60.4	CTCGAACCATCAGCAATACAAG	59.7	GSMUA_Achr6P21780_001	4533
PAHANG_RES_PAHANG_LOCUS23	(ATT)8	GCCTACTTTCAAGACTAAACGC	57.4	GTCAATTCTAATCCCAGCTTCA	58.1	GSMUA_Achr1P14030_001	5314
PAHANG_RES_PAHANG_LOCUS24	(GT)10	TTCTGTCAAATTAAGGGCAGGT	60.0	TACGGCCAAGATGCAATAAGTA	59.6	GSMUA_Achr6P21780_001	5676
PAHANG_RES_PAHANG_LOCUS25	(TC)6	CATGAACCGATGTGGCTAAATA	59.8	TTCAATGGTCCTAAGTGGCTCT	60.1	GSMUA_Achr7P20230_001	7268
PAHANG_RES_PAHANG_LOCUS26	(CT)7	ACATCAGGTAACATCAAACCCC	59.9	TTCTAGCGCCAATCTATTACGG	60.5	GSMUA_AchrUn_randomP19270_001	9140
PAHANG_RES_PAHANG_LOCUS27	(GAA)6	TAGTCGATGATGATGCTGCTTT	59.8	TGTCTCTGTTACCACTGTTGGG	60.0	GSMUA_AchrUn_randomP28320_001	DENTRO
PAHANG_RES_PAHANG_LOCUS28	(GA)7	AAAAGTTCATGGTGGCTTCTTA	60.0	GGCTCTCATTTTCTTTGAATGC	60.2	GSMUA_AchrUn_randomP27810_001	DENTRO
PAHANG_RES_PAHANG_LOCUS29	(GAT)15	TTATATCGGCTAGAGAGGGCAG	59.8	TCTGTGCTGCTCCTTATTTCCT	60.1	GSMUA_Achr10P07100_001	DENTRO
PAHANG_RES_PAHANG_LOCUS30	(GAA)10	TAGTCGATGATGATGCTGCTTT	59.8	CAAGACATTTCACTGCCAACAT	60.0	GSMUA_Achr3P27790_001	DENTRO
PAHANG_RES_PAHANG_LOCUS31	(CTT)7	TCATCTGTTGTTTCGCTCAGTT	59.9	AAAACCTTCCGCTACCCAAT	59.8	GSMUA_AchrUn_randomP27050_001	DENTRO
PAHANG_RES_PAHANG_LOCUS32	(GAG)8	TTCTGAGCTTAAATCGTTGGT	60.1	AGGACAATCTTTGGTTTGCAGT	60.0	GSMUA_Achr1P14020_001	DENTRO
PAHANG_RES_PAHANG_LOCUS33	(TTC)13	TAATGGGTGCTGTGTTTCCTT	58.4	ACTACATCAAGCATTCCTGCAC	59.2	GSMUA_Achr3P28030_001	DENTRO
PAHANG_RES_PAHANG_LOCUS34	(AGA)11	GGAAACACGGAGGAGAAGTAG	59.2	AAAGAATGGAGAAGGTGTTGGA	59.9	GSMUA_Achr3P28030_001	DENTRO

PAHANG_RES_PAHANG_LOCUS35	(CGG)6	GTTATGGGGAAGAACATGAAGC	59.8	GTCCTCCTTGTGGTAGATCAGC	60.1	GSMUA_Achr9P02520_001	DENTRO
PAHANG_RES_PAHANG_LOCUS36	(CTT)7	AAAGAATCGAGAAGGTGTTGGA	60.1	CTTAGACTCGGAAACAACGGAG	60.2	GSMUA_Achr3P27830_001	DENTRO
PAHANG_RES_PAHANG_LOCUS37	(TTC)9	GCAAAAGATGATTGTGAGTTGC	59.7	GCACCTTCGAGACAAACCTAGT	59.8	GSMUA_AchrUn_randomP19290_001	DENTRO
PAHANG_RES_PAHANG_LOCUS38	(AGA)7	TTCTGCTAGTGGCGTATGGAAT	60.9	CGTCATCTTTTCTGCTCTGATG	60.0	GSMUA_Achr6P11450_001	DENTRO
PAHANG_RES_PAHANG_LOCUS39	(AGG)6	AGAAAGCGAGTAGACCACATCC	59.7	ACGAATAAACATTCTCGTCCGT	59.8	GSMUA_Achr8P30710_001	DENTRO

TM: Temperatura de *melting* (°C); *Primer FW* (*Primer forward*); *Primer RV* (*Primer reverso*).

3.1.3. Desenvolvimento de marcadores moleculares SSR *in silico* derivados de sequências de transcritos (sequenciamento 454) em *M. acuminata* Calcutta 4 mapeados no genoma de *M. acuminata* DH Pahang

Foram identificados um total de 152 SSRs de transcritos de *M. acuminata* Calcutta 4 flanqueando 34 genes NB-ARC no genoma do *M. acuminata* DH Pahang. Os genes que codificam as proteínas NBS-LRR são principalmente organizados em arranjos de sequências estreitamente relacionados, resultantes de eventos de duplicação em *tandem*. Em *M. acuminata* DH Pahang eles estão presentes em maior número nos cromossomos 3, 6, 9, 10 (D'Hont *et al.*, 2012). Foi observado neste estudo a predominância dos locos SSR nesses mesmos cromossomos, o que era realmente esperado já que a escolha do loco se deu com base na proximidade de genes pertencentes a esta família (Figura 21).

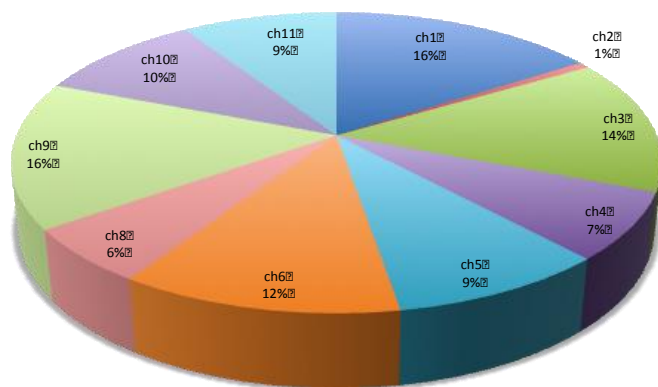


Figura 21: Distribuição de locos SSR derivados de transcritos NBS-LRR de *M. acuminata* Calcutta 4 mapeados ao longo do genoma de *M. acuminata* DH Pahang.

Um total de 50 SSRs localizados mais próximos dos genes NB-ARC e com valor de *score* superior a 0,5; ou seja, um máximo de três inserções para os dois *primers* durante PCR eletrônico contra a sequência genômica de *M. acuminata* DH Pahang, foram utilizados para análise de polimorfismo nos 22 genótipos diplóides de *M. acuminata* (Tabela 15). Informações adicionais do loco, tais como a anotação gênica para o SSR em *M. acuminata* DH Pahang, bem como a distância do microssatélite ao gene NB-ARC em *M. acuminata* DH Pahang, localização cromossômica, identificação do transcrito correspondente (anotação de dados do sequenciamento 454 de RNA da interação *M. musicola* e *M. acuminata* Calcutta 4) e número do *score* do *primer* podem ser visualizados no Anexo 4.

Tabela 15: Informações dos primers SSR desenvolvidos a partir de *Unigenes* derivados do transcrito de *M. acuminata* Calcutta 4 em interação com *M. musicola* mapeados no genoma de *M. acuminata* DH Pahang.

Loco	Motivos de repetição	Primer FW	TM	Primer RV	TM	Tamanho produto (pb)
454_RES_PAHANG_locus3299	ATG	GCTGCCGACCAAGTTTGTAT	60,1	ATTACTCTGCAAGCCCATCG	60,2	174
454_RES_PAHANG_locus1444	GTC	GTTTCAGCATGAGCAGAGTCG	59,7	AACACAAATGGCAAGGGAAG	60,0	144
454_RES_PAHANG_locus3316	GATA	GGTTTCCTCGTCACGTTCTC	59,7	GTTGAGAAAAGGATTGCCCA	60,1	202
454_RES_PAHANG_locus1458	CCT	ACCCAAACAGAGGCATAACG	60,0	CAGAAGAGGAAGAGGCATCG	60,1	297
454_RES_PAHANG_locus1835	CAC	ATGGGAATGCAAAGAACTCG	60,1	TCACATGTGCAGTGTTCG	59,3	242
454_RES_PAHANG_locus3952	GCT	CTCTTCCTCTGCCTTCGCTA	59,9	GGCGCAGAAAGAAAGACTCAC	60,1	302
454_RES_PAHANG_locus2300	CCG	CCCTAATCTAGCCGGACCTC	60,1	AAGGGGTGGAAGGATCTGAG	60,4	328
454_RES_PAHANG_locus407	TGC	GGCTCTACCTGAAGCGAATG	60,0	GCCTACACAAGCACCCAAAT	60,0	471
454_RES_PAHANG_locus2811	CCG	CTCGTCCAACCGAGAAGAAG	60,0	GACAAACGAGGGAAGAGGTG	59,7	262
454_RES_PAHANG_locus2810	AGA	TCGCTCCCTCTCATCTTGTT	60,0	CCTCGGAGATGCAGGAGTAG	60,0	185
454_RES_PAHANG_locus513	TTCA	TTTTTCTGTGGCCCTTTCAC	60,1	AGCCAATGGTGACATCATAAA	58,0	197
454_RES_PAHANG_locus1102	TC	GGTTAAACGCCCTCTTTCTC	57,9	GTACACCACCTTGATCGGCT	60,0	129
454_RES_PAHANG_locus2274	GAG	TCGATGATGATAAGGCCTCC	60,0	ATATCGGCGGTAAAAATCCC	60,0	433
454_RES_PAHANG_locus2273	CGA	CTTCGCTTTGCTTGTTCCTC	60,0	CTCCTCGCCATCTACTACGC	60,0	327
454_RES_PAHANG_locus1096	TATT	CGCGTCACCTGTTTCTTGTA	59,9	CTTCACACCATTGCCTTCCT	60,1	284
454_RES_PAHANG_locus3926	AT	ACGTATTCCCAACCCATCAA	60,1	CCCCAACGATAAAGGCATCT	59,9	151
454_RES_PAHANG_locus3698	GAA	TTTGTTCAGATAAATCGGGG	58,9	GTCATCGGAGTCATCGAGGT	60,1	253
454_RES_PAHANG_locus311	GGAG	TAAAAGTTGGATTTTGGGCG	59,9	GCATCCAGAGCGAGTAGAGG	60,1	189
454_RES_PAHANG_locus14	GCT	ATCATGCATCCTCTCCCAAG	60,0	TTCTCAACATTCACCAGCCA	60,2	148
454_RES_PAHANG_locus3990	GAT	TATTGGACTGATGGGGCATT	60,2	TCCTTGTGGAGGATGACTTTG	60,1	316
454_RES_PAHANG_locus2243	CTC	CACTCCACACCACCATCATC	59,8	TGGATCGTACAGGTAGCAGAG	58,0	132
454_RES_PAHANG_locus549	GAAAAA	CAGGGCTAAGGAGAATGCAG	60,0	TTTTACCACCGCAAGGATA	60,3	291
454_RES_PAHANG_locus3360	CTG	TATTTTATGGTGGCCCTGCAA	59,0	GCACCTAAAGGGCCATTACA	60,0	364
454_RES_PAHANG_locus1635	CCA	GAAGCCCACTGATCTTTCCTCA	60,2	CTCTTCCTTTGGTGAGGCAG	60,0	253
454_RES_PAHANG_locus3561	GCT	AAACACCCATATCGGGACAA	60,1	TCATGTCCTTGAATCCACCA	59,9	250
454_RES_PAHANG_locus3050	GA	CTACTCCAAAGACCCCGTCA	60,1	TGCAGGAGTGACAGGATGAG	60,0	261

454_RES_PAHANG_locus2334	CT	TGCCTCGAGGACCCTAATAA	59,7	CAAAGCCAAGTGCAGAAACA	60,0	388
454_RES_PAHANG_locus1633	GAG	TTAGAAGGAGAGAGCGCGAG	60,0	GGAGGTTATCGTCGACTGGA	60,1	238
454_RES_PAHANG_locus1634	CCA	TCCAGTCGACGATAACCTCC	60,1	GTGAAGATCTCGCCGTTCTC	60,0	240
454_RES_PAHANG_locus1890	AT	ATGCGCATGCATGAGTACAG	60,9	GACCACAAAGTTAGCTCCGC	59,9	176
454_RES_PAHANG_locus3310	CTC	CTACCGCTCTACGATCCAGC	60,0	GAGTATGTGGCAAAGGGAG	59,6	203
454_RES_PAHANG_locus852	CAG	ACGCCTCCACAAGATCCTAA	59,7	TATGCACCACTCAGGTCTGC	59,9	134
454_RES_PAHANG_locus4029	CT	CCACAGCATAACACGACC	60,0	CCAACGGAAGAAAGTCCAAA	60,1	144
454_RES_PAHANG_locus3231	AGAT	GCTTGTGATGGATGGTGTG	60,0	GGTCAAAGGTGCTGAGAAGC	60,0	267
454_RES_PAHANG_locus1101	TAA	GAGGGCGACATCTTTGTGAT	60,1	AGTCTGAATCACGTCGGGTT	59,6	286
454_RES_PAHANG_locus38	CGA	ACTCCTTACGGACCCGTTCT	60,0	CGGAGGTAAAAGGGACATCA	59,9	164
454_RES_PAHANG_locus3080	ATG	AGCTGGAGGAGACAAGTGGAA	60,0	AGCCAAAAAGGGGAGAAAAA	60,0	290
454_RES_PAHANG_locus2246	TCG	TAATGGAGAAGATCCGCGAC	60,2	GCAGCAGAGGCTTTCTTGTT	59,8	201
454_RES_PAHANG_locus82	TGTC	AGTGAGGAGGCTGCCTATGA	60,0	GCAATATGCCTTTGAAATAAACC	59,0	192
454_RES_PAHANG_locus1081	GAT	TATCCTCGCCAGCAAGTTCT	60,0	ATTCTGCTGCCATTTGAAC	60,1	172
454_RES_PAHANG_locus1593	ATG	AGAACCCCAACAACAAGTCG	60,0	TCATCCTCTTCTTCTCTCCA	59,0	350
454_RES_PAHANG_locus2809	GCG	AGATGGGTCATCAGGTGGG	60,8	TAGCTGCATCATGGAAGACG	60,0	105
454_RES_PAHANG_locus3083	GGA	TCTGTTCTTTCCGGTGGGTTT	60,1	AGGTACCATCCAGGTGCTTG	60,0	255
454_RES_PAHANG_locus2616	GGA	TCAGAACCCAACCTCGCTCTT	60,0	CCACCTGGTGTCTTGAAC	60,0	272
454_RES_PAHANG_locus2722	TCC	GTCGACTGTGCTGAAGATGG	59,4	TGATGCTGCGCAACTTATTC	60,0	562
454_RES_PAHANG_locus2641	AAG	AAAAACAGAGCAAACCCCT	60,0	CCGCTTCTGTTCTGAGGTC	60,0	374
454_RES_PAHANG_locus2642	CATA	GACCTCAGAACAAGAAGCGG	60,0	TGCATTGTTCAAAGGAAAAGG	60,1	235
454_RES_PAHANG_locus2488	TGCC	CGTTTCCGCTTTCTTCTACG	60,0	TATGGTTCCGGTTGGAAGGAC	59,8	395
454_RES_PAHANG_locus1168	CAG	GGCTGCTGTCTCAACTTCC	60,0	CTTCTGCTTGCAAGGCTG	60,0	137
454_RES_PAHANG_locus3160	GGCG	AAACCTCTCTCGATCCCCTC	60,6	CGAGTTGAACTTGTGATCGG	59,3	287

TM: Temperatura de *melting* (°C); *Primer FW* (*Primer forward*) ; *Primer RV* (*Primer reverso*)

Dentre os 50 locos SSR de transcritos de *M. acuminata* Calcutta 4 que foram selecionados para validação de polimorfismo, 33 estão localizados dentro de regiões gênicas em *M. acuminata* DH Pahang, com a maioria situados proximamente a genes R já anotados, conforme Figura 22. Os demais locos, apesar de estarem próximos a genes R, estão localizados em posições não anotadas do genoma e não estão representados na figura.

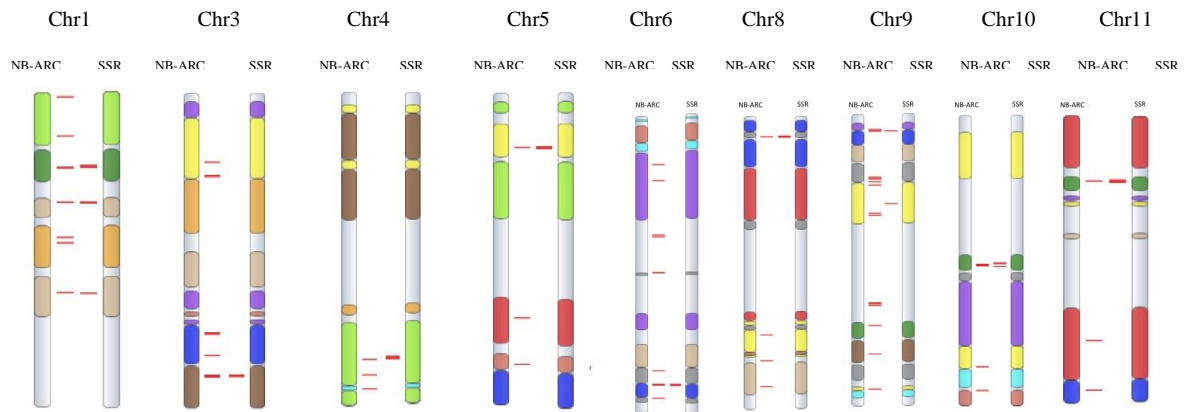


Figura 22: Representação genômica de *M. acuminata* DH Pahang com destaque para SSRs e genes NB-ARC. O cromossomo da esquerda corresponde a distribuição dos genes NB-ARC ao longo do cromossomo e o da direita corresponde a distribuição dos microssatélites co-localizados ao gene de interesse.

3.1.4. Desenvolvimento de marcadores moleculares SSR *in silico*, derivados de sequências de transcritos (sequenciamento 454) em *M. acuminata* Calcutta 4, de genes R de diversas famílias.

Foi identificado um total de 52 marcadores SSR pertencentes à diversas famílias de genes R derivados de transcritos 454 em *M. acuminata* Calcutta 4 (Tabela 16), sendo que 16 deles já foram testados em trabalhos anteriores (Passos *et al.*, 2013). No estudo atual, um total de 36 marcadores SSR, distribuídos nas diferentes famílias de genes R, foram avaliados quanto a polimorfismo (Figura 23).

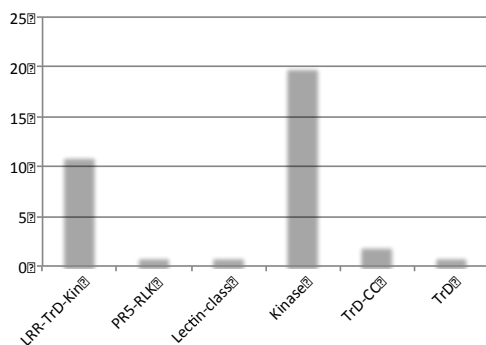


Figura 23: Distribuição de SSRs derivados de diferentes famílias de genes R, identificados em transcritos (sequenciamento 454) em *M. acuminata* Calcutta 4

Tabela 16: Informações dos locos SSR derivados de transcritos (sequenciamento 454) de *M. acuminata* Calcutta 4 pertencentes à diferentes famílias de genes R.

Loco	Família Gene R	Unigene ID	Repetição	Primer FW	TM	Primer RV	TM	Produto
loco2225	LRR-TrD-Kin	musa_c4_small_rep_c2434	(CTT)3	CATTAGCGCCCACTTTCATT	60,1	AGCCATCGCTGCACTCTATT	60,0	312
loco25	LRR-TrD-Kin	musa_c4_small_c10217	(CTTC)4	GAGCGCTTAGGCTTCTCTCA	60,0	GGTGCTTTATGCTGCTGTCA	60,0	98
loco929	LRR-TrD-Kin	musa_c4_small_c25095	(CGC)3	ACGGAAGTGGCAGTCAAAAAT	59,6	TTGCTGGTCTTCACATCTCG	60,0	459
loco1008	LRR-TrD-Kin	musa_c4_small_c26328	(GAGAAT)6	GGTTTGCACAGGAGATTGGT	60,0	AGGACACTGCTTGCTTGGAT	59,9	272
loco2225	LRR-TrD-Kin	musa_c4_small_rep_c2434	(CTT)3	CATTAGCGCCCACTTTCATT	60,1	AGCCATCGCTGCACTCTATT	60,0	312
loco2364	LRR-TrD-Kin	musa_c4_small_rep_c3064	(CTT)3	GGACTCACGAGTCCACAT	60,0	AGGAAGAGGTGGGTGAGGTT	60,0	557
loco2645	LRR-TrD-Kin	musa_c4_small_rep_c566	(TCGCCT)6	CCCACCCACCTACATCTCTC	59,4	AGTCCTTCTGGTTCAGCGA	60,0	244
loco3334	LRR-TrD-Kin	C4NIGN96RJG02GGFK3 435	(GCT)3	TATGGATCCCTCTCCTGTGC	60,0	CCGAGGAGGAAGAGGAGTCT	59,9	98
loco3826	LRR-TrD-Kin	C4NIGN96RJG02GR7TQ 377	(CTT)3	GCCCATTCTACTTTTCCCC	59,8	ATTCGAGAGGTTCCAGCTCA	60,0	306
loco1628	LRR-TrD-Kin	musa_c4_small_rep_c1220	(GCT)3	CTCCCTATCCCCTTTCCTCA	60,4	TCGTACCAGATCACAGCAGC	60,0	336
loco1205	LRR-TrD-Kin	musa_c4_small_c6272	(AG)2	AGATTGGTGGTTGCCAAGAC	60,0	TCATGCTTCCAAGCAGACAG	60,1	504
loco79	LRR-TrD-Kin	musa_c4_small_c10952	(CGC)3	TTCTCAAATCCGACCACCTC	60,1	CTTGAAGTCTTGGTAGCCGC	60,0	226
loco1351	LRR-PEST-ECS	musa_c4_small_c8037	(GCA)3	AGGCCACCAAATCATGAGAG	60,1	GCAAAGCTGTTGGGGTCTT	60,2	444
loco38	Kinase/Hm1	musa_c4_small_c10369	(CGA)3	ACTCCTTACGGACCCGTCT	60,0	CGGAGGTAAAAGGGACATCA	59,9	164
loco136	Kinase/Hm1	musa_c4_small_c11784	(GCG)3	AAGATGGACGTCTTCGCCTA	59,8	TTACACGGTGATAGCACCA	60,0	522
loco282	Kinase/Hm1	musa_c4_small_c13618	(GGA)3	ATAGCTTTGGGATGCTGCTG	60,4	CTCCGATGGCTTTATTTGGA	60,0	215
loco761	Kinase/Hm1	musa_c4_small_c21858	(GTG)3	TGGGGTTGGTCTGTGTTTT	60,2	GTCATGAGGTCCTACCACGC	60,5	393
loco1025	Kinase/Hm1	musa_c4_small_c3248	(AGC)3	ACAAATTTCTACCATCGCC	59,9	GTCCACTCTGCACTTCACCA	59,9	389
loco1062	Kinase/Hm1	musa_c4_small_c4114	(TCA)3	GATCCCCTCTCCAGACTTC	60,0	CCATGTCAATCGAGTTTGGGA	59,5	281
loco1301	Kinase/Hm1	musa_c4_small_c7392	(TGTT)4	TGCTGACGAAAGGATCACTG	60,0	AGCAAAACAAATCGACAGCA	59,5	575
loco1432	Kinase/Hm1	musa_c4_small_c9029	(CGC)3	ATTGGTGTCTGTGGTACA	60,0	TGGAAGGACTGGTTGCAGTT	60,7	597
loco1609	Kinase/Hm1	musa_c4_small_rep_c1177	(TGG)3	AGCTTACGGTGAGGAAGAA	60,0	GCTCTACTCGGTATCGGACG	59,9	258
loco1613	Kinase/Hm1	musa_c4_small_rep_c11853	(TTA)3	TGAAGTTGTATTCTCGGGC	60,1	ACCTACCAAGCCAAACATGC	60,0	332
loco1957	Kinase/Hm1	musa_c4_small_rep_c1776	(GTTTGGCT)8	TTCCAGGTTACATTGCACCA	60,0	CTTCGTTTCGTCATGCACATT	59,7	593
loco2281	Kinase/Hm1	musa_c4_small_rep_c2609	(GTG)3	CTAGAAACGGAATGAAGCCG	59,8	ACCCACTCACCCTCCAGAC	60,0	435
loco2754	Kinase/Hm1	musa_c4_small_rep_c717	(AAG)3	TTGTAATGTGGAACGCCAA	60,0	TCGGTGTATGATGGTGTGCT	60,0	587

loco2836	Kinase/Hm1	musa_c4_small_rep_c8269	(CTG)3	GCTGTCCAGCTCTCAAACC	60,0	CAAAGCATCAAAGCTTGCTGTC	59,9	244
loco3271	Kinase/Hm1	C4IGN96RJG02IYGQ8 381	(TCC)3	AATGGCAAAGAAGAAGCCCT	60,2	TGTCGCATCGGATCTTGTAG	59,8	189
loco3543	Kinase/Hm1	C4IGN96RJG01CL9T2 374	(CCG)3	CCAATTTCAAGTCGTCCTCC	59,5	GAGGAGGAGGGTTTAGTCGC	60,2	130
loco3567	Kinase/Hm1	C4IGN96RJG02JLFBN 350	(AAGG)4	AGGGCAATTTGGTGTGAG	60,0	GGATCTCAGGGCAATGTAA	59,9	344
loco3931	Kinase/Hm1	C4NIGN96RJG02J2H2R 438	(AAG)3	GTCCAGAGCCCACTTCAGAG	60,0	AAAGCCGTAGCACTTCTCA	60,0	337
loco3978	Kinase/Hm1	C4NIGN96RJG02I97VT 439	(CT)2	AATCCATCCCAGTCAAACG	59,8	AGAGGCAGAGGGAGAAGAGG	60,1	397
loco2971	Kinase/Hm1	musa_c4_clusters_pos_c1409_1	(AAG)3	GCCATTGAAGGTGACTTGGT	60,0	CGGTGATCTCAGCTTTCCTC	60,0	472
loco1826	Kinase/Hm1	musa_c4_small_rep_c1558	(GAT)3	TCCATCGTGACATCAAGTCC	59,5	CTTTGTGTGCCCATGTCAC	60,0	300
loco2028	Trd-CC	musa_c4_small_rep_c1930	(GGT)3	ACAACCCCTCTTGAGACCT	60,0	TCATGGTGTGAGCTTCTTG	60,0	233
loco2029	Trd-CC	musa_c4_small_rep_c1930	(TCG)3	CAAGAAGCTCGACACCATGA	60,0	AGGGAGAGTTAGCGGGTAGC	59,9	162

TM: Temperatura de *melting* (°C); *Primer FW* (*Primer forward*) ; *Primer RV* (*Primer reverso*); Produto (pb).

3.2. Desenvolvimento de marcadores SSR para genes potencialmente envolvidos em respostas de defesa

Foi identificado um total de 56 marcadores SSR em *M. acuminata* Calcutta 4, derivados de transcritos (sequenciamento 454), em genes potencialmente envolvidos em respostas de defesa (Tabela 17).

Tabela 17: Informações de primers SSR derivados de *Unigenes* de transcritos 454 de *M. acuminata* Calcutta 4 potencialmente envolvidos em respostas de defesa.

Loco	Unigene ID	Repetição	TM	Primer FW	TM	Primer RV	Produto	Descrição KOG
Loco 21	musa_c4_small_c10181	(CCT)3	59,8	GCCTGAGTACATC CACCCAT	60,2	GTCGGAGGAGAAG ATGGTGA	133	Predicted endo-1,3-beta-glucanase
Loco 22	musa_c4_small_c10184	(TTTA)4	60,0	AAGGGTGCCCAA CTGACTG	59,8	CACACGGGAGAGT CTGATGA	571	FOG: Predicted E3 ubiquitin ligase
Loco 24	musa_c4_small_c10207	(ACC)3	60,5	GGCGAAGACTTAC CCTCCAT	60,0	GAGGACTGAGAGT TGTCGGC	311	FOG: PPR repeat
Loco 64	musa_c4_small_c10840	(GGA)3	60,1	GGTGAGAAGGCGT CAGAGAG	60,0	AATCCGTCACGTC AAGTTCC	209	FOG: Predicted E3 ubiquitin ligase
Loco 117	musa_c4_small_c11561	(CAC)3	60,0	AGTCCAGCTCCAG AACCTCA	60,0	TCGACAACCAGAA GGAGCTT	144	FOG: Predicted E3 ubiquitin ligase
Loco 119	musa_c4_small_c11613	(GTT)3	60,0	GAGTCGTCGGAAA GATCGAG	60,1	GTCGACATGGAAG AAGGCAT	122	FOG: Predicted E3 ubiquitin ligase
Loco 120	musa_c4_small_c11621	(CAGCAT)6	59,8	ACTACAGGTGCAC ACATCGC	60,0	TGCTAATGAAGCC ACTGTCTG	357	A WRKY superfamily transcription factors
Loco 128	musa_c4_small_c11685	(CT)2	60,4	AAGAGGGAAAGG AATGACCG	60,1	AAATGGCTCTCCA TGTTTGC	192	FOG: Predicted E3 ubiquitin ligase
Loco 160	musa_c4_small_c12075	(CCA)3	59,7	CACATTCTCGAG CTCTGC	59,7	GTAAGGGCATCTC GTCGAAC	392	FOG: PPR repeat
Loco 187	musa_c4_small_c12418	(AGC)3	60,0	CTGAGCAGGGAGA GGTCAAC	60,0	TGGATCGTAGTGC AGAGTGC	226	FOG: Predicted E3 ubiquitin ligase
Loco 202	musa_c4_small_c12585	(TC)2	59,8	CGACCCCTTTTCAT TAACCA	60,0	ATCATCGATGTGC GTGTTGT	279	1,3-beta-glucan synthase/callose synthase catalytic subunit
Loco 248	musa_c4_small_c13136	(GTG)3	60,0	CTACTCTCGGCG TCTTCAC	60,0	CCAAGTAGAGTGC CACAGCA	375	N-methyl-D-aspartate receptor glutamate-binding subunit
Loco 256	musa_c4_small_c13270	(TCG)3	59,8	AGGGAAGGAATCG AAAGAGC	60,2	CGGGAGATATTTG CCTCTCA	455	FOG: Predicted E3 ubiquitin ligase
Loco 275	musa_c4_small_c13519	(AGT)3	59,9	GAGAGTCTTGCT GCTCGGT	60,1	AATACGCAGAAGA ACCGTGG	240	FOG: Predicted E3 ubiquitin ligase
Loco 571	musa_c4_small_c18303	(CAA)3	60,0	CTCAACAGCTGAA ACGGTGA	60,5	ATCGCTAGCTGCA GGATTTG	188	A WRKY superfamily transcription factors
Loco 872	musa_c4_small_c23888	(CAC)3	59,9	ATCCTCCTCCACAT AGCCCT	60,0	AACTCGCACAGGC AGACC	284	FOG: Predicted E3 ubiquitin ligase
Loco 956	musa_c4_small_c25569	(CAAT)4	60,0	TCAATCTCACCGA CCAATGA	59,2	TGTGCTATTGGAA ATTCGACA	121	A WRKY superfamily transcription factors
Loco 975	musa_c4_small_c25861	(CCGCCT)6	60,1	GACAGAAACCGAT CAAGGGA	59,8	GGCTACATGGAAG CCTGAAG	251	A WRKY superfamily transcription factors
Loco 1017	musa_c4_small_c2913	(GGA)3	60,0	GCTGAGGTTGATC TGTGCAA	60,0	TCCATCAGTATCCT CCAGGC	588	A No apical meristem (NAM) developmental protein and related proteins
Loco 1018	musa_c4_small_c2942	(TC)2	59,2	TCTACTCTCGCTCT CTCTCGC	60,0	ATCCAAGGTTTTG ACGCAAC	268	Glutathione peroxidase
Loco 1018	musa_c4_small_c2942	(TC)2	59,2	TCTACTCTCGCTCT CTCTCGC	60,0	ATCCAAGGTTTTG ACGCAAC	268	Glutathione peroxidase
Loco 1026	musa_c4_small_c3274	(TGACA)5	59,9	GATGCGGACAAAAG GTTTCAT	60,0	CCAGTGGTGTAC AAGGATG	252	Glutathione synthetase
Loco 1026	musa_c4_small_c3274	(TGACA)5	59,9	GATGCGGACAAAAG GTTTCAT	60,0	CCAGTGGTGTAC AAGGATG	252	Glutathione synthetase

Loco 1063	musa_c4_small_c4120	(CAA)3	60,1	GCAAGCAACTACG ACGACAA	59,9	CAAAGAATAATGC GGGGAAA	169	A WRKY superfamily transcription factors
Loco 1068	musa_c4_small_c4154	(ATCGC)5	59,9	AGCTGCAGTTTGA CATCCCT	60,1	CCGAGAAAACCA ATTCAGT	259	UDP-glucuronosyl and UDP-glucosyl transferase
Loco 1116	musa_c4_small_c4754	(GGC)3	60,0	GCACCAAAGGCTA AAAGCTG	60,2	GGAGAAAATGAAA GCGGTGA	598	Epithelin/granulin
Loco 1132	musa_c4_small_c5016	(CCCT)4	60,0	CTCTTGTCAGCAC CCCTCTC	59,9	AAGGAAGTCCGAC CAAGGAT	503	UDP-glucuronosyl and UDP-glucosyl transferase
Loco 1262	musa_c4_small_c6836	(TGT)3	60,0	TGGATCGTTGCTC ACTCAAG	59,5	TGCCTTATATCATT GATCACGAA	198	Chorismate synthase
Loco 1516	musa_c4_small_rep_c10393	(TCCTCT)6	59,9	TACGCTTACCGTG CTTTCCT	59,9	TAGAGGTGGAGGT TCATGGG	295	Predicted glutathione S-transferase
Loco 1520	musa_c4_small_rep_c10443	(AG)2	60,4	AGCAACATCCTCG CAGAAAC	60,0	GGAGAACCCTTCGT TGAGCTG	392	UDP-glucuronosyl and UDP-glucosyl transferase
Loco 1687	musa_c4_small_rep_c1307	(GAGGTG)6	59,9	ACTGCGGGTAGGA GTTCTGA	60,0	GCAGACCACCCAG TCATTTT	579	A No apical meristem (NAM) developmental protein and related proteins
Loco 1688	musa_c4_small_rep_c1307	(TCC)3	60,0	ATGTTGGAAGCAG GGAACAC	60,0	CCGACGAGAGATC AGTCACA	142	A No apical meristem (NAM) developmental protein and related proteins
Loco 1747	musa_c4_small_rep_c14140	(CAG)3	60,0	TCACAAGCATGGT GTGGTTT	60,1	CCCGCTTACCAC CATAAGA	596	SNARE protein Syntaxin 18/UFE1
Loco 1803	musa_c4_small_rep_c15097	(CCT)3	59,9	CCGAGAGTGAGCT TCTGCTT	60,0	CTTGAACCCCATC TCTTCCA	146	Flavonol reductase/cinnamoyl-CoA reductase
Loco 2115	musa_c4_small_rep_c2133	(AT)2	60,0	ACACCAGAGAAAG AAGCGGA	60,0	GTCGTTGGTGGTG ACATGAG	199	A Germin/oxalate oxidase
Loco 2116	musa_c4_small_rep_c2133	(TC)2	60,0	CTCATGTCACCAC CAACGAC	60,0	CAAACCCATCACA ACAGACG	578	A Germin/oxalate oxidase
Loco 2122	musa_c4_small_rep_c217	(GA)2	60,0	TCAGAAGTCCGCA TCTCCTT	60,0	CTCACGTTGGTGG AAAGGTT	89	Macrophage migration inhibitory factor
Loco 2122	musa_c4_small_rep_c217	(GA)2	60,0	TCAGAAGTCCGCA TCTCCTT	60,0	CTCACGTTGGTGG AAAGGTT	89	Macrophage migration inhibitory factor
Loco 2180	musa_c4_small_rep_c2331	(CTT)3	59,9	TTCTCCACATCT AATCGG	60,0	CCTGGTCTCCTTCT GTCGAG	402	Chorismate synthase
Loco 2440	musa_c4_small_rep_c362	(CTC)3	60,0	ATTGTAGGTGGTG GAGACGC	60,0	TCATGAAGCTGGT GAAGTGC	224	Cysteine proteinase Cathepsin F
Loco 2462	musa_c4_small_rep_c3783	(GCA)3	60,0	TACAAACCGCAAA TCCAACA	60,1	TGCCATCCTCCTAT GCTACC	534	A No apical meristem (NAM) developmental protein and related proteins
Loco 2469	musa_c4_small_rep_c3819	(ATG)3	60,5	GTGCCAAATTTAA CCCTCCC	59,9	ATGGCTTCCGTGTT GTTACC	81	Peroxidase
Loco 2470	musa_c4_small_rep_c3819	(TCTCC)5	60,0	GAGGGGATCTCGA TGAATGA	59,8	GCGCCAAATGGTG AATTTAT	368	Peroxidase
Loco 2615	musa_c4_small_rep_c529	(TCCTC)5	60,2	TCAAGAACAACCA GGCCTTC	59,2	CGAATTCTTCTTCC ATTTGACA	291	
Loco 2637	musa_c4_small_rep_c5576	(CT)2	60,0	TGTCCATTTGAGG AGGGAAG	59,6	GAGAGTAGCGAGG TCGTCTG	84	Lipoxygenase
Loco 2649	musa_c4_small_rep_c5730	(GTC)3	60,1	AAGCAGAAAACCA TGGATCG	60,0	CAAAACCTTCTTC ATCCCGA	498	A No apical meristem (NAM) developmental protein and related proteins
Loco 2650	musa_c4_small_rep_c5730	(TCT)3	60,0	CCTTTGAGCTGAA CTCTGGG	60,1	TCGAACATTCGTT GCACATT	136	A No apical meristem (NAM) developmental protein and related proteins
Loco 2879	musa_c4_small_rep_c891	(GAAA)4	60,0	GTGCACATACACT TGGACGG	59,9	CTGCCACCAAGTG CTTCATA	469	Peroxidase

Loco 2966	musa_c4_clusters_pos_c1375_1	(CT)2	58,9	CCTTTTCCTTCTCT TCACCG	60,0	CCACAGCAGCAAT AGTCGAA	576	Cysteine proteinase Cathepsin L
Loco 2967	musa_c4_clusters_pos_c1375_1	(TCC)3	60,1	CTTCACCGTCTCTG CCTCTC	60,0	TCCACAGCAGCA ATAGTCG	566	Cysteine proteinase Cathepsin L
Loco 2971	musa_c4_clusters_pos_c1409_1	(AAG)3	60,0	GCCATTGAAGGTG ACTTGGT	60,0	CGGTGATCTCAGC TTTCCTC	472	A Seven transmembrane protein Mlo
Loco 2992	musa_c4_clusters_pos_c1541_1	(CCT)3	60,0	AATGGTTGCACTT ACCCTGC	60,0	TCGATTCTGCTGTC CACTTG	401	A 1, 3-beta-glucanase
Loco 3085	musa_c4_clusters_pos_c545_1	(TG)2	60,1	ATAAGCATGAACC AGGACGC	57,8	TCCTATGCATGAG ATTGAGATG	318	Lipoxygenase
Loco 3356	C4IGN96RJG02JAHAV 342	(CGC)3	61,4	GATGTAAACGCCG ATGAACG	60,2	AATCCCTTCACTTC CTCGCT	229	A beta-1, 3 glucanase
Loco 3356	C4IGN96RJG02JAHAV 342	(CGC)3	61,4	GATGTAAACGCCG ATGAACG	60,2	AATCCCTTCACTTC CTCGCT	229	A beta-1, 3 glucanase
Loco 3634	C4IGN96RJG01DXNUA 365	(AAG)3	60,7	GACGCTGTCGTCC TGGTAAC	60,0	ACGAGGCAGAAAAG GAACAGA	149	Peroxidase

TM: Temperatura de *melting* (°C); *Primer FW* (*Primer forward*) ; *Primer RV* (*Primer reverso*); Produto (pb)

3.3. Caracterização de marcadores microssatélites para locos potencialmente envolvidos em resposta aos estresses bióticos e análise de polimorfismos em genótipos de *M. acuminata*

A partir de diferentes estratégias de busca, foram identificados e selecionados para validação de polimorfismo um subgrupo de 156 marcadores SSR derivados de regiões expressas de *M. acuminata* Calcutta 4, Cavendish Grande Naine e DH Pahang. Recentemente, Mudalkar e colaboradores (2014) encontraram 19.379 SSRs dentre mais de 53.854 *contigs* transcritos em *Camelina sativa*, e Tan e colaboradores (2013) identificaram 12.582 SSRs distribuídos em 10.290 *unigenes* de *Camellia sinensis*. Tais trabalhos, dentre muitos outros, indicam a existência de microssatélites com frequência alta em regiões expressas do genoma de espécies cultivadas.

Os 156 SSRs selecionados neste estudo foram organizados em seis grupos com base no número de nucleotídeos de cada marcador: 22 (14,1%) foram compostos por motivos dinucleotídicos, 105 (67,3%) por motivos trinucleotídicos, 16 (10,2%) por motivos tetranucleotídicos, 5 (3,2%) por motivos pentanucleotídicos, 7 (4,4%) por motivos hexanucleotídicos e apenas 1 (0,6%) por motivos octanucleotídicos. Os mais abundantes dos motivos foram AGA, presentes em 5,12% do total, seguido por CT presentes em 4,4% dos SSRs. Em conformidade, Lagercrantz e colaboradores (1993) identificaram em estudos sobre abundância de marcadores microssatélites, os motivos AT e CT como sendo os mais abundantes em plantas, porém, neste estudo a maioria dos motivos AT foi eliminada devido a dificuldades de amplificação deste motivo. Passos e colaboradores (2013) identificou o dinucleotídeo CT como sendo o mais abundante e os trinucleotídeos AAG e AGA presentes com alta frequência em genótipos de banana *M. acuminata* Calcutta 4 e Cavendish Grande Naine.

Os marcadores microssatélites foram inicialmente avaliados em *bulks* de *M. acuminata* contrastantes em resistência, para a detecção preliminar de polimorfismo através de visualização em gel de poliacrilamida corado com nitrato de prata (Figura 24).

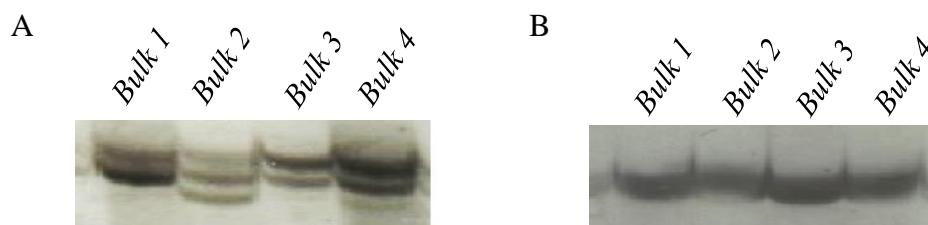


Figura 24: Análise eletroforética em gel de poliacrilamida 4% de amplicons do loco polimórfico 454_DEF_locus120 (A) e do loco monomórfico 454_DEF_locus2116 (B) em *bulks* de genótipos de *M. acuminata* contrastantes em resistência. *Bulk 1* – genótipos resistentes a Sigatoka negra, *Bulk 2* – genótipos suscetíveis a Sigatoka negra, *Bulk 3* – genótipos resistentes a Sigatoka amarela e *Bulk 4* – genótipos suscetíveis a Sigatoka amarela

A estratégia de montagem de *bulks* contrastantes em resistência, permite uma otimização e análise de polimorfismo de um maior número de marcadores SSR, sendo amplamente utilizada em mandioca (Olasanmi *et al.*, 2014), algodão (Marangoni *et al.*, 2013), milho (Xu *et al.*, 1999), trigo (Ghazvini *et al.*, 2013), tomate (Joshi *et al.*, 2013) dentre outros. A análise em géis de poliacrilamida se deu devido a alta precisão e baixo custo, como justificada por Menksen & Kahl (2005) que afirmam que a precisão obtida com géis de poliacrilamida é maior que a obtida com agarose e próxima da obtida em eletroforese capilar o qual possui custo mais elevado.

Os locos que apresentaram comportamento polimórfico foram testados nos 20 parentais contrastantes de *M. acuminata*, além das duas progênes (geração F1) entre Calcutta 4 e Pisang Berlin, com os resultados confirmando os dados prévios obtidos nas análises em *bulks* (exemplo na Figura 25).

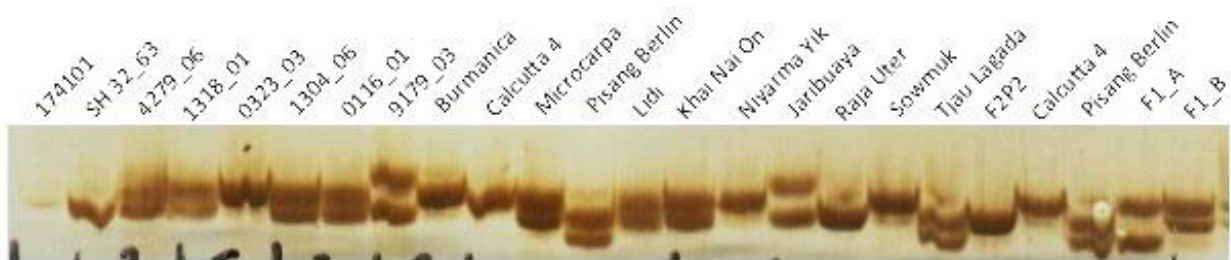


Figura 25: Análise eletroforética em gel de poliacrilamida 4% de amplicons polimórficos em parentais de *M. acuminata*, loco 454_DEF_locus120. F1_A e F1_B são progênes de *M. acuminata* Calcutta 4 e Pisang Berlin, com os dois parentais repetidos ao lado das progênes para facilitar interpretação.

O polimorfismo dos microssatélites pode ter diversas origens e os principais mecanismos apontados para a alta taxa de mutação nessas regiões são o “deslizamento” da polimerase durante a replicação e a recombinação desigual. O primeiro se refere ao “deslizamento” da DNA polimerase durante a replicação (*DNA “slippage”*) (Tachida *et al.*, 1992). Esses deslizamentos da DNA polimerase em regiões microssatélites acontecem quando a mesma avança ou recua durante a replicação de uma região repetitiva, aumentando ou diminuindo o número de repetições do bloco de DNA. O segundo mecanismo se refere ao processo de recombinação desigual onde o polimorfismo surge como consequência do emparelhamento equivocado entre os cromossomos homólogos, causado pela presença de seqüências repetitivas resultando numa fita dupla mais curta e outra mais longa após a recombinação (Harding *et al.*, 1992).

Do total de 156 marcadores microssatélites para locos potencialmente envolvidos em respostas a estresses bióticos, 21 (13,46%) tiveram polimorfismo confirmado pela análise individual. Os demais apresentaram características monomórficas (Tabela 18), amplificação inespecífica ou falhas na amplificação.

As falhas de amplificação podem refletir divergências nas sequências que flanqueiam os microssatélites, produzindo alelos nulos. Esses alelos surgem quando ocorre uma mutação no DNA alvo, precisamente na sequência homóloga à sequência do *primer* utilizado na reação de PCR. Isto reduz ou inibe completamente a capacidade de anelamento do mesmo com a fita molde (Callen *et al.*, 1993). Esses alelos mutantes não são visíveis nos géis submetidos às técnicas de revelação de bandas normalmente utilizadas para produtos de PCR separados por eletroforese. Apesar de serem amplamente empregados nos estudos de populações, a possível presença de alelos nulos é considerada uma restrição à aplicação dos marcadores microssatélites em análises genéticas (Pemberton *et al.*, 1995) e, por esta razão, os locos com problemas na amplificação foram descartados. Em *Musa*, essas falhas também foram constatadas por Crouch *et al.* (1998) e Creste *et al.* (2003).

Todos os *primers* foram desenhados para microssatélites presentes dentro de regiões gênicas, o que explica o baixo grau de polimorfismo das amostras. Sequências que não codificam produtos protéicos são potencialmente mais variáveis entre alelos de uma mesma espécie, pois as consequências da variação de sequência nessas regiões é esperado que seja menor do que as variações que ocorrem dentro de regiões codantes. Pelo fato dos níveis de variação não serem iguais entre as regiões gênicas (promotor, exon, íntron), espera-se que as consequências fenotípicas também não sejam iguais. Sequências que têm efeitos amplos e diretos sobre fenótipos tendem a manter o mínimo de variação sob seleção (Holland *et al.*, 2001), além do que as sequências de íntrons evoluem mais rapidamente do que as sequências de exons em plantas (Small e Wendel, 2000).

Tabela 18: Marcadores microssatélites derivados de *unigenes* potencialmente envolvidos em respostas a estresses bióticos. He- heterozigidade esperada. Ho - Heterozigidade observada. PIC- Conteúdo de Informação Polimórfica

Loco	Polimorfismo	Amplitude Alélica (pb)	Nº de alelos	He	Ho	PIC
MaC139M12_loc115	Polimórfico	347-353	3	0,4298	0,3636	0,3855
Ma4052E23_loc3	Polimórfico	344-350	3	0,4680	0,5455	0,4050
MaC063A04_loc47	Polimórfico	347-353	2	0,4163	0,5000	0,3297
MaC063A04_loc62	Polimórfico	394-403	4	0,4963	0,4500	0,4517
454_DEF_locus1516	Polimórfico	272-284	3	0,5527	0,5909	0,4704
454_DEF_locus3356	Polimórfico	228-243	4	0,4990	0,2273	0,4646
454_DEF_locus3634	Polimórfico	144-159	3	0,6333	0,7727	0,5609
454_DEF_locus2637	Polimórfico	84-88	3	0,6505	0,2941	0,5757
454_DEF_locus120	Polimórfico	331-343	4	0,6002	0,5909	0,5179
454_DEF_locus2650	Polimórfico	146-155	4	0,5465	0,3636	0,5058
454_DEF_locus187	Polimórfico	235-253	4	0,6188	0,5455	0,5661
454_DEF_locus275	Polimórfico	283-300	2	0,4990	0,6818	0,3745
454_DEF_locus1262	Polimórfico	197-206	4	0,7263	0,6500	0,6775
454_RES_LRR_RLK_locus2225	Polimórfico	323-326	2	0,3967	0,5455	0,3180
454_RES_LRR_RLK_locus2645	Polimórfico	240-258	4	0,5785	0,6364	0,5258
454_RES_LRR_RLK_locus3826	Polimórfico	312-321	4	0,6849	0,3182	0,6335
454_RES_PAHANG_locus2811	Polimórfico	266-275	3	0,5548	0,6818	0,4592
454_RES_PAHANG_locus1081	Polimórfico	170-179	3	0,6281	0,5455	0,5511
454_RES_PAHANG_locus2809	Polimórfico	100-103	2	0,4163	0,5000	0,3297
454_RES_PAHANG_locus3083	Polimórfico	247-253	3	0,6281	1,0000	0,5511
454_RES_PAHANG_locus2642	Polimórfico	232-244	4	0,6994	0,7273	0,6491
MaC139M12_loc110	Monomórfico	285	1	0,5582	0,5491	0,4906
MaC139M12_loc112	Monomórfico	280	1		Média	
MaC091O16_loc82	Monomórfico	161	1			
MaC063A04_loc75	Monomórfico	280	1			
454_DEF_locus2116	Monomórfico	578	1			
454_DEF_locus2122	Monomórfico	89	1			

454_DEF_locus1132	Monomórfico	503	1
454_DEF_locus1520	Monomórfico	392	1
454_DEF_locus2992	Monomórfico	401	1
454_DEF_locus571	Monomórfico	188	1
454_DEF_locus956	Monomórfico	121	1
454_DEF_locus1063	Monomórfico	169	1
454_DEF_locus1688	Monomórfico	142	1
454_DEF_locus1803	Monomórfico	146	1
454_DEF_locus2470	Monomórfico	368	1
454_DEF_locus256	Monomórfico	455	1
454_DEF_locus2462	Monomórfico	534	1
454_DEF_locus117	Monomórfico	144	1
454_DEF_locus119	Monomórfico	122	1
454_DEF_locus872	Monomórfico	284	1
454_DEF_locus2440	Monomórfico	224	1
454_RES_LRR_RLK_locus929	Monomórfico	459	1
454_RES_LRR_RLK_locus2225	Monomórfico	312	1
454_RES_LRR_RLK_locus3334	Monomórfico	98	1
454_RES_LRR_RLK_locus1628	Monomórfico	336	1
454_RES_LRR_RLK_locus1205	Monomórfico	504	1
454_RES_PR5_RLK_locus79	Monomórfico	226	1
454_RES_LLECTIN_CLASS_locus1351	Monomórfico	444	1
454_RES_STK_locus136	Monomórfico	522	1
454_RES_STK_locus282	Monomórfico	215	1
454_RES_STK_locus761	Monomórfico	393	1
454_RES_STK_locus1025	Monomórfico	389	1
454_RES_STK_locus1432	Monomórfico	597	1
454_RES_STK_locus1609	Monomórfico	258	1
454_RES_STK_locus1613	Monomórfico	332	1
454_RES_STK_locus1957	Monomórfico	593	1
454_RES_STK_locus2281	Monomórfico	435	1

454_RES_STK_locus3567	Monomórfico	344	1
454_RES_STK_locus3931	Monomórfico	337	1
454_RES_STK_locus3978	Monomórfico	397	1
454_RES_TRD_CC_locus2028	Monomórfico	233	1
454_RES_PAHANG_locus3316	Monomórfico	202	1
454_RES_PAHANG_locus1458	Monomórfico	297	1
454_RES_PAHANG_locus3952	Monomórfico	302	1
454_RES_PAHANG_locus407	Monomórfico	471	1
454_RES_PAHANG_locus513	Monomórfico	197	1
454_RES_PAHANG_locus1102	Monomórfico	129	1
454_RES_PAHANG_locus2274	Monomórfico	433	1
454_RES_PAHANG_locus2273	Monomórfico	327	1
454_RES_PAHANG_locus1096	Monomórfico	284	1
454_RES_PAHANG_locus3698	Monomórfico	253	1
454_RES_PAHANG_locus311	Monomórfico	189	1
454_RES_PAHANG_locus3990	Monomórfico	316	1
454_RES_PAHANG_locus2243	Monomórfico	132	1
454_RES_PAHANG_locus549	Monomórfico	291	1
454_RES_PAHANG_locus3360	Monomórfico	364	1
454_RES_PAHANG_locus3561	Monomórfico	250	1
454_RES_PAHANG_locus1633	Monomórfico	238	1
454_RES_PAHANG_locus1634	Monomórfico	240	1
454_RES_PAHANG_locus3310	Monomórfico	203	1
454_RES_PAHANG_locus1101	Monomórfico	286	1
454_RES_PAHANG_locus82	Monomórfico	192	1
454_RES_PAHANG_locus2616	Monomórfico	272	1
454_RES_PAHANG_locus2641	Monomórfico	374	1
454_RES_PAHANG_locus2488	Monomórfico	395	1
454_RES_PAHANG_locus1168	Monomórfico	137	1
454_RES_PAHANG_locus3160	Monomórfico	287	1

He: heterozigidade esperada Ho: Heterozigidade observada PIC: Conteúdo de informação polimórfica

Foram detectados 68 alelos nos 21 locos SSRs polimórficos em que o número de alelos por loco variou de dois a quatro, com média de 3,23 alelos por loco SSR nos 20 genótipos de *M. acuminata*. Variação alélica semelhante foi encontrada em caracterizações de locos SSRs nos mesmos genótipos de *M. acuminata*, obtendo variação alélica média de 4,41 (Passos *et al.*, 2013), 3,23 (Passos *et al.*, 2012), e 2,8 (Miller *et al.*, 2010). Samarasinghe e colaboradores (2010) obtiveram média de variação alélica de 6,3 na caracterização de locos SSR em genótipos de ambos, genoma A e B de *Musa*.

O tamanho dos alelos variou de 84 a 403pb nos locos 454_DEF_locus2637 e MaC063A04_loc62, respectivamente. A média da heterozigiosidade observada dos 21 locos foi 0,54, com os locos 454_DEF_locus3634 e 454_RES_PAHANG_locus2642 apresentando as maiores frequências (0,77 e 0,72, respectivamente). Dos locos analisados, a heterozigiosidade observada foi maior que a esperada em onze locos, porém a média de H_e foi superior a de H_o , sugerindo excesso de homozigotos em relação ao modelo de equilíbrio de Hardy-Weinberg.

Segundo Botstein *et al.* (1980) o conteúdo de informação polimórfica (PIC) pode ser definido como um indicador de qualidade do marcador estudado. Dessa forma, o valor de PIC variou de 0,31 (454_RES_LRR_RLK_locus2225) a 0,67 (454_DEF_locus1262), sendo que mais da metade dos marcadores obtiveram PIC superior a 0,5. Segundo tais autores, marcadores serão considerados altamente informativos quando apresentarem PIC superior a 0,5; razoavelmente informativos, quando possuírem valores entre 0,25 e 0,50; e, levemente informativos, os que obtiverem valores inferiores a 0,24. Os locos que obtiveram PIC mais baixo, 454_RES_LRR_RLK_locus2225 (0,31), 454_RES_PAHANG_locus2809 (0,32) foram razoavelmente informativos. O menor conteúdo informativo, observado para esses dois locos, deve-se ao fato de as frequências alélicas não terem sido distribuídas de forma equitativa, com poucos alelos concentrando grande parte das frequências alélicas. Em relação ao loco 454_RES_LRR_RLK_locus2225, 72,7% da frequência alélica foi concentrada em apenas um dos alelos. O mesmo ocorreu no loco 454_RES_PAHANG_locus2809, em que 70,4% da frequência alélica se concentrou em apenas um alelo.

As aplicações de microssatélites em *Musa* concentram-se em taxonomia, como por exemplo em Lagoda *et al.*, (1998), genotipagem, como em Christelová *et al.*, (2011) e saturação de mapas genéticos, como por exemplo em Hippolyte *et al.*, (2010). SSRs derivados de genes podem ser associados com a variação genética funcional, oferecendo um potencial para seleção assistida por marcadores, com marcadores ou provenientes de um gene para uma característica fenotípica desejável, ou co-localizados com um determinado loco de característica quantitativa. Com esses marcadores isolados de regiões codantes, a conservação também é potencialmente maior, aumentando a possibilidade de transferência de espécies relacionadas (Gupta *et al.*, 2013). Em

comparação com outras culturas, relativamente poucos marcadores SSR foram desenvolvidos para o gênero *Musa*, refletindo os recursos de sequências limitados disponíveis até recentemente. Embora a maioria tenha sido isolado a partir de bibliotecas de DNA genômico ou clones BAC (Cheung e Town, 2007; Lagoda et al., 1998; Kaemmer et al., 1997; Ravishankar et al., 2012), apenas alguns SSR derivados de genes foram caracterizadas até hoje, como, por exemplo, em Wang et al., (2008) e Amorim et al., (2012).

Este trabalho representa uma contribuição para o desenvolvimento de novos marcadores SSR gênicos para aplicação no melhoramento genético de *Musa*. Com os esforços em curso para o desenvolvimento de populações segregantes para características de interesse (Amorim et al., 2009; Rekha, et al., 2011), estes marcadores baseados em genes funcionais são aplicáveis para associação de QTL e seleção assistida por marcadores, bem como a análise da evolução, avaliação de parentesco e aplicações gerais de genotipagem em programas de melhoramento.

CONCLUSÃO

A identificação de genes relacionados com a resistência de plantas a pragas e doenças é de grande importância para o entendimento dos mecanismos de defesa envolvidos durante o ataque de patógenos.

Neste trabalho, a anotação funcional, por meio de análise BLAST de dados de transcrito (pirosequenciamento 454) de folha do genótipo selvagem diplóide *M. acuminata* ssp. *burmannicoides* Calcutta 4 (resposta incompatível) e triplóides cv *M. acuminata* subgrupo Cavendish comercial Grande Naine (resposta compatível), ambos não infectados e desafiados com *M. musicola*, permitiu identificar 14 genes R expressando proteínas com sítios de ligação de nucleotídeo e região rica em leucina (NBS-LRR) em tecidos foliares de Calcutta 4 e 25 em Cavendish Grande Naine, ambos infectados e não-infectados. Mapeamento dos *contigs Unigenes* derivados de 454 com modelos de genes que contêm o domínio NB-ARC no genoma de referência de *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang) identificou 38 *contigs* mapeando a 40 modelos de genes completos, no caso de dados de transcrito de Calcutta 4 e 43 *contigs* mapeando em 40 modelos de genes em Cavendish Grande Naine.

Os genes potencialmente envolvidos em respostas de resistência e defesa identificados pela análise transcritômica foram utilizados como fonte de recursos para o desenvolvimento de marcadores SSR e análise de expressão desses genes.

Dos 156 marcadores microssatélites desenvolvidos e avaliados em genótipos de *M. acuminata*, 21 apresentaram comportamento polimórfico, com cinco deles flanqueando genes NB-ARC já descritos em *M. acuminata* DH Pahang. Esses marcadores, com alto conteúdo de informação polimórfica (PIC), são considerados os mais úteis para aplicação em seleção assistida por marcadores. O padrão alélico para os 68 alelos revelados em 21 locos SSR neste trabalho é um esforço para o emprego de marcadores SSR em situações de proteção de cultivares para o agronegócio de banana no Brasil, em adição aos marcadores fenotípicos. Os marcadores microssatélites são específicos para cada espécie, no entanto, pelo fato de as regiões flanqueadoras serem conservadas entre espécies ou gêneros próximos, é possível que esses iniciadores desenvolvidos possam ser transferidos para espécies correlacionadas.

As análises baseadas em PCR em tempo real da expressão de 51 genes potencialmente envolvidos em respostas de defesa e resistência durante a interação entre *M. acuminata* e *M. musicola* demonstraram que 19 obtiveram especificidade na reação de PCR

em tempo real e que a maioria dos genes envolvidos em processos de defesa não foram modulados, com exceção do gene da corismato sintase e NAC. O gene NPR1 foi expresso apenas no cultivar suscetível e os RGAs 37, 43, M09A31, M09A21, B03A51 e B03A11 tiveram um aumento de expressão quando inoculados com o patógeno. Os RGAs 43 e RGA 37 mostraram indução significativa em *M. acuminata* Calcutta 4 quando inoculada com o patógeno e os RGAs 9330, 7100, 6160, M09A31 e M09A21 apresentaram regulação negativa em Calcutta 4.

A caracterização contínua de genes envolvidos em respostas ao estresse biótico em *M. acuminata* durante a interação com *M. musicola*, bem como o desenvolvimento de marcadores moleculares, irá contribuir para o desenvolvimento de um controle eficaz da doença Sigatoka com base no melhoramento genético através da transformação de plantas ou seleção assistida por marcadores.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Agarwal M, Shrivastava N, Padh H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 2008 Apr;27(4):617–31.
2. Agarwal PK, Agarwal P, Reddy MK, Sopory SK. Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.* 2006 Dec 1;25(12):1263–74.
3. Agrios GN. *Plant Pathology.* Academic Press. 2005;(5):633.
4. Aguilar Morán J. Improvement of cavendish banana cultivars through conventional breeding. *Acta Hort (ISHS).* 2013;986:205–8.
5. Allen C, Prior P, Hayward AC. Bacterial wilt disease and the *Ralstonia solanacearum* species complex. *American Phytopathological Society;* 2005. 540 p.
6. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997 Sep 1;25(17):3389–402.
7. Alzate-Marin AL, Cervigni GDL, Moreira MA, Barros EG. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. *Fitopatologia Brasileira.* 2005 Aug;30(4):333–42.
8. Amorim EP, Silva PH, Ferreira CF, Amorim VBO, Santos VJ, Vilarinhos AD, et al. New microsatellite markers for bananas (*Musa* spp). *Genet Mol Res.* 2012;11(2):1093–8.
9. Amorim EP, Lessa LS, Ledo CA da S, Amorim VB de O, Reis RV dos, Santos-Serejo JA dos, et al. Caracterização agronômica e molecular de genótipos diplóides melhorados de bananeira. *Revista Brasileira de Fruticultura.* 2009 Mar;31(1):154–61.
10. Arzanlou M, Abeln ECA, Kema GHJ, Waalwijk C, Carlier J, Vries I de, et al. Molecular diagnostics for the sigatoka disease complex of banana. *Phytopathology.* 2007 Sep;97(9):1112–8.
11. Azhar M, Heslop-Harrison JS. Genomes, diversity and resistance gene analogues in *Musa* species. *Cytogenet Genome Res.* 2008;121(1):59–66.
12. Backiyarani S, Uma S, Arunkumar G, Saraswathi M, Sundararaju P. Cloning and characterization of NBS-LRR resistance gene analogues of *Musa* spp. and their expression profiling studies against *Pratylenchus coffeae*. *Academic Journals.* 2013 Jul 3;12(27):4256–68.
13. Bai T-T, Xie W-B, Zhou P-P, Wu Z-L, Xiao W-C, Zhou L, et al. Transcriptome and Expression Profile Analysis of Highly Resistant and Susceptible Banana Roots Challenged with *Fusarium oxysporum* f. sp. *cubense* Tropical Race 4. *PLoS ONE.* 2013;8(9):e73945.

14. Ballvora A, Ercolano MR, Weiss J, Meksem K, Bormann CA, Oberhagemann P, et al. The R1 gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *Plant J.* 2002 May;30(3):361–71.
15. Bar M, Sharfman M, Ron M, Avni A. BAK1 is required for the attenuation of ethylene-inducing xylanase (Eix)-induced defense responses by the decoy receptor LeEix1. *Plant J.* 2010 Sep;63(5):791–800.
16. Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. SNP discovery via 454 transcriptome sequencing. *The Plant Journal.* 2007;51(5):910–8.
17. Baurens F-C, Bocs S, Rouard M, Matsumoto T, Miller RN, Rodier-Goud M, et al. Mechanisms of haplotype divergence at the RGA08 nucleotide-binding leucine-rich repeat gene locus in wild banana (*Musa balbisiana*). *BMC Plant Biology.* 2010 Jul 16;10(1):149.
18. Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, et al. RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science.* 1994 Sep 23;265(5180):1856–60.
19. Beveraggi A, Mourichon X, Salle G. Etude comparee des premieres etapes de l'infection chez les bananiers sensibles et resistants infectes par le *Cercospora fijiensis* (*Mycosphaerella fijiensis*) agent responsable de la maladie des raies noires. *Canadian Journal of Botany.* 1995;73:1328–37.
20. Beveraggi A, Mourichon X, Sallé G. Étude comparée des premières étapes de l'infection chez des bananiers sensibles et résistants infectés par le *Cercospora fijiensis* (*Mycosphaerella fijiensis*) agent responsable de la maladie des raies noires. *Can J Bot.* 1995 Sep 1;73(9):1328–37.
21. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res.* 2004 May;14(5):988–95.
22. Bittner-Eddy PD, Crute IR, Holub EB, Beynon JL. RPP13 is a simple locus in *Arabidopsis thaliana* for alleles that specify downy mildew resistance to different avirulence determinants in *Peronospora parasitica*. *Plant J.* 2000 Jan;21(2):177–88.
23. Boller T, He SY. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science.* 2009 May 8;324(5928):742–4.
24. Botella MA, Parker JE, Frost LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, et al. Three genes of the *Arabidopsis* RPP1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. *Plant Cell.* 1998 Nov;10(11):1847–60.
25. Brandt SA, Spring A, Hiebsch C, Tabogie E, Diro M, Yntiso G, et al. The “tree Against Hunger”: Enset-based Agricultural System in Ethiopia. *American Association for the Advancement of Science*; 1997. 70 p.

26. Brueggeman R, Rostoks N, Kudrna D, Kilian A, Han F, Chen J, et al. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proc Natl Acad Sci USA*. 2002 Jul 9;99(14):9328–33.
27. Brugmans DW, Wouters D, van Os H, Hutten R, van der Linden G, Visser RGF, et al. Genetic mapping and transcription analyses of resistance gene loci in potato using NBS profiling. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik*. 2008;117(8):1379–88.
28. Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R. Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Molecular Ecology Notes*. 2005;5(2):327–30.
29. Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, et al. Incidence and origin of “null” alleles in the (AC)_n microsatellite markers. *Am J Hum Genet*. 1993 May;52(5):922–7.
30. Carvajal SLB. El cultivo del plátano (*Musa AAB* Simmonds) en el Trópico. Instituto Colombiano Agropecuario ICA, Subgerencia de Investigación, División Producción Cultivos, Sección Frutícolas-Plátano y Banano; 1991. 392 p.
31. Cenci A, Guignon V, Roux N, Rouard M. Genomic analysis of NAC transcription factors in banana (*Musa acuminata*) and definition of NAC orthologous groups for monocots and dicots. *Plant Mol Biol*. 2014 May;85(1-2):63–80.
32. Center for History and New Media. Guia de Inicio Rápido [Internet]. Available from: http://zotero.org/support/quick_start_guide
33. Cheesman EE. Classification of the Bananas: The Genus *Ensete* Horan. *Kew Bulletin*. 1947 Jan 1;2(2):97–106.
34. Chen X, Shang J, Chen D, Lei C, Zou Y, Zhai W, et al. A B-lectin receptor kinase gene conferring rice blast resistance. *Plant J*. 2006 Jun;46(5):794–804.
35. Cheung F, Town CD. A BAC end view of the *Musa acuminata* genome. *BMC Plant Biol*. 2007;7:29.
36. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, et al. Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. *Genome Res*. 2004 Jun 1;14(6):1147–59.
37. Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*. 2006 Feb 24;124(4):803–14.
38. Churchill ACL. *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Mol Plant Pathol*. 2011 May;12(4):307–28.

39. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*. 2005 Jan 1;142(1-2):169–96.
40. Collinge M, Boller T. Differential induction of two potato genes, *Stprx2* and *StNAC*, in response to infection by *Phytophthora infestans* and to wounding. *Plant Mol Biol*. 2001 Jul;46(5):521–9.
41. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005 Sep 15;21(18):3674–6.
42. Creste S, Neto AT, Figueira A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep*. 2001 Dec 1;19(4):299–306.
43. Creste S, Neto AT, Silva S de O, Figueira A. Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica*. 2003 Jul 1;132(3):259–68.
44. Cronquist A. The divisions and classes of plants. *Bot Rev*. 1960 Oct 1;26(4):425–82.
45. Crouch HK, Crouch JH, Jarret RL, Cregan PB, Ortiz R. Segregation at Microsatellite Loci in Haploid and Diploid Gametes of *Musa*. *Crop Science*. 1998;38(1):211.
46. D'Hont A, Denoeud F, Aury J-M, Baurens F-C, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*. 2012 Aug 9;488(7410):213–7.
47. Damalas CA, Eleftherohorinos IG. Pesticide exposure, safety issues, and risk assessment indicators. *Int J Environ Res Public Health*. 2011 May;8(5):1402–19.
48. Daniells J, Jenny C, Karamura DA, Sharrock,. *Musalogue: a catalogue of Musa germplasm. Diversity in the genus Musa*. Montpellier: INIPAB; 2001. 213 p.
49. Dantas JLL, Shepherd K, Silva SO, Soares Filho WS. Classificação botânica, origem, evolução e distribuição geográfica. In: ALVES EJ, editor. *A cultura da banana – aspectos técnicos, socioeconômicos e agroindustriais*. Cruz das Almas: Embrapa-CNPMPF; 1997. p. 27–34.
50. Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J. A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids. *BMC Genomics*. 2013 Oct 5;14(1):683.
51. De Wit PJGM. How plants recognize pathogens and defend themselves. *Cell Mol Life Sci*. 2007 Nov;64(21):2726–32.

52. Deller S, Hammond-Kosack KE, Rudd JJ. The complex interactions between host immunity and non-biotrophic fungal pathogens of wheat leaves. *J Plant Physiol.* 2011 Jan 1;168(1):63–71.
53. Desclos-Theveniau M, Arnaud D, Huang T-Y, Lin GJ-C, Chen W-Y, Lin Y-C, et al. The Arabidopsis Lectin Receptor Kinase LecRK-V.5 Represses Stomatal Immunity Induced by *Pseudomonas syringae* pv. tomato DC3000. *PLoS Pathog* [Internet]. 2012 Feb [cited 2013 Dec 2];8(2). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3276567/>
54. Deslandes L, Olivier J, Theulieres F, Hirsch J, Feng DX, Bittner-Eddy P, et al. Resistance to *Ralstonia solanacearum* in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. *Proc Natl Acad Sci USA.* 2002 Feb 19;99(4):2404–9.
55. Diener AC, Ausubel FM. Resistance to *Fusarium oxysporum* 1, a dominant Arabidopsis disease-resistance gene, is not race specific. *Genetics.* 2005 Sep;171(1):305–21.
56. Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JD. The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *Plant Cell.* 1998 Nov;10(11):1915–25.
57. Dodds PN, Lawrence GJ, Ellis JG. Six amino acid changes confined to the leucine-rich repeat beta-strand/beta-turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell.* 2001 Jan;13(1):163–78.
58. Donzelli BGG, Churchill ACL. A Quantitative Assay Using Mycelial Fragments to Assess Virulence of *Mycosphaerella fijiensis*. *Phytopathology.* 2007 Aug;97(8):916–29.
59. Droc G, Lariviere D, Guignon V, Yahiaoui N, This D, Garsmeur O, et al. The Banana Genome Hub. Database. 2013 May 23;2013(0):bat035–bat035.
60. Durroux V. New banana disease to Africa found in Mozambique [Internet]. RTB-CGIAR. 2013 [cited 2014 Feb 28]. Available from: <http://www.rtb.cgiar.org/new-banana-disease-to-africa-found-in-mozambique/>
61. Eckardt NA. Positive and Negative Feedback Coordinate Regulation of Disease Resistance Gene Expression. *Plant Cell.* 2007 Sep;19(9):2700–2.
62. El-basyoni I, Baenziger PS, Dweikat I, Wang D, Eskridge K, Saadalla M. Using DArT Markers to Monitor Genetic Diversity throughout Selection: A Case Study in Nebraska's Winter Wheat Breeding Nurseries. *Crop Science.* 2013;53(6):2363.
63. Elphinstone J. The current Bacterial Wilt situation: a global overview. In: Allen C, Prior P, Hayward AC, editors. *Bacterial Wilt Disease and the Ralstonia Solanaceum Species Complex.* St. Paul, MN: The American Phytopathological Society.; 2005. p. 9–28.
64. Ernst K, Kumar A, Kriseleit D, Kloos D-U, Phillips MS, Ganai MW. The broad-spectrum potato cyst nematode resistance gene (Hero) from tomato is the only member of a

large gene family of NBS-LRR genes with an unusual amino acid repeat in the LRR region. *Plant J.* 2002 Jul;31(2):127–36.

65. FAOSTAT. FAOSTAT Agriculture Data [Internet]. 2013. Available from: <http://faostat.fao.org>

66. Faux CE, McInnes JC, Jarman SN. High-throughput real time PCR and melt curve analysis for sexing southern ocean seabirds using faecal samples. *Theriogenology* [Internet].

67. Ferrari JT, Nogueira EM de C. Sigatoka Negra: como identificar e combater a Sigatoka Negra da bananeira [Internet]. Centro de Pesquisa e Desenvolvimento de Sanidade Vegetal; 2008

68. Filion M. Quantitative real-time PCR in applied microbiology. Norfolk, UK: Caister Academic; 2012. 257 p.

69. Fullerton RA. Sigatoka Leaf Diseases. In: Ploetz RC, Zentmyer GA, Nishijima WT, Rohrbach KG, Ohr HD, editors. *Compendium of Tropical Fruit Diseases*. St. Paul, MN: American Phytopathological Society; 1994. p. 12–4.

70. Gao Y, Guo W, Wang L, Zhang T. Isolation and characterization of resistance and defense gene analogs in cotton (*Gossypium barbadense* L.). *Sci China, C, Life Sci.* 2006 Dec;49(6):530–42.

71. Gao Y, Xu Z, Jiao F, Yu H, Xiao B, Li Y, et al. Cloning, structural features, and expression analysis of resistance gene analogs in tobacco. *Mol Biol Rep.* 2010 Jan;37(1):345–54.

72. Gassmann W, Hinsch ME, Staskawicz BJ. The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *Plant J.* 1999 Nov;20(3):265–77.

73. Gawel NJ, Jarret RL, Whittemore AP. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. *Theoret Appl Genetics.* 1992 Jul 1;84(3-4):286–90.

74. Ghazvini H, Hiebert CW, Thomas JB, Fetch T. Development of a multiple bulked segregant analysis (MBSA) method used to locate a new stem rust resistance gene (Sr54) in the winter wheat cultivar Norin 40. *Theor Appl Genet.* 2013 Feb 1;126(2):443–9.

75. Glazebrook J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol.* 2005;43:205–27.

76. Głowacki S, Macioszek VK, Kononowicz AK. R proteins as fundamentals of plant innate immunity. *Cell Mol Biol Lett.* 2011 Mar;16(1):1–24.

77. Gómez-Gómez L, Boller T. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell.* 2000 Jun;5(6):1003–11.

78. Gong C, Cao S, Fan R, Wei B, Chen G, Wang X, et al. Identification and Phylogenetic Analysis of a CC-NBS-LRR Encoding Gene Assigned on Chromosome 7B of Wheat. *Int J Mol Sci*. 2013 Jul 24;14(8):15330–47.
79. Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, et al. Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science*. 1995 Aug 11;269(5225):843–6.
80. Grant PBC, Woudneh MB, Ross PS. Pesticides in blood from spectacled caiman (*Caiman crocodilus*) downstream of banana plantations in Costa Rica. *Environ Toxicol Chem*. 2013 Nov;32(11):2576–83.
81. Grapin A, Noyer J-L, Carreel F, Dambier D, Baurens F-C, Lanaud C, et al. Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *ELECTROPHORESIS*. 1998;19(8-9):1374–80.
82. Grushetskaya ZE, Lemesh VA, Poliksenova VD, Khotyleva LV. Mapping of the Cf-6 tomato leaf mould resistance locus using SSR markers. *Russ J Genet*. 2007 Nov 1;43(11):1266–70.
83. Gu K, Yang B, Tian D, Wu L, Wang D, Sreekala C, et al. R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*. 2005 Jun 23;435(7045):1122–5.
84. Guo Z, Bonos S, Meyer WA, Day PR, Belanger FC. Transgenic creeping bentgrass with delayed dollar spot symptoms. *Molecular Breeding*. 2003 Feb 1;11(2):95–101.
85. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics*. 2003 Dec;270(4):315–23.
86. Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW. Plant disease resistance genes: Current status and future directions. *Physiological and Molecular Plant Pathology*. 2012 Apr;78:51–65.
87. Häkkinen M. Reappraisal of sectional taxonomy in *Musa* (Musaceae). *Taxon*. 2013 Aug 21;62(4):809–13.
88. Hammond-Kosack KE, Kanyuka K. Resistance Genes (R Genes) in Plants. eLS; 2001.
89. Hammond-Kosack KE, Jones JDG. Plant disease resistance genes. *Annu Rev Plant Physiol Plant Mol Biol*. 1997 Jun;48:575–607.
90. Harding RM, Boyce AJ, Clegg JB. The evolution of tandemly repetitive DNA: recombination rules. *Genetics*. 1992 Nov;132(3):847–59.

91. Hayashi Y, Toyomasu T, Hirose Y, Onodera Y, Mitsunashi W, Yamane H, et al. Comparison of the enzymatic properties of ent-copalyl diphosphate synthases in the biosynthesis of phytoalexins and gibberellins in rice. *Biosci Biotechnol Biochem*. 2008 Feb;72(2):523–30.
92. Hegedus D, Yu M, Baldwin D, Gruber M, Sharpe A, Parkin I, et al. Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress. *Plant Mol Biol*. 2003 Oct;53(3):383–97.
93. Heslop-Harrison JS, Schwarzacher T. Domestication, genomics and the future for banana. *Ann Bot*. 2007 Nov;100(5):1073–84.
94. Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci A-M, et al. A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biol*. 2010;10:65.
95. Holland JB, Helland SJ, Sharopova N, Rhyne DC. Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and introns and repeat sequences in oat. *Genome*. 2001 Dec;44(6):1065–76.
96. Hoorn RAL van der, Kamoun S. From Guard to Decoy: A New Model for Perception of Plant Pathogen Effectors. *Plant Cell*. 2008 Aug 1;20(8):2009–17.
97. Hrahsel L, Basu A, Sahoo L, Thangjam R. In Vitro Propagation and Assessment of the Genetic Fidelity of *Musa acuminata* (AAA) cv. Vaibalhla Derived from Immature Male Flowers. *Appl Biochem Biotechnol*. 1–10.
98. Hřibov E, Christelov P, Roux N, Doležel J. A simple and robust approach for genotyping in Musaceae. *Acta Hort*. 2013;986(1):241–246.
99. Hřibová E, Doležel J. Different approaches to discover the repetitive part of the banana nuclear genome. *Acta Hort*. 2011;897:63–8.
100. Hwang T-Y, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, et al. High-density Integrated Linkage Map Based on SSR Markers in Soybean. *DNA Res*. 2009 Aug 1;16(4):213–25.
101. IBGE. Produção Agrícola Municipal: culturas temporárias e permanentes. Instituto Brasileiro de Geografia e Estatística. 2011;38:1–97.
102. INIBAP. Networking Banana and Plantain. Bioversity International; 2000. 64 p.
103. Johal GS, Briggs SP. Reductase activity encoded by the HM1 disease resistance gene in maize. *Science*. 1992 Nov 6;258(5084):985–7.
104. Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Wiel C van de, et al. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*. 1997 Oct 1;3(5):381–90.

105. Jones JDG, Dangl JL. The plant immune system. *Nature*. 2006 Nov 16;444(7117):323–9.
106. Joshi BK, Panthee DR, Louws FJ, Yencho GC, Sosinski B, Arellano C. RAPD Markers Linked to Late Blight Resistance in Tomato. *Nepal Journal of Science and Technology*. 2013 Oct 10 ;14(1).
107. Joshi RK, Kar B, Nayak S. Survey and characterization of NBS-LRR (R) genes in *Curcuma longa* transcriptome. *Bioinformatics*. 2011 Jul 19;6(9):360–3.
108. Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJ, Wright F, et al. Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics*. 2012 Feb 15;13(1):75.
109. Kaemmer D, Fischer D, Jarret RL, Baurens F-C, Grapin A, Dambier D, et al. Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. *Euphytica*. 1997 Jul 1;96(1):49–63.
110. Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK. Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. 2011 Feb 1;177(3):309–34.
111. Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. *BMC Plant Biology*. 2012 Aug 9;12(1):139.
112. Karakousis A, Barr AR, Chalmers KJ, Ablett GA, Holton TA, Henry RJ, et al. Potential of SSR markers for plant breeding and variety identification in Australian barley germplasm. *Aust J Agric Res*. 2003 Jan 1;54(12):1197–210.
113. Kawchuk LM, Hachey J, Lynch DR, Kulcsar F, Rooijen G van, Waterer DR, et al. Tomato Ve disease resistance genes encode cell surface-like receptors. *PNAS*. 2001 May 22;98(11):6511–5.
114. Kema GHJ. The pesticide reduction plan for banana. *Anais XVII Reunião Internacional ACORBAT*. Joinville, Santa Catarina: Soprano, E., Tcacenco, F.A., Lichtemberg, L.A. and Silva, M.C.; 2006. p. 3–4.
115. Kim E Hammond–Kosack KK. *Resistance Genes (R Genes) in Plants*. 2007;
116. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature*. 1995 Mar 9;374(6518):183–6.
117. Koeck M, Hardham AR, Dodds PN. The role of effectors of biotrophic and hemibiotrophic fungi in infection. *Cell Microbiol*. 2011 Dec;13(12):1849–57.
118. Kolpakov R, Bana G, Kucherov G. mreps: efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res*. 2003 Jul 1;31(13):3672–8.

119. Lagercrantz U, Ellegren H, Andersson L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* 1993 Mar 11;21(5):1111–5.
120. Lagoda PJ, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C. Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Mol Ecol.* 1998 May;7(5):659–63.
121. Lapin D, Van den Ackerveken G. Susceptibility to plant disease: more than a failure of host immunity. *Trends Plant Sci.* 2013 Oct;18(10):546–54.
122. Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG. The L6 gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene RPS2 and the tobacco viral resistance gene N. *The Plant cell.* 1995;7(8):1195–206.
123. Lescot M, Piffanelli P, Ciampi AY, Ruiz M, Blanc G, Leebens-Mack J, et al. Insights into the Musa genome: Syntenic relationships to rice and between Musa species. *BMC Genomics.* 2008 Jan 30;9(1):58.
124. Li C, Shao J, Wang Y, Li W, Guo D, Yan B, et al. Analysis of banana transcriptome and global gene expression profiles in banana roots in response to infection by race 1 and tropical race 4 of *Fusarium oxysporum* f. sp. *cubense*. *BMC Genomics.* 2013;14:851.
125. Li C, Chen S, Zuo C, Sun Q, Ye Q, Yi G, et al. The use of GFP-transformed isolates to study infection of banana with *Fusarium oxysporum* f. sp. *cubense* race 4. *Eur J Plant Pathol.* 2011 Oct 1;131(2):327–40.
126. Librelon SS, Costa MR, Nietzsche S, Pereira MCT. Genetic diversity of “Prata Anã” (AAB) banana tree clones by using SSR markers. *Revista Brasileira de Fruticultura.* 2013 Sep;35(3):809–17.
127. Lin C-Y, Trinh NN, Lin C-W, Huang H-J. Transcriptome analysis of phytohormone, transporters and signaling pathways in response to vanadium stress in rice roots. *Plant Physiol Biochem.* 2013 May;66:98–104.
128. Lin R, Zhao W, Meng X, Wang M, Peng Y. Rice gene OsNAC19 encodes a novel NAC-domain transcription factor and responds to infection by *Magnaporthe grisea*. *Plant Science.* 2007 Jan;172(1):120–30.
129. Lister R, Gregory BD, Ecker JR. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol.* 2009 Apr;12(2):107–18.
130. Liu G, Kennedy R, Greenshields DL, Peng G, Forseille L, Selvaraj G, et al. Detached and attached Arabidopsis leaf assays reveal distinctive defense responses against hemibiotrophic *Colletotrichum* spp. *Mol Plant Microbe Interact.* 2007 Oct;20(10):1308–19.
131. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* 2005 May 1;21(9):2128–9.

132. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol.* 2012;2012:251364.
133. Liu P, Goh C-J, Loh C-S, Pua E-C. Differential expression and characterization of three metallothionein-like genes in Cavendish banana (*Musa acuminata*). *Physiol Plant.* 2002 Feb;114(2):241–50.
134. Loh JP, Kiew R, Set O, Gan LH, Gan YY. Amplified fragment length polymorphism fingerprinting of 16 banana cultivars (*Musa* cvs.). *Mol Phylogenet Evol.* 2000 Dec;17(3):360–6.
135. Lozano-Torres JL, Wilbers RHP, Gawronski P, Boshoven JC, Finkers-Tomczak A, Cordewener JHG, et al. Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proc Natl Acad Sci USA.* 2012 Jun 19;109(25):10119–24.
136. Macheroux P, Schmid J, Amrhein N, Schaller A. A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway. *Planta.* 1999 Jan;207(3):325–34.
137. Marangoni M, Girotto L, Nunes M, Almeida W, Galbieri R, Schuster I, et al. Search for a Microsatellite Marker Linked with Resistance Gene to *Xanthomonas axonopodis* pv. *malvacearum* in Brazilian Cotton. *American Journal of Plant Sciences.* 2013;4(10):2039–42.
138. Marcel Scheideler NLS. Monitoring the switch from housekeeping to pathogen defense metabolism in *Arabidopsis thaliana* using cDNA arrays. *The Journal of biological chemistry.* 2002;277(12):10555–61.
139. Marone D, Russo MA, Laidò G, De Leonardis AM, Mastrangelo AM. Plant Nucleotide Binding Site-Leucine-Rich Repeat (NBS-LRR) Genes: Active Guardians in Host Defense Responses. *Int J Mol Sci.* 2013;14(4):7302–26.
140. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganai MW, Spivey R, et al. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science.* 1993 Nov 26;262(5138):1432–6.
141. Martin GB, Bogdanove AJ, Sessa G. Understanding the Functions of Plant Disease Resistance Proteins. *Annual Review of Plant Biology.* 2003;54(1):23–61.
142. Martins WS, Lucas DCS, Neves KF de S, Bertioli DJ. WebSat--a web software for microsatellite marker development. *Bioinformatics.* 2009;3(6):282–3.
143. McDowell JM, Dhandaydham M, Long TA, Aarts MG, Goff S, Holub EB, et al. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. *Plant Cell.* 1998 Nov;10(11):1861–74.
144. McDowell JM, Woffenden BJ. Plant disease resistance genes: recent insights and potential applications. *Trends in Biotechnology.* 2003 Apr;21(4):178–83.

145. Meksem K, Kahl G. The handbook of plant genome mapping: genetic and physical mapping. Weinheim; [Great Britain]: Wiley-VCH; 2005. 402 p.
146. Menendez T, Shepherd K. Breeding new bananas. *World Crops*. 1975;27(3):104–12.
147. Meredith DS. Banana Leaf Spot Disease (Sigatoka) Caused by *Mycosphaerella Musicola* Leach. England: Commonwealth Mycological Institute; 1970. 7 p.
148. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-Wide Analysis of NBS-LRR-Encoding Genes in *Arabidopsis*. *Plant Cell*. 2003 Apr 1;15(4):809–34.
149. Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW. Receptor-like Genes in the Major Resistance Locus of Lettuce Are Subject to Divergent Selection. *Plant Cell*. 1998 Nov 1;10(11):1833–46.
150. Miller RN, Bertioli DJ, Baurens FC, Santos CM, Alves PC, Martins NF, et al. Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: Isolation, RFLP marker development, and physical mapping. *BMC Plant Biology*. 2008 Jan 30;8(1):15.
151. Miller RN, Passos MA, Menezes NN, Souza MT, Costa MM do C, Azevedo VCR, et al. Characterization of novel microsatellite markers in *Musa acuminata* subsp. *burmannicoides*, var. *Calcutta 4*. *BMC Research Notes*. 2010 May 27;3(1):148.
152. Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM. The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. *Plant Cell*. 1998 Aug;10(8):1307–19.
153. Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, et al. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding*. 1997 Apr 1;3(2):87–103.
154. Mohr TJ, Mammarella ND, Hoff T, Woffenden BJ, Jelesko JG, McDowell JM. The *Arabidopsis* downy mildew resistance gene *RPP8* is induced by pathogens and salicylic acid and is regulated by *W* box cis elements. *Mol Plant Microbe Interact*. 2010 Oct;23(10):1303–15.
155. Monaco MK, Stein J, Naithani S, Wei S, Dharmawardhana P, Kumari S, et al. *Gramene 2013: comparative plant genomics resources*. *Nucleic Acids Res*. 2014 Jan;42(Database issue):D1193–1199.
156. Moreau M, Tian M, Klessig DF. Salicylic acid binds NPR3 and NPR4 to regulate NPR1-dependent defense responses. *Cell Res*. 2012 Dec;22(12):1631–3.
157. Mourichon, Fullerton RA. Geographical distribution of the two species *Mycosphaerella musicola* Leach (*Cercospora musae*) and *M. fijiensis* Morelet (*C. fijiensis*), respectively agents of Sigatoka Disease and Black Leaf Streak Disease in bananas and plantains. *Fruits*. 1990;45:213–8.

158. Mudalkar S, Golla R, Ghattay S, Reddy AR. De novo transcriptome analysis of an imminent biofuel crop, *Camelina sativa* L. using Illumina GAII-X sequencing platform and identification of SSR markers. *Plant Mol Biol*. 2014 Jan 1;84(1-2):159–71.
159. Mukunthakumar S, Padmesh P, Vineesh PS, Skaria R, Kumar KH, Krishnan PN. Genetic diversity and differentiation analysis among wild antecedents of banana (*Musa acuminata* Colla) using RAPD markers. *NISCAIR-CSIR*. 2013 Oct;12(4):493–8.
160. Muthamilarasan M, Prasad M. Plant innate immunity: an updated insight into defense mechanism. *J Biosci*. 2013 Jun;38(2):433–49.
161. Nakashima K, Tran L-SP, Van Nguyen D, Fujita M, Maruyama K, Todaka D, et al. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. *Plant J*. 2007 Aug;51(4):617–30.
162. Namukwaya B, Tripathi L, Tripathi JN, Arinaitwe G, Mukasa SB, Tushemereirwe WK. Transgenic banana expressing Pflp gene confers enhanced resistance to *Xanthomonas* wilt disease. *Transgenic Res*. 2012 Aug;21(4):855–65.
163. Ning S-P, Xu L-B, Lu Y, Huang B-Z, Ge X-J. Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. *Scientia Horticulturae*. 2007 Nov 20;114(4):281–8.
164. Norman DJ, Zapata M, Gabriel DW, Duan YP, Yuen JMF, Mangravita-Novo A, et al. Genetic Diversity and Host Range Variation of *Ralstonia solanacearum* Strains Entering North America. *Phytopathology*. 2009 Sep;99(9):1070–7.
165. Nwakanma DC, Pillay M, Okoli BE, Tenkouano A. PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in *Musa* L. *Theor Appl Genet*. 2003 Dec 1;108(1):154–9.
166. Olanmi B, Akoroda MO, Okogbenin E, Egesi C, Kahya SS, Onyegbule O, et al. Bulk segregant analysis identifies molecular markers associated with early bulking in cassava (*Manihot esculenta* Crantz). *Euphytica*. 2014 Jan 1;195(2):235–44.
167. Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell*. 2000 Nov;12(11):2019–32.
168. Ori N, Eshed Y, Paran I, Presting G, Aviv D, Tanksley S, et al. The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. *Plant Cell*. 1997 Apr;9(4):521–32.
169. Parker JE, Coleman MJ, Szabò V, Frost LN, Schmidt R, Biezen EA van der, et al. The Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. *Plant Cell*. 1997 Jun 1;9(6):879–94.

170. Passos MAN, Cruz V de O, Emediato FL, Teixeira C de C, Souza MT, Matsumoto T, et al. Development of expressed sequence tag and EST-SSR marker resources for *Musa acuminata*. *AoB PLANTS* [Internet]. 2012 Sep 26 [cited 2013 Jun 10]; Available from: <http://aobpla.oxfordjournals.org/content/early/2012/09/26/aobpla.pls030>
171. Passos MAN, Cruz VO de, Emediato FL, Teixeira CC de, Azevedo VCR, Brasileiro ACM, et al. Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. *BMC Genomics*. 2013 Feb 5;14(1):78.
172. Pavla Christelová MV. A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB plants*. 2011;2011:plr024.
173. Peeters N, Guidot A, Vailleau F, Valls M. *Ralstonia solanacearum*, a widespread bacterial plant pathogen in the post-genomic era. *Mol Plant Pathol*. 2013 Sep;14(7):651–62.
174. Pemberton JM, Slate J, Bancroft DR, Barrett JA. Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol Ecol*. 1995 Apr;4(2):249–52.
175. Pennisi E. Armed and dangerous. *Science*. 2010 Feb 12;327(5967):804–5.
176. Peraza-Echeverria S, Dale JL, Harding RM, Smith MK, Collet C. 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. *Mol Breeding*. 2008 Nov 1;22(4):565–79.
177. Petre B, Major I, Rouhier N, Duplessis S. Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar. *BMC Plant Biology*. 2011 Feb 15;11(1):33.
178. Petutschnig EK, Jones AME, Serazetdinova L, Lipka U, Lipka V. The Lysin Motif Receptor-like Kinase (LysM-RLK) CERK1 Is a Major Chitin-binding Protein in *Arabidopsis thaliana* and Subject to Chitin-induced Phosphorylation. *J Biol Chem*. 2010 Sep 10;285(37):28902–11.
179. Pfaffl M. *Livestock Transcriptomics: Quantitative MRNA Analytics in Molecular Endocrinology and Physiology*. 2003. book p.
180. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucl Acids Res*. 2002 May 1;30(9):e36–e36.
181. Pieterse CMJ, Pelt JA van, Wees ACM van, Ton J, Verhagen BWM, Léon-Kloosterziel K, et al. Indução de resistência sistêmica por rizobactérias e comunicação na rota de sinalização para uma defesa refinada. *Revisão anual de Patologia de Plantas*. 2005;13:277–319.

182. Pieterse CMJ, Leon-Reyes A, Van der Ent S, Van Wees SCM. Networking by small-molecule hormones in plant immunity. *Nat Chem Biol.* 2009 May;5(5):308–16.
183. Piffanelli P, Vilarinhos AD, Safar J, Sabau X, Dolezel J. Construction of bacterial artificial chromosome (BAC) libraries of banana (*Musa acuminata* and *Musa balbisiana*). *Fruits.* 2008;63(06):375–9.
184. Portal O, Izquierdo Y, De Vleeschauwer D, Sánchez-Rodríguez A, Mendoza-Rodríguez M, Acosta-Suárez M, et al. Analysis of expressed sequence tags derived from a compatible *Mycosphaerella fijiensis*-banana interaction. *Plant Cell Rep.* 2011 May;30(5):913–28.
185. PROMEGA. GenePrint® STR Systems (Silver Stain Detection). 2006.
186. Qi Y, Tsuda K, Glazebrook J, Katagiri F. Physical association of pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) immune receptors in *Arabidopsis*. *Mol Plant Pathol.* 2011 Sep;12(7):702–8.
187. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010 Mar 15;26(6):841–2.
188. Ramos HC, Rumbo M, Sirard J-C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol.* 2004 Nov;12(11):509–17.
189. Ravishankar KV LV. Development of SSR markers based on a survey of genomic sequences and their molecular analysis in banana (*Musa* spp.). *Journal of Horticultural Science and Biotechnology.* 2012;87:84–8.
190. Rekha A, Ravishankar KV, Ambika DS. Generation of segregating F1 populations for mapping the *Musa* genome. *Acta Hort.* 2011;897:277–8.
191. Ron M, Avni A. The Receptor for the Fungal Elicitor Ethylene-Inducing Xylanase Is a Member of a Resistance-Like Gene Family in Tomato. *Plant Cell.* 2004 Jun;16(6):1604–15.
192. Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM. The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. *Proc Natl Acad Sci USA.* 1998 Aug 18;95(17):9750–4.
193. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol.* 2000;132:365–86.
194. Sallé G, Pichard V, Mourichon X. Cytological studies of the interaction between *Mycosphaerella fijiensis* Morelet and three cultivars of *Musa* presenting different levels of resistance. In: Fullerton R, Stover R, editors. 'Sigatoka leaf spot diseases of bananas: Proceedings of an international workshop held at San José, Costa Rica. San José, Costa Rica: INIBAP; 1989. p. 180–90.

195. Samarasinghe WLG, Wickramasinghe IP, Rajapakse S, Perera ALT. Morphological and Molecular Characterization of Musa Germplasm in Sri Lanka and Selection of Superior Genotypes. *Acta Horticulturae*. 2010;879:571–6.
196. Santos CMR, Martins NF, Hörberg HM, Almeida ERP de, Coelho MCF, Togawa RC, et al. Analysis of expressed sequence tags from *Musa acuminata* ssp. *burmannicoides*, var. *Calcutta 4* (AA) leaves submitted to temperature stresses. *Theor Appl Genet*. 2005 May 1;110(8):1517–22.
197. Saraiva L de A, Castelan FP, Shitakubo R, Hassimotto NMA, Purgatto E, Chillet M, et al. Black leaf streak disease affects starch metabolism in banana fruit. *J Agric Food Chem*. 2013 Jun 12;61(23):5582–9.
198. Shah J. The salicylic acid loop in plant defense. *Curr Opin Plant Biol*. 2003 Aug;6(4):365–71.
199. Shan W, Kuang J, Chen L, Xie H, Peng H, Xiao Y, et al. Molecular characterization of banana NAC transcription factors and their interactions with ethylene signalling component EIL during fruit ripening. *J Exp Bot*. 2012 Sep;63(14):5171–87.
200. Shang J, Tao Y, Chen X, Zou Y, Lei C, Wang J, et al. Identification of a new rice blast resistance gene, *Pid3*, by genomewide comparison of paired nucleotide-binding site--leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics*. 2009 Aug;182(4):1303–11.
201. Sharrock, Frison. *Musa* production around the world - trends, varieties and regional importance. INIBAP Annual Report 1998. Montpellier.: INIBAP; 1998. p. 42–47.
202. Shekhawat UKS, Ganapathi TR. *Musa*WRKY71 Overexpression in Banana Plants Leads to Altered Abiotic and Biotic Stress Responses. *PLoS ONE*. 2013 Oct 8;8(10):e75506.
203. Shewry PR, Lucas JA. *Plant Proteins that Confer Resistance to Pests and Pathogens*. Academic Press; 1997. 192 p.
204. Silva S de O, Morais L., Santos-Serejo J. Melhoramento genético de bananeira para resistência a doenças. Recursos genéticos vegetais no Estado da Bahia. Feira de Santana: ROMÃO, R.L.; RAMOS, S.R.R; 2005. p. 49–67.
205. Silva S de O e, Amorim EP, Santos-Serejo JA dos, Ferreira CF, Rodriguez MAD. Genetic improvement of banana: strategies and available technologies. *Revista Brasileira de Fruticultura*. 2013 Sep;35(3):919–31.
206. Simmonds NW, Shepherd K. The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society of London, Botany*. 1955;55(359):302–12.
207. Simmonds NW. *The Evolution of the Bananas*. Longmans; 1962. 192 p.

208. Small RL, Wendel JF. Copy number lability and evolutionary dynamics of the Adh gene family in diploid and tetraploid cotton (*Gossypium*). *Genetics*. 2000 Aug;155(4):1913–26.
209. Solovyev VV, Salamov AA, Lawrence CB. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res*. 1994 Dec 11;22(24):5156–63.
210. Song WY, Pi LY, Wang GL, Gardner J, Holsten T, Ronald PC. Evolution of the rice Xa21 disease resistance gene family. *Plant Cell*. 1997 Aug;9(8):1279–87.
211. Ssekiwoko F, Taligoola HK& T. *Xanthomonas campestris* pv *musacearum* Host Range in Uganda. *African Crop Science Journal*. 14(2):111–20.
212. Stover RH. Sigatoka Leaf Spots of Bananas and. *Plant Disease*. 1980;64(8):750.
213. Stover RH, Simmonds NW. *Bananas*. Harlow: Longman Scientific & Technical,; 1987.
214. Swarupa V, Ravishankar KV, Rekha. Characterization of tolerance to *Fusarium oxysporum* f.sp., cubense infection in banana using suppression subtractive hybridization and gene expression analysis. *Physiological and Molecular Plant Pathology*. 2013;83(1):1–7.
215. Tachida H, Iizuka M. Persistence of repeated sequences that evolve by replication slippage. *Genetics*. 1992 Jun;131(2):471–8.
216. Tai TH, Dahlbeck D, Clark ET, Gajiwala P, Pasion R, Whalen MC, et al. Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. *PNAS*. 1999 Nov 23;96(24):14153–8.
217. Tan L-Q, Wang L-Y, Wei K, Zhang C-C, Wu L-Y, Qi G-N, et al. Floral Transcriptome Sequencing for SSR Marker Development and Linkage Map Construction in the Tea Plant (*Camellia sinensis*). *PLoS ONE*. 2013 Nov 26;8(11):e81611.
218. Tan X, Meyers BC, Kozik A, West MA, Morgante M, St Clair DA, et al. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biol*. 2007 Oct 23;7:56.
219. Tao Y, Xie Z, Chen W, Glazebrook J, Chang H-S, Han B, et al. Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell*. 2003 Feb;15(2):317–30.
220. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 2003 Sep 11;4:41.
221. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.):

frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 2001 Aug;11(8):1441–52.

222. Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, et al. Characterization of the tomato Cf-4 gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *Plant Cell.* 1997 Dec;9(12):2209–24.

223. Thomas D, Wu CKW. GMAP: a genomic mapping and alignment program for mRNA and EST sequence. *Bioinformatics.* 2005;21:1859–75.

224. Thomma BP, Penninckx IA, Broekaert WF, Cammue BP. The complexity of disease signaling in *Arabidopsis*. *Curr Opin Immunol.* 2001 Feb;13(1):63–8.

225. Thomma BPHJ, Nürnberger T, Joosten MHJ. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell.* 2011 Jan;23(1):4–15.

226. Tor M, Lotze MT, Holton N. Receptor-mediated signalling in plants: molecular patterns and programmes. *J Exp Bot.* 2009 Sep;60(13):3645–54.

227. Tripathi L, Mwangi M, Abele S, Aritua V, Tushemereirwe WK, Bandyopadhyay R. *Xanthomonas Wilt: A Threat to Banana Production in East and Central Africa.* *Plant Disease.* 2009 May;93(5):440–51.

228. Truong HTH, Choi H, Cho MC, Lee HE, Kim JH. Use of Cf-9 gene-based markers in marker-assisted selection to screen tomato cultivars with resistance to *Cladosporium fulvum*. *Hortic Environ Biotechnol.* 2011 Apr 1;52(2):204–10.

229. Tsuda K, Katagiri F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr Opin Plant Biol.* 2010 Aug;13(4):459–65.

230. Ude G, Pillay M, Nwakanma D, Tenkouano A. Genetic Diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP Markers. *Theor Appl Genet.* 2002 Jun;104(8):1246–52.

231. Vajna L. Phytopathogenic *Fusarium oxysporum* Schlecht, as a Necrotrophic Mycoparasite. *Journal of Phytopathology.* 1985;114(4):338–47.

232. Van der Biezen EA, Freddie CT, Kahn K, Parker JE, Jones JDG. *Arabidopsis* RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *Plant J.* 2002 Feb;29(4):439–51.

233. Van der Vossen EA, van der Voort JN, Kanyuka K, Bendahmane A, Sandbrink H, Baulcombe DC, et al. Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. *Plant J.* 2000 Sep;23(5):567–76.

234. Van der Biezen EA, Jones JD. Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem Sci.* 1998 Dec;23(12):454–6.

235. Wachholtz M, Heng-Moss T, Twigg P, Baird L, Lu G, Amundsen K. Transcriptome analysis of two buffalograss cultivars. *BMC Genomics*. 2013;14:613.
236. Wang J-Y, Chen Y-Y, Liu W-L, Wu Y-T. [Development and application of EST-derived SSR markers for bananas (*Musa nana* Lour.)]. *Yi Chuan*. 2008 Jul;30(7):933–40.
237. Wang LX, Li HB, Gu TC, Liu LH, Pang BS, Qiu J, et al. Assessment of wheat variety stability using SSR markers. *Euphytica*. 2014 Feb 1;195(3):435–52.
238. Wang L, Zhu W, Fang L, Sun X, Su L, Liang Z, et al. Genome-wide identification of WRKY family genes and their response to cold stress in *Vitis vinifera*. *BMC Plant Biology*. 2014 Apr 22;14(1):103.
239. Wang X-L, Chiang T-Y, Roux N, Hao G, Ge X-J. Genetic diversity of wild banana (*Musa balbisiana* Colla) in China as revealed by AFLP markers. *Genet Resour Crop Evol*. 2007 Aug 1;54(5):1125–32.
240. Wang X, Hadrami AE, Adam LR, Daayf F. Local and distal gene expression of pr-1 and pr-5 in potato leaves inoculated with isolates from the old (US-1) and the new (US-8) genotypes of *Phytophthora infestans* (Mont.) de Bary. *Environmental and Experimental Botany*. 2006 Aug;57(1–2):70–9.
241. Wang Z, Zhang J, Jia C, Liu J, Li Y, Yin X, et al. De Novo characterization of the banana root transcriptome and analysis of gene expression under *Fusarium oxysporum* f. sp. *Cubense* tropical race 4 infection. *BMC Genomics*. 2012 Nov 21;13(1):650.
242. Wang Z-X, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, et al. The Pib gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *The Plant Journal*. 1999;19(1):55–64.
243. Wani SH. Inducing Fungus-Resistance into Plants through Biotechnology. *Notulae Scientia Biologicae*. 2010 Jun 13;2(2):14–21.
244. Warren RF, Henk A, Mowery P, Holub E, Innes RW. A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene RPS5 partially suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell*. 1998 Sep;10(9):1439–52.
245. Weaver LM, Herrmann KM. Dynamics of the shikimate pathway in plants. *Trends in Plant Science*. 1997 Sep;2(9):346–51.
246. Wong C, Kiew R, Loh JP, Gan LH, Set O, Lee SK, et al. Genetic Diversity of the Wild Banana *Musa acuminata* Colla in Malaysia as Evidenced by AFLP. *Ann Bot*. 2001 Dec 1;88(6):1017–25.
247. Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, et al. Broad-Spectrum Mildew Resistance in *Arabidopsis thaliana* Mediated by RPW8. *Science*. 2001 Jan 5;291(5501):118–20.

248. Xu ML, Melchinger AE, Xia XC, Lübberstedt T. High-resolution mapping of loci conferring resistance to sugarcane mosaic virus in maize using RFLP, SSR, and AFLP markers. *Mol Gen Genet*. 1999 Apr 1;261(3):574–81.
249. Yang S, Zhang X, Yue J-X, Tian D, Chen J-Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Genet Genomics*. 2008 Sep;280(3):187–98.
250. Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang Z-X, Kono I, et al. Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *PNAS*. 1998 Feb 17;95(4):1663–8.
251. YOUNG M, CONIE J. Life Cycle of (*Mycosphaerella fijiensis*) and Epidemiology of Black Sigatoka Diseases. In: Thomas-Louisy L, editor. IICA Report on Caribbean Black Sigatoka Management Workshop. Kingston; 2001. p. 116.
252. Zdobnov EM, Apweiler R. InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*. 2001 Sep 1;17(9):847–8.
253. Zhang J, Zhou J-M. Plant immunity triggered by microbial molecular signatures. *Mol Plant*. 2010 Sep;3(5):783–93.
254. Zhao S, Fernald RD. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J Comput Biol*. 2005 Oct;12(8):1047–64.
255. Zheng Y, Li C, Liu Y, Yan C, Zhang T, Zhuang W, et al. Cloning and Characterization of a NBS-LRR Resistance Gene from Peanut (*Arachis hypogaea* L.? *Journal of Agricultural Science*. 2012 Nov 11;4(12):p243.
256. Zhou JM, Trifa Y, Silva H, Pontier D, Lam E, Shah J, et al. NPR1 differentially interacts with members of the TGA/OBF family of transcription factors that bind an element of the PR-1 gene required for induction by salicylic acid. *Mol Plant Microbe Interact*. 2000 Feb;13(2):191–202.
257. Zipfel C. Pattern-recognition receptors in plant innate immunity. *Curr Opin Immunol*. 2008 Feb;20(1):10–6.
258. Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JDG, Boller T, et al. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell*. 2006 May 19;125(4):749–60.
259. Zvereva AS, Pooggin MM. Silencing and Innate Immunity in Plant Defense Against Viral and Non-Viral Pathogens. *Viruses*. 2012 Oct 29;4(11):2578–97.

ANEXOS

ANEXO 1: Artigos publicados em periódicos científicos

1. Passos, MAN; Cruz, VO; Emediato, FL; Teixeira CC, Azevedo, VCR; Brasileiro, ACM; Amorim, EP; Ferreira, CF; Martins, NF; Togawa, RC; Pappas, GJ; Silva, OB; and Miller, RNG. 2013. Analysis of the leaf transcriptome of *M. acuminata* during interaction with *M. musicola*: gene assembly, annotation and marker development. BMC Genomics 14: 78. doi:10.1186/1471-2164-14-78
2. F. L. Emediato, M. A. N. Passos, C. de Camargo Teixeira, G. J. Pappas Jr, R. N. G. Miller 2013. Analysis of expression of NBS-LRR Resistance Gene Analogs in *M. acuminata* during compatible and incompatible interactions with *M. musicola*. Acta Hort. 986: 255-258.
3. Passos, M.A.N; Emediato, F. L; Cruz, V. O; Teixeira, CC; Figueiredo, L. F. de A; Martins, N. F; Togawa, R. C; Costa, M. M. C; Silva Jr, O; Pappas Jr. G. J.; Miller, R. N. G. 2013. Understanding Plant Immunity: Transcriptome Profiling in *Musa*-Pathogen Interactions using Next Generation Sequencing. Acta Hort. 986: 227-240.
4. M. A. N. Passos, V. de Oliveira Cruz, F. L. Emediato, C. de Camargo Teixeira, M. T. Souza Jr, T. Matsumoto, V. C. Rennó Azevedo, C. F. Ferreira, E. P. Amorim, L. Flavio de Alencar Figueiredo, N. F. Martins, M. de Jesus Barbosa Cavalcante, F.-C. Baurens, O. Bonfim da Silva Jr, G. J. Pappas Jr, L. Pignolet, C. Abadie, A. Y. Ciampi, P. Piffanelli and R. N. G. Miller. 2012. Development of expressed sequence tag and EST-SSR marker resources for *M. acuminata*. AoB Plants - Advance Access published September 26, 2012. doi: 10.1093/aobpla/pls030
5. R.N.G. Miller, M.A.N. Passos, F.L. Emediato, C. de Camargo Teixeira, G.J. Pappas Júnior. 2011. Candidate Resistance Gene Discovery: Resistance Gene Analog Characterization and Differential Gene Expression Analysis in *Musa-Mycosphaerella* Host-Pathogen Interactions. Acta Hort. (ISHS) 897: 179-186.

RESEARCH ARTICLE

Open Access

Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development

Marco A N Passos¹, Viviane Oliveira de Cruz¹, Flavia L Emediato¹, Cristiane Camargo de Teixeira², Vânia C Rennó Azevedo³, Ana C M Brasileiro³, Edson P Amorim⁴, Claudia F Ferreira⁴, Natalia F Martins³, Roberto C Togawa³, Georgios J Pappas Júnior¹, Orzenil Bonfim da Silva Jr³ and Robert NG Miller^{1*}

Abstract

Background: Although banana (*Musa* sp.) is an important edible crop, contributing towards poverty alleviation and food security, limited transcriptome datasets are available for use in accelerated molecular-based breeding in this genus. 454 GS-FLX Titanium technology was employed to determine the sequence of gene transcripts in genotypes of *Musa acuminata* ssp. *burmannicoides* Calcutta 4 and *M. acuminata* subgroup Cavendish cv. Grande Naine, contrasting in resistance to the fungal pathogen *Mycosphaerella musicola*, causal organism of Sigatoka leaf spot disease. To enrich for transcripts under biotic stress responses, full length-enriched cDNA libraries were prepared from whole plant leaf materials, both uninfected and artificially challenged with pathogen conidiospores.

Results: The study generated 846,762 high quality sequence reads, with an average length of 334 bp and totalling 283 Mbp. *De novo* assembly generated 36,384 and 35,269 unigene sequences for *M. acuminata* Calcutta 4 and Cavendish Grande Naine, respectively. A total of 64.4% of the unigenes were annotated through Basic Local Alignment Search Tool (BLAST) similarity analyses against public databases.

Assembled sequences were functionally mapped to Gene Ontology (GO) terms, with unigene functions covering a diverse range of molecular functions, biological processes and cellular components. Genes from a number of defense-related pathways were observed in transcripts from each cDNA library. Over 99% of contig unigenes mapped to exon regions in the reference *M. acuminata* DH Pahang whole genome sequence. A total of 4068 genic-SSR loci were identified in Calcutta 4 and 4095 in Cavendish Grande Naine. A subset of 95 potential defense-related gene-derived simple sequence repeat (SSR) loci were validated for specific amplification and polymorphism across *M. acuminata* accessions. Fourteen loci were polymorphic, with alleles per polymorphic locus ranging from 3 to 8 and polymorphism information content ranging from 0.34 to 0.82.

Conclusions: A large set of unigenes were characterized in this study for both *M. acuminata* Calcutta 4 and Cavendish Grande Naine, increasing the number of public domain *Musa* ESTs. This transcriptome is an invaluable resource for furthering our understanding of biological processes elicited during biotic stresses in *Musa*. Gene-based markers will facilitate molecular breeding strategies, forming the basis of genetic linkage mapping and analysis of quantitative trait loci.

Keywords: *Musa acuminata*, *Mycosphaerella musicola*, Biotic stress, Transcriptome, 454 pyrosequencing, SSR

* Correspondence: robertmiller@unb.br

¹Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Departamento de Biologia Celular, CEP 70.910-900, Brasília, D.F, Brazil

Full list of author information is available at the end of the article

Background

Cultivated edible banana and plantains are derived from the progenitor species *Musa acuminata* Colla (A genome) and *Musa balbisiana* Colla (B genome), which are both members of the section *Eumusa*. These commodity fruit crops are amongst the most important across tropical and sub-tropical regions, contributing to food security, nutrition and poverty alleviation. Global annual production from over 120 countries on 5 continents is estimated to be approximately 102 million tonnes [1].

Global movement of *Musa* germplasm from its centre of origin in Southeast Asia and the Pacific region has resulted in a spread of pests and disease, causing major constraints to banana production. Biotic stresses include fungi, bacteria, viruses, nematodes and insects. In excess of 40 fungal pathogens cause disease in banana [2], with three species of the genus *Mycosphaerella* recognized as important foliar pathogens. As causal members of the Sigatoka disease complex, *Mycosphaerella fijiensis* is responsible for black leaf streak disease, *Mycosphaerella musicola* for Sigatoka leaf spot disease, and *Mycosphaerella eumusae* for Eumusae leaf spot. *M. musicola* was the first *Mycosphaerella* pathogen to be recorded on banana, spreading from Java (Indonesia) in 1902 to most of the world's production areas in the 1960s [3]. With a preference for higher altitudes and cooler temperatures, it is typically a greater problem during rainy seasons in subtropical banana growing regions [4,5]. Foliar necrotic lesions and diminished photosynthetic capacity cause reductions in fruit number and size per bunch, with premature fruit ripening observed in the field and post-harvest. Estimated production losses vary between 50 and 100%, given that infected fruits have no commercial value [6,7].

As disease resistance is absent in most cultivated varieties, in particular members of the Cavendish subgroup, control is largely based upon agronomic management practices and application of protectant and systemic fungicides. In addition to increasing production costs, long term dependence upon agrochemical control increases selection pressure for fungicide resistance or tolerance development in pathogen populations. Resistance to benzimidazole, triazole and strobilurin systemic fungicides has been reported in the genus *Mycosphaerella* [8,9].

Given the susceptibility to an ever-increasing range of pests and diseases, development of resistant cultivars through genetic improvement is of fundamental importance for sustainable disease management. In contrast to fertility in wild diploid *Musa* genotypes, commercial triploid and diploid cultivars are seedless and parthenocarpic, with fruit development via parthenocarpy. With maintenance of such plants by vegetative propagation [10], somatic mutation-driven evolution has resulted in a crop with a narrow genetic base, with many genotypes lacking resistance to pests and disease. Conventional

genetic improvement is hindered principally by male and female sterility, with approaches time-consuming and demanding in terms of land use. For example, current hybridization strategies for the development of resistant tetraploid varieties rely upon sexually active wild or improved fertile *M. acuminata* diploids, which provide sources of resistance to biotic and abiotic stresses, for crossing with established semi-fertile triploid genotypes [11,12]. Success can be limited, however, given low numbers or absence of seeds. Complementary strategies for resolving constraints for *Musa* improvement are also under development, with molecular and tissue-culture approaches including mutagenesis, somaclonal variation, somatic hybridization and genetic modification via plant transformation (for review see [13]).

Genetic modification of *Musa* requires access to genomic information, including expression analysis of gene models under different conditions. Nuclear genome size has been reported to range from 534–615 Mbp in the genus, with variation observed between species and among *M. acuminata* genotypes [14]. The publication in July 2012 of a 90% complete draft of a reference whole genome sequence for a double haploid of *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang) reported a genome size of 523 Mbp, with 36,542 predicted gene models [15]. Although banana is one of the world's most important edible crops, comprehensive diverse transcriptome datasets, complementary to a whole genome reference sequence, are required for use in accelerated molecular-based breeding in this genus. Publically available datasets currently contain 15560 ESTs for *M. acuminata* and 5320 for *M. balbisiana* (accessed July 2012), numbers which represent only a fraction of the total number of unigene sequences expected to be present in the whole transcriptome. Examples of developed datasets include those from different genotypes, plant tissues [16] and during ripening [17,18]. Examination of gene expression in relation to drought tolerance has also been reported [19-21]. Only limited analysis of gene expression in response to fungal biotic stresses has been reported. Examples for economically important pathosystems include *Musa-Fusarium oxysporum* f. sp. *cubense* [22] and *Musa-M. fijiensis* [15,23,24].

Next Generation Sequencing (NGS) of uncloned cDNA is appropriate for whole genome transcriptome characterization and gene discovery. Today's available 454 GS FLX platform with Titanium chemistry allows for read lengths of 400 bases, such that accurate *de novo* assembly of transcripts can be achieved. 454 transcriptome pyrosequencing has now been conducted in numerous important plants including *Arabidopsis thaliana* [25], *Oryza longistaminata* [26] *Medicago sativa* [27] and *Phaseolus vulgaris* [28]. 454 transcriptome analysis in plant-fungi pathosystems is also now being reported, for example in [29-33].

In order to develop a functional genomics resource for *M. acuminata*, including transcriptome response data in relation to infection by the fungal pathogen *M. musicola*, we performed Roche 454 Pyrosequencing of expressed genes in genotypes contrasting in resistance to Sigatoka leaf spot disease. Total RNA was extracted from whole plant leaf material from the wild diploid genotype *M. acuminata* ssp. *burmannicoides* Calcutta 4 (incompatible response) and the commercial triploid *M. acuminata* subgroup Cavendish cv. Grande Naine (compatible response), both uninfected and challenged with the pathogen. Transcriptome datasets were also exploited for large scale gene-based marker development.

Results and discussion

The objectives of this work were to generate a transcriptome resource for *M. acuminata* which includes genes expressed in banana-*M. musicola* interactions using highly susceptible (Cavendish Grande Naine) and completely resistant (Calcutta 4) genotypes. Calcutta 4 is a wild fertile diploid widely employed in breeding programs for improvement of commercial cultivars (e.g. [34]). As a donor species, it is considered an important source of resistance to important fungal pathogens and nematodes. Given this importance, it has been adopted as a model for comparative genomics with rice [35,36], with functional genomics applications [19] and candidate resistance gene discovery also reported [37,38]. Cavendish subgroup bananas, such as Grande Naine, by contrast, are sterile triploids, which, although representing more than 40% of global production, lack resistance to biotic stresses, such that regular pesticide application is necessary for commercial production. In addition to unigene discovery for each genotype during this pathosystem interaction, large scale isolation of microsatellites and genic-SSR marker development was also conducted, for application in genetic mapping, genotyping and marker-assisted selection of specific traits in breeding populations.

454 sequencing statistics and assembly

Emulsion PCR and 454 pyrosequencing were conducted according to Roche standard protocols using GS FLX technology and Titanium series chemistry. Each cDNA library was sequenced on a ¼ segment of a single plate run, generating 978,133 raw sequence reads for the two genotypes, totalling over 466 megabases of sequence data. Following adaptor sequence trimming and short read (< 50 bp) removal, a total of 846,762 high quality reads (283 megabases) were processed. Table 1 shows a summary of size distribution for both genotype datasets.

Calcutta 4 sequence reads were *de novo* assembled into 36,384 unique unigene sequences, which included 24,259 contigs and 12,125 singletons. Of these, a total of 25,381

Table 1 Size distribution of the high quality *M. acuminata* Calcutta 4 and Cavendish Grande Naine 454-derived sequence reads

<i>M. acuminata</i> Calcutta 4				
EST Length summary (bp)	Min	Mean	Median	Max
	40	332	353	908
EST Length distribution(bp)	40 – 100	20472		
	101 – 200	39009		
	201 – 300	78901		
	301 – 400	152325		
	401 – 500	129795		
	>500	3737		
<i>M. acuminata</i> Cavendish Grande Naine				
EST Length summary (bp)	Min	Mean	Median	Max
	40	336	362	1161
EST Length distribution(bp)	40 – 100	22665		
	101 – 200	38635		
	201 – 300	71477		
	301 – 400	140099		
	401 – 500	143815		
	>500	5832		

unigenes were represented in the transcriptome dataset from non-infected leaves and 25,154 in the dataset from pathogen-challenged leaves. In the case of Cavendish Grande Naine sequences, a similar assembly pattern was observed, with a total of 35,269 unigenes, composed of 23,729 contigs and 11,540 singletons. 18,611 unigenes were represented in the transcriptome dataset from non-infected leaves and 29,223 in the dataset from pathogen-challenged leaves. Singleton sequences which represent unique low level transcripts were generally of sufficient length to enable annotation, with an average length of 343 bp in Calcutta 4 and 345 bp in Cavendish Grande Naine. Unigene contig length distributions for each genotype are shown in Figure 1A. Average contig lengths of 552 bp and 548 bp were observed for Calcutta 4 and Cavendish Grande Naine, respectively. Distribution of the number of reads in a unigene contig (depth of a contig) were also similar in both assembled sequence datasets (Figure 1B).

Considering both contigs and singletons, the total number of genes per genotype corresponds well with the most recent estimate of 36,542 protein-coding gene models in the reference genome sequence for *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang) [15], which may reflect the stringent quality analysis and assembly parameters adopted. It must be recognized, however, that *de novo* assembly may over-estimate gene numbers as a result of

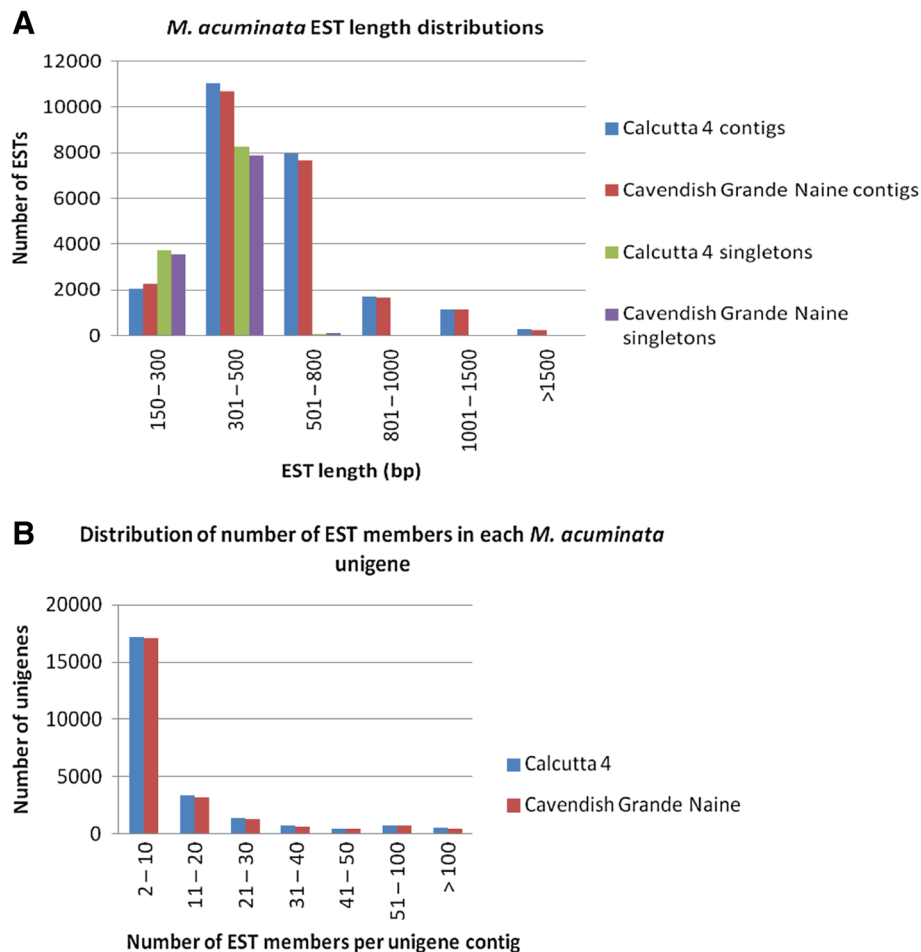


Figure 1 Summary of sequence assemblies from *M. acuminata* Calcutta 4 and Cavendish Grande Naine datasets. **A**, Length distribution of the assembled *M. acuminata* unigene contigs and singletons; **B**, Distribution of number of sequence reads in assembled *M. acuminata* unigene contigs.

non-overlapping sequence read members present for a single gene. Alignment of contigs against the reference genome revealed such examples, with 24,097 of the 24,259 contigs in Calcutta 4 (99.3%) mapping to 16,519 gene models, and similarly 23,548 of the 23,729 contigs in Cavendish Grande Naine (99.2%) mapping to 16,402 reference genome gene models (Additional files 1 and 2).

The high percentages of unigenes mapped to gene model exons validates both the *de novo* contig assemblies and gene annotation of the reference *M. acuminata* genome, overlaying important information in relation to host expression during this plant-pathogen interaction. Through TBLASTX analysis of unmapped contigs, positive hits to genes in the NCBI EST (others) database identified a further 162 potential unigenes in Calcutta 4 and 181 in Cavendish Grande Naine (Additional file 1). Although these unaligned gene sequence may be specific to the *M. acuminata* genotypes, this data may also indicate additional genes requiring further curation in the reference genome.

In order to gain insights into broad similarities and differences between transcriptome unigene datasets (contigs and singletons) for the two *M. acuminata* genotypes evaluated in this study, all gene models in the reference genome were used as a base for identification of common and distinct genes. A Venn diagram (Additional file 3) illustrates overlap between the genotypes, with 16,386 common gene models identified to which mapped coverage of the query unigene sequence was greater than 90% and the percentage identity of the sequence relative to the genome was greater than 95%. This number represents 82.0% of the mapped gene models for each genotype. Although mapped unigenes specific to each genotype may be attributed to different evolutionary distances from *M. acuminata* ssp. malaccensis var. Pahang, overlap between data sets is likely to correlate with 454 sequence coverage.

All sequence data from the study are available for each genotype in the Sequence Read Archive (SRA) at the

National Center for Biotechnology Information (NCBI) (submission SRA055816).

Functional annotation and classification

Annotation of assembled unigene sequences was conducted by sequence similarity searches against the NCBI non-redundant protein sequence database (nr). BLASTX criteria were that the alignment length should be greater than 100 amino acids and the E-value cut-off at 10^{-5} . Of the total estimated unigene sequences for Calcutta 4, 10,080 displayed significant identity to genes encoding proteins with known or putative function, 1,633 to genes encoding proteins with unknown function, and 13,513 showed no significant identity to any sequences in the database. Similar results were observed for Cavendish Grande Naine, with 10,645 unigenes displaying significant identity to genes encoding proteins with known or putative function, 1,800 to genes for proteins with unknown function, and 11,971 showing no significant identity to any database sequences. Unigene sequences and Blast annotations are summarized in Additional file 4. The protein domain-searching tool InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>) was used to further annotate sequences. A total of 14,826 Calcutta 4 unigenes contained interpro domains, with 3,949 distinct domains represented in the unigene set. Of the 13,513 translated sequences with no significant Blast hits, 192 showed functional protein domains. Similarly, 14,006 Cavendish Grande Naine unigenes possessed interpro domains, with a total of 3,974 interpro domains represented. A total of 223 out of the 11,971 sequences with no significant hits to Genbank database sequences were identified with interpro domains. The most representative domains in both unigene datasets are summarized in Additional file 5. Novelty rates for unigene datasets based upon Blast and interpro analyses were 36.6% in the case of Calcutta 4 and 33.3% for Cavendish Grande Naine.

Unigene sequences with significant homology to known plant proteins from the public database were abundant. On the basis of best blast hit analyses, respectively for Calcutta 4 and Cavendish Grande Naine unigenes, totals of 8,518 and 8,948 matched genes in rice (*Oryza sativa*), 5,718 and 6,078 matched genes in maize (*Zea mays*), 6,626 and 7,026 matched genes in sorghum (*Sorghum bicolor*), 9,060 and 9,579 matched genes in grape (*Vitis vinifera*), and 3,771 and 3,927 matched genes in *Arabidopsis thaliana* (Figure 2). Similar distributions were observed for *M. acuminata* DH-Pahang gene families [15]. For both Calcutta 4 and Cavendish Grande Naine datasets, only 2.8% of Blast hits to homologous plant sequences matched those from the GenBank NR protein database with a taxonomic filter for the genus *Musa*, indicating that our data represents a considerable contribution to expressed unigenes for the genus.

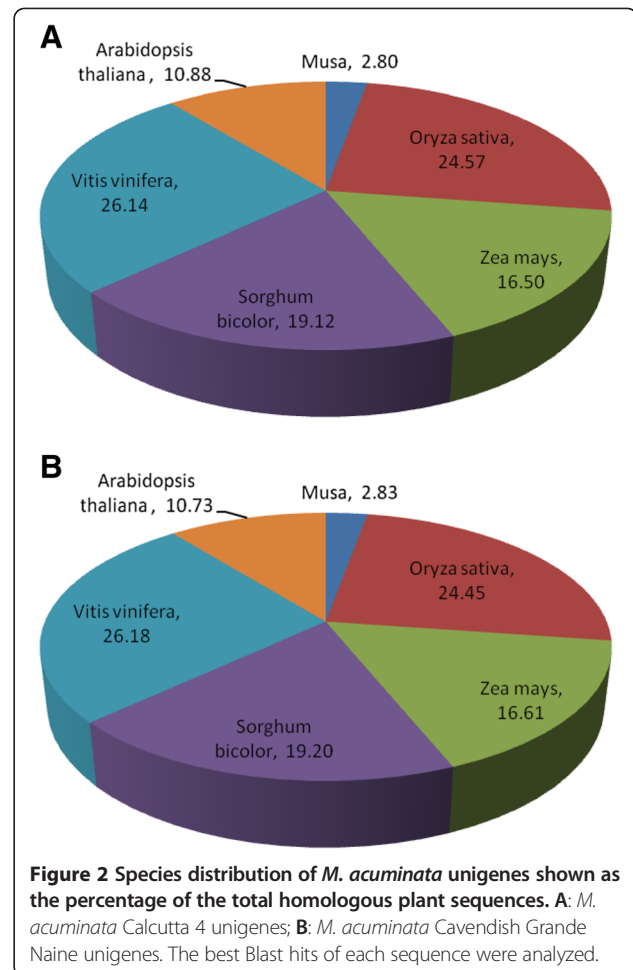


Figure 2 Species distribution of *M. acuminata* unigenes shown as the percentage of the total homologous plant sequences. **A:** *M. acuminata* Calcutta 4 unigenes; **B:** *M. acuminata* Cavendish Grande Naine unigenes. The best Blast hits of each sequence were analyzed.

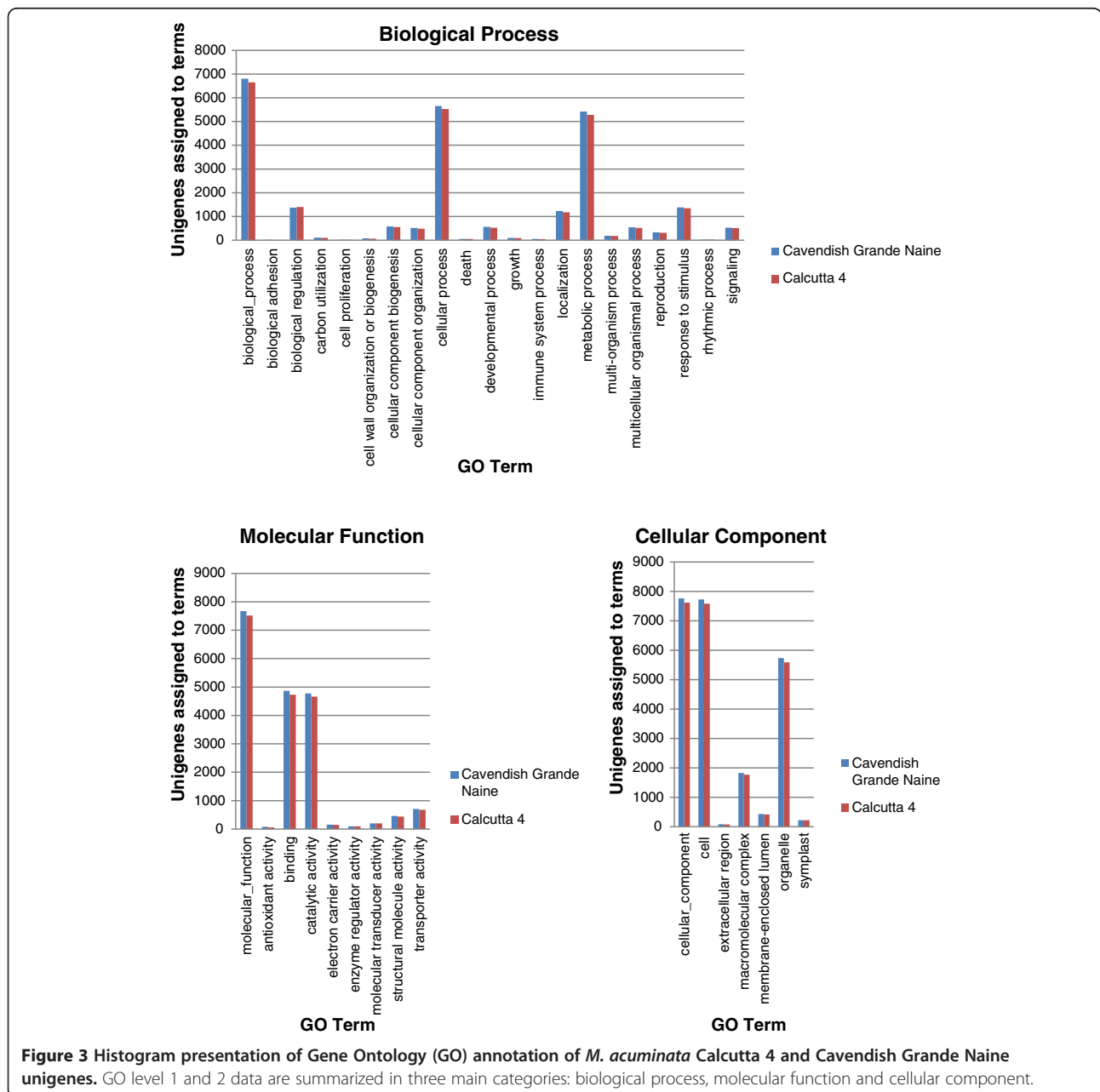
Functional classification of unigene sets was conducted through Gene Ontology (GO) assignment using Blast2GO. Data was categorized for each genotype across the three main GO categories of biological process, cellular component and molecular function. Calcutta 4 unigenes were assigned a total of 341,244 GO term annotations, with 162,468 biological process terms representing 10 GO levels, 69,390 molecular function terms from 10 levels and 109,386 cellular component terms from 9 levels. Similar assignments were seen with Cavendish Grande Naine unigene data, with 351,220 GO terms assigned across the three main categories. A total of 168,015 biological process terms were assigned, 70,878 molecular function terms and 112,327 cellular component terms. Within the biological process category, At GO level two, the majority of unigenes were assigned to “cellular process” (5,530 for C4, 5,649 for CAV), metabolic process (5,276 for C4, 5,418 for CAV), “biological regulation” (1,397 for C4, 1,377 for CAV), “response to stimulus” (1,343 for C4, 1,383 for CAV), and “localization” (1,179 for C4, 1,227 for CAV). Similarly, for Molecular Function, terms “binding” (4,732 for C4, 4,863

for CAV), “catalytic activity” (4,659 for C4, 4,771 for CAV) and “transporter activity” (678 for C4, 708 for CAV) were the most abundant assigned terms for the unigene datasets from both genotypes. Across the cellular function category, the most abundant terms were “cellular component” (7,620 for C4, 7,759 for CAV), “cell” (7,580 for C4, 7,725 for CAV), “organelle” (5,589 for C4, 5,735 for CAV) and “macromolecular complex” (1,772 for C4, 1,829 for CAV). Figure 3 summarises level 1 and 2 GO annotation of unigenes.

Musa-Mycosphaerella interactions

M. musicola is a hemibiotrophic pathogen which penetrates leaf tissues through stomatal pores, following a

period of epiphytic growth on the leaf surface. Once inside the host, the pathogen colonizes the intercellular space within mesophyll tissue layers and palisade tissues, without forming haustoria or infecting host cells. This biotrophic phase can last for a number of weeks before the onset of symptoms of necrotic lesions in mesophyll cells. Growth of conidiophores out through stomata then enables conidiospore development on the leaf surface. Notable differences in disease development are seen in totally resistant genotypes, with early necrosis of stomatal guard cells and death of only a limited number of host cells at the site of infection. No fungal sporulation is observed in such incompatible interactions.



Analysis of gene expression during banana-*Mycosphaerella* interactions has been limited to date [15,23,24], with our study, to the authors knowledge, the first massal transcriptome analysis for the *M. acuminata*-*M. musicola* pathosystem. A sequencing strategy for identification of transcripts from pathogen-challenged and unchallenged leaves was employed, with the preparation of two cDNA libraries per genotype (infected and non-infected). Given reports that during the early biotrophic phase of infection of banana with *M. fijiensis*, germ tubes penetrate stomata from 3 to 6DAI [39], a similar timecourse was employed for the *M. acuminata*-*M. musicola* pathosystem.

The use of detached leaf tissues in disease bioassays with *Mycosphaerella* banana pathogens has been reported on occasion to give inconsistent results (e.g. [40,41]), with development of disease symptoms not always correlating with those typically observed on intact plants. Additionally, [40] suggested that the hemibiotrophic *Mycosphaerella* banana pathogens require healthy banana plants for disease development, with [42] suggesting that the physiological state of detached leaves is not comparable with those on whole plants with an intact root system. Host gene expression in detached leaves during interaction with biotrophs has also been reported in *Arabidopsis* to closer reflect plant senescence [43]. For these reasons, bioassays were conducted using intact young leaves from 6 month old plants, with optimal temperature and humidity conditions employed during the experiment.

On the basis of BlastX search results against nr, Blast2GO annotation and assignment of unigenes to GO terms related to defense (Additional files 4 and 6) many unigenes potentially involved in plant Effector-Triggered Immunity (ETI) and PAMP-triggered Immunity (PTI) were identified across both genotypes. PTI is a branch of plant immunity which involves interactions between host pattern recognition receptor-like kinases (PRRs) and pathogen-associated molecular patterns (PAMPs) [44]. This branch involves a mitogen-associated protein (MAP) kinase cascade and WRKY transcription factors, and is responsible for resistance to the majority of potential pathogens. ETI [45], is based upon the co-evolution of plant resistance R protein receptors and specific pathogen effector molecules, responsible for disease resistance at the intra-specific level. Downstream signal transduction components can overlap between PTI and ETI, including an oxidative burst via the production of ROS, alterations in plant hormone production and MAPK signaling cascades.

Expressed genes in Calcutta 4 and Cavendish Grande naine

In silico analysis of gene expression together with Blastx-derived annotation nomenclature revealed genes potentially involved in biotic stress responses in Calcutta 4. Numerous transcription factors were identified, which are

typically involved in regulation of plant development, signaling and response to environment, amongst other roles. Many are also known to be involved in signal transduction and expression regulation of stress-responsive genes. Transcripts observed in infected leaf tissues in Calcutta 4 included a NAC domain protein and ethylene insensitive-like protein 4, a probable transcription factor acting as a positive regulator in the ethylene response pathway. Plant R-gene-mediated recognition of a biotrophic pathogen *Avr* gene product is associated with the hypersensitive response (HR), which results in programmed cell death at the site of infection to limit pathogen spread [46]. In inoculated Calcutta 4 plants, unigenes potentially involved in plant detoxification included a considerable presence of metallothionein-like proteins. These low molecular weight polypeptides sequester metal ions and are associated with regulation of intracellular redox potential and oxygen detoxification [47,48], protecting cells from damaging effects of reactive oxygen species (ROS). ROS generation may indicate HR activity, following pathogen infection and recognition. Four distinct types (MT 1 to MT4) are known in plants, according to distribution of cysteine residues. [49] reported isolation of types MT2 and MT3 in banana, with expression influenced in response to ethylene and metals. More recent examination has reported their abundance in *M. acuminata* Calcutta 4 [19]. Our study confirmed this, with MT2 and MT3 unigene contigs with considerable sequence depth. Superoxide dismutase enzymes (SODs) were similarly observed in infected Calcutta 4 leaves. Like metallothioneins, these also act as antioxidants, protecting plant cell components from oxidation by ROS. Glutathione-S-transferases were also represented. These may be involved in cell signaling pathways as well as in detoxification of products of oxidative stress during HR. Their expression in *M. fijiensis*-*M. acuminata* late stage compatible interactions has also recently been reported [23]. Reported peaks in accumulation of H₂O₂ and peroxidase activity in Calcutta 4 up to 10DAI with *M. fijiensis* [50] are consistent with our observations of abundance of transcripts for genes involved in ROS detoxification and HR during interaction with *M. musicola*. Phenylpropanoids in plants are involved in a number of defense responses, including biosynthesis of antimicrobial compounds such as phytoalexins and molecules involved in signaling. An abundance of gene transcripts for phenylalanine ammonia-lyase (PAL) was observed in Calcutta 4 following challenge with the pathogen. This enzyme catalyses a reaction in the phenylpropanoid biosynthetic pathway, with deamination of phenylalanine to cinnamic acid, leading to downstream synthesis of phytoalexins, as well as production of salicylic acid, a signal molecule involved in systemic acquired resistance (SAR). 4-coumarate-CoA ligase, another important enzyme in the biosynthesis of flavonoids and isoflavonoids, was also observed.

Blast nr-based functional annotation of expressed unigenes in Cavendish Grande Naine revealed an abundance with homology to transcripts poorly characterized according to the NCBI. However, where descriptions could be used as a guide, numerous unigenes in infected leaves were identified as potentially involved in plant responses to biotic stress. These included transcription factors, metallothionein-like proteins and superoxide dismutases (plant detoxification), 4-coumarate:coA ligase 2, cinnamic acid 4-hydroxylase and isoflavone reductase-like protein (phenylpropanoid pathway), and disease-related F-box protein and calmodulin binding protein (defense response). Examples of unigenes with fewer counts in infected leaves when compared with non infected tissues included WRKY transcription factor 17 and MAP kinase BIMK1 (defense signaling), putative callose synthase 1 catalytic subunit (plant callose synthesis), xyloglucan endotransglycosylase (xyloglucan-cellulose framework modification and strength of plant cell walls), endochitinases and putative chitinases (degradation of fungal cell walls), pathogenesis-related protein 1, F-box, wd40 domain protein, hypersensitive-induced response protein (plant defense), glutathione S-transferase 1 (plant detoxification), type III polyketide synthase 4, cinnamate-4-hydroxylase (phenylpropanoid pathway), CTR1-like protein kinase, ethylene receptor-like protein, ethylene response factor 11 and ethylene-responsive transcriptional coactivator (ethylene and defence signaling), respiratory burst oxidases (ROS signaling, signal transduction and cell death) and thaumatin-like proteins (degradation of fungal cell walls).

A number of distinct plant disease resistance *R*-gene families are recognized as involved in ETI and PTI, based upon protein domains and cell function. The most abundant class code for cytoplasmic receptor proteins with nucleotide binding site-leucine-rich repeat (NBS-LRR) domains [51]. In rice, approximately 400 NBS-LRR genes have been characterized, with 150 present in the *Arabidopsis* genome [52], and 89 identified in *M. acuminata* DH-Pahang [15]. Conservation of motifs within nucleotide-binding site leucine rich repeat domains has also enabled analyses of NBS-LRR *R*-gene family diversity across the genus (e.g. [38,53]). In the current study, 14 expressed NBS-LRR genes were identified through Blast analysis from both infected and non-infected leaf tissues for Calcutta 4 and 25 in Cavendish Grande Naine. Mapping of unigene contigs to gene models containing the NB-ARC domain in the reference *M. acuminata* genome identified 38 contigs mapping to 40 gene models in the case of Calcutta 4 transcriptome data and 43 Cavendish Grande Naine contigs mapping to 40 gene models (Additional file 1). Other known plant *R*-gene classes include extracellular LRRs anchored by transmembrane domains (receptor-like proteins), extracellular LRRs linked to cytoplasmic serine-threonine kinase domains (receptor-like kinases), intra-

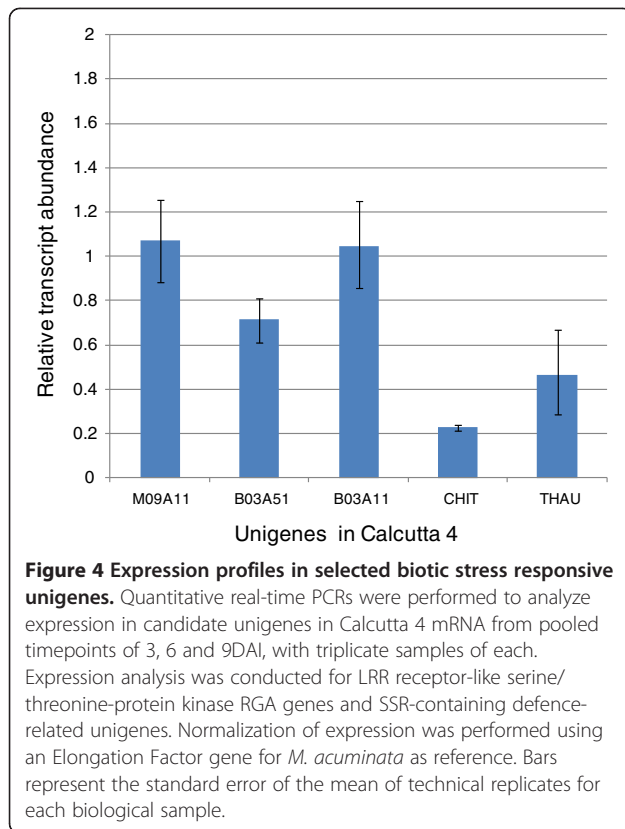
cellular serine-threonine kinases, and proteins with a coiled-coil domain anchored to the cell membrane. Blast analysis predicted numerous transcripts for these classes among the two genotype datasets.

Assignment to GO terms related to defense and Blast2go annotation (Additional file 3) provide a further level for mining candidate genes involved in host defense responses, complementing Blast nr-derived annotation. For example, a number of interesting unigenes in pathogen-inoculated Calcutta 4 leaf tissues were identified. These include glucan synthase components, which are associated with callose deposition in host cell walls during defense response, Rpm1 interacting protein 4, reported in plant defense involving R proteins RPM1 and RPS2, Mac perforin domain-containing proteins, which are associated with the salicylic acid (SA)-mediated pathway of programmed cell death, brassinosteroid insensitive 1-associated receptor kinase 1, which is involved in PTI and programmed cell death, and mlo-like protein 1, known to be involved in plant defense and cell wall strengthening.

Quantitative real-time PCR analysis of gene expression

Receptor-like protein kinases (RLKs) are transmembrane proteins with an N-terminal signal sequence, a specific receptor extracellular domain, and an intracellular C-terminal kinase domain. With over 600 RLKs characterized in *Arabidopsis* [54], only few are known to be involved in plant immunity [55,56]. Expression of LRR receptor-like serine/threonine-protein kinase RGA genes was evaluated in Calcutta 4, following alignment of 454 transcriptome data to annotated BAC sequences and specific primer design for mapped expressed genes. Quantitative real-time PCR revealed no significant change in expression in the three selected RGAs in relation to basal expression in non-inoculated controls (Figure 4). Although constant transcript levels during pathogen attack have been reported for different classes of host receptor R genes, where constitutive expression may mediate pathogen recognition and activation of signal transduction and defence responses, upregulation has been demonstrated in certain pathosystems in response to infection [57]. Expression analysis of potential host defence-related genes modulated during interaction with *M. musicola* was also conducted on randomly selected SSR-containing unigenes in Calcutta 4 (KOG descriptions: Pathogenesis related protein group 5, putative chitinase). Both genes were downregulated following pathogen challenge over the timecourse (Figure 4), with abundance of raw EST counts for each of the examined genes in inoculated and non-inoculated controls reflecting the changes observed with qRT-PCR data.

The overlaying of the 454 transcripts identified in the *M. acuminata* – *M. musicola* interaction onto gene models on the *M. acuminata* DH Pahang reference



genome will facilitate ongoing validation of candidate gene expression via qRT-PCR. An in-depth understanding of the mechanisms by which R genes and downstream defence mechanisms function in *Musa* is necessary for development of novel strategies for durable resistance.

Data mining against *M. fijiensis* gene models

Data mining for *M. musicola* pathogen transcripts amongst the pre-processed 454 sequence data derived from infected leaves was performed through genome mapping and alignment against the *Mycosphaerella fijiensis* v2.0 All Gene Models (transcripts) database. Based upon rDNA ITS sequence analysis, this species forms a monophyletic group with *M. musicola* [58]. Although pathogen transcript abundance is likely dependent on 454 depth of sequencing, a total of 10 unigenes from transcripts in pathogen-infected Calcutta 4 and 12 in Cavendish Grande Naine mapped to *M. fijiensis* gene models (Additional file 7). In addition to hypothetical proteins and no hits which were identified in both C4 and CAV, potential pathogen genes identified in transcriptome data for CAVI included a SAP family cell cycle dependent phosphatase-associated protein, two Hsp70 family proteins, a FAD binding domain protein, and a calcium channel. In the case of C4I data, positively mapped *M. fijiensis* gene models also included an extracellular protein 6 (Ecp6), and two 60S ribosomal proteins.

A total of seven Ecps have been identified in *Cladosporium fulvum* (syn. *Passalora fulva*), which are secreted during interaction with tomato [59]. The *C. fulvum* effector protein Ecp6, crucial for virulence, contains carbohydrate-binding Lysin (LysM) domains that may be involved in binding to chitin released from fungal cell walls during the infection process [60]. Such PAMP-binding may prevent induction of plant basal defense responses. Orthologs of Ecp6 have been identified in EST and whole genome sequence data in a number of fungal genera. Within the genus *Mycosphaerella*, LysM-containing Ecp6-like proteins have been identified in *M. fijiensis* [59] and *Mycosphaerella graminicola* [59,61], indicating their likely presence in other members of the genus. Whilst homologues of the *C. fulvum* Ecp2 effectors have been described in *M. fijiensis* transcripts from *in vitro* culture [62], to date there have been no reports of *in vivo* Ecp effector homolog expression for the *Mycosphaerella* banana pathogens during host-pathogen interaction. Our identification of Ecp6 effector protein homolog transcripts in *M. musicola* may contribute towards identification of pathogen effectors and cognate disease resistance genes in *Musa*.

Genic SSR markers

Microsatellite markers or simple sequence repeats (SSRs) are defined as tandem DNA repeats which are dispersed in both coding and non-coding regions along eukaryotic genomes. As molecular markers, they are typically somatically stable, polymorphic, co-dominant, and multi-allelic. Applications in *Musa* have focused on taxonomy (e.g. [63]), genotyping (e.g. [64]), and genetic map saturation (e.g. [65]). Gene-derived SSRs can be associated with functional genetic variation, as opposed to non-coding SSRs, with presence in transcribed regions potentially influencing gene function, transcription or translation [66]. Consequently they offer potential for marker assisted breeding, with markers either originating from a gene for a desirable phenotypic trait, or co-localizing with a particular quantitative trait locus. With such markers isolated from coding regions, conservation is also potentially greater, increasing transferability to related species (e.g. [67]). In comparison with other crops, relatively few SSR markers have been developed for the genus *Musa*, reflecting the limited sequence resources available until recently. Whilst the majority have been isolated from genomic libraries or BAC clones [35,63,68-71], only few gene-derived SSRs have been characterized (e.g. [72,73]).

We screened the Calcutta 4 and Cavendish Grande Naine unigene datasets for the presence of SSR motifs. A total of 4,068 potentially PCR amplifiable genic-SSR loci were identified in the 36,384 Calcutta 4 consensus sequences, representing an EST-SSR frequency of 11.18%. Similar numbers were observed for the 35,269 Cavendish Grande Naine consensi, with a total of 4,095 SSRs,

representing a frequency of 11.61%. In terms of SSR class distribution, for Calcutta 4 unigenes, trinucleotide repeats were the most abundant (61.4%), followed by di- (18.26%), tetra- (9.7%), hexa- (5.9%) and penta-nucleotide repeats (3.7%). Hepta-nucleotide repeats and above accounted for 1.04%. A similar distribution was observed in Cavendish Grande Naine unigenes, with trinucleotide core motifs representing 61.68% of repeats, followed by dinucleotides (19.4%), tetra- (8.6%), hexa- (5.8%) and pentanucleotides (3.5%). Again, heptanucleotide repeats and longer SSRs made up only 1.02% of the total. A predominance of trinucleotides in transcriptome sequence data is common [74,75], with the presence of such motifs in gene regions avoiding frameshift mutation introduction and changes at the protein level. As expected, the shorter the nucleotide core sequence, the greater the number of repeats present. For Calcutta 4 an average of 7.74 repeats were observed for di-nucleotide motifs, 4.99 for tri-, 3.81 for tetra-, 3.56 for penta-, and 3.63 for hexa-motifs. Similarly for Cavendish Grande Naine there were an average of 8.55 repeats for di-, 4.99 for tri-, 3.75 for tetra-, 3.59 for penta-, and 3.70 for hexa- motifs. Motif frequency distribution for trinucleotide and dinucleotide core motifs, which were the most abundant repeats for both genotypes are shown in Figure 5.

The details of all 8,163 *M. acuminata* genic-SSR primers, including SSR motif, primer sequences, PCR amplification information, predicted gene function and *in silico* expression data are provided in Additional file 8. A subset of 95 potential defense-related gene-derived SSR loci, selected on the basis of Blast similarities and KOG descriptors, were validated for specific amplification and polymorphism across 20 diploid *M. acuminata* accessions, to complement previous work by our group [70]. A total of 73 (76.8%) amplified specific products of expected size from genomic DNA originating from the tested *M. acuminata* genotypes. Polymorphism was observed in 14 loci (14.7%), with alleles per polymorphic locus ranging from 3 to 8 and a total of 66 alleles scored across the polymorphic loci. Polymorphism information content ranged from 0.34 to 0.82, with an average of 0.63 (Table 2). This limited diversity can be expected for genic-SSR markers, as a result of DNA sequence conservation of transcribed regions [76,77].

Our work represents a large scale development of new genic-SSR markers for application in genetic improvement, complementing the approximately 2,000 SSRs identified in BAC-end and sequence scaffolds of DH-Pahang [15]. With efforts underway towards development of segregant populations for traits of interest [78-80], these functional gene-based markers are applicable for association to trait loci and downstream marker-assisted selection, as well as evolution analysis, parentage assessment and general genotyping applications in breeding programs.

Conclusions

This NGS-based investigation of the transcriptome in the *M. acuminata* – *M. musicola* interaction provides useful data on expressed genes during plant immune responses in this pathosystem, with 36,384 and 35,269 unigene sets identified, respectively, for contrasting *M. acuminata* genotypes Calcutta 4 and Cavendish Grande Naine. Genes were characterized according to Blast annotation, GO category assignment and Interpro-based domain identification. The data represent a global transcriptome level resource for *M. acuminata*, with identification of candidate genes expressed during infection contributing to our understanding of host defense mechanisms against this important pathogen. The recently published reference whole genome for DH-Pahang represents approximately 90% of the estimated genome size. Genome annotation updates will be facilitated by the availability and mapping of comprehensive sets of expressed gene sequences for this species. Our large scale characterization of genic-SSRs and marker development are a resource for application in genetic map enrichment, diversity characterization and downstream marker assisted breeding in *Musa*.

Methods

Plant material preparation

M. acuminata Calcutta 4 and Cavendish Grande Naine whole plants (*Musa* International Transit Centre accessions ITC0249 and ITC0654) were selected for transcriptome analysis based upon their known contrasting resistance to *M. musicola*. A total of 36 six month old plants were maintained in a greenhouse under a 12-h light/12-h dark photoperiod at 25°C and 85% relative humidity. Both infected and non-infected control plants were maintained under identical conditions. A strain of *M. musicola*, isolated from Cavendish Grande Naine leaf lesion sporodochia at Embrapa Cassava and Tropical Fruits, Brazil, was used for artificial inoculation of the abaxial surface of the youngest emerged leaf of each plant. Inoculation was conducted by spraying the entire leaf with a suspension of 2×10^4 conidiospores ml^{-1} , with addition of surfactant Tween 20 at 0.05%. Control samples comprised water-surfactant-treated leaves incubated under the same growth conditions. Sprayed leaves were covered with transparent plastic bags to ensure high humidity. Three independent replicates were collected for each sample.

Scanning electron microscopy was conducted to confirm *M. musicola* germination and infection, and used to determine time points for leaf harvesting. A total of three replicates per sample were prepared for analysis. Tissues were fixed for 2.5 h in a fresh solution of 0.05 M cacodylate buffer at pH 6.8, containing 2.5% glutaraldehyde fixer, washed in 0.1 M cacodylate buffer at pH 6.8, and postfixed for 1 h through addition of 2% osmium tetroxide to the buffer solution. Sample dehydration was conducted at 4°C

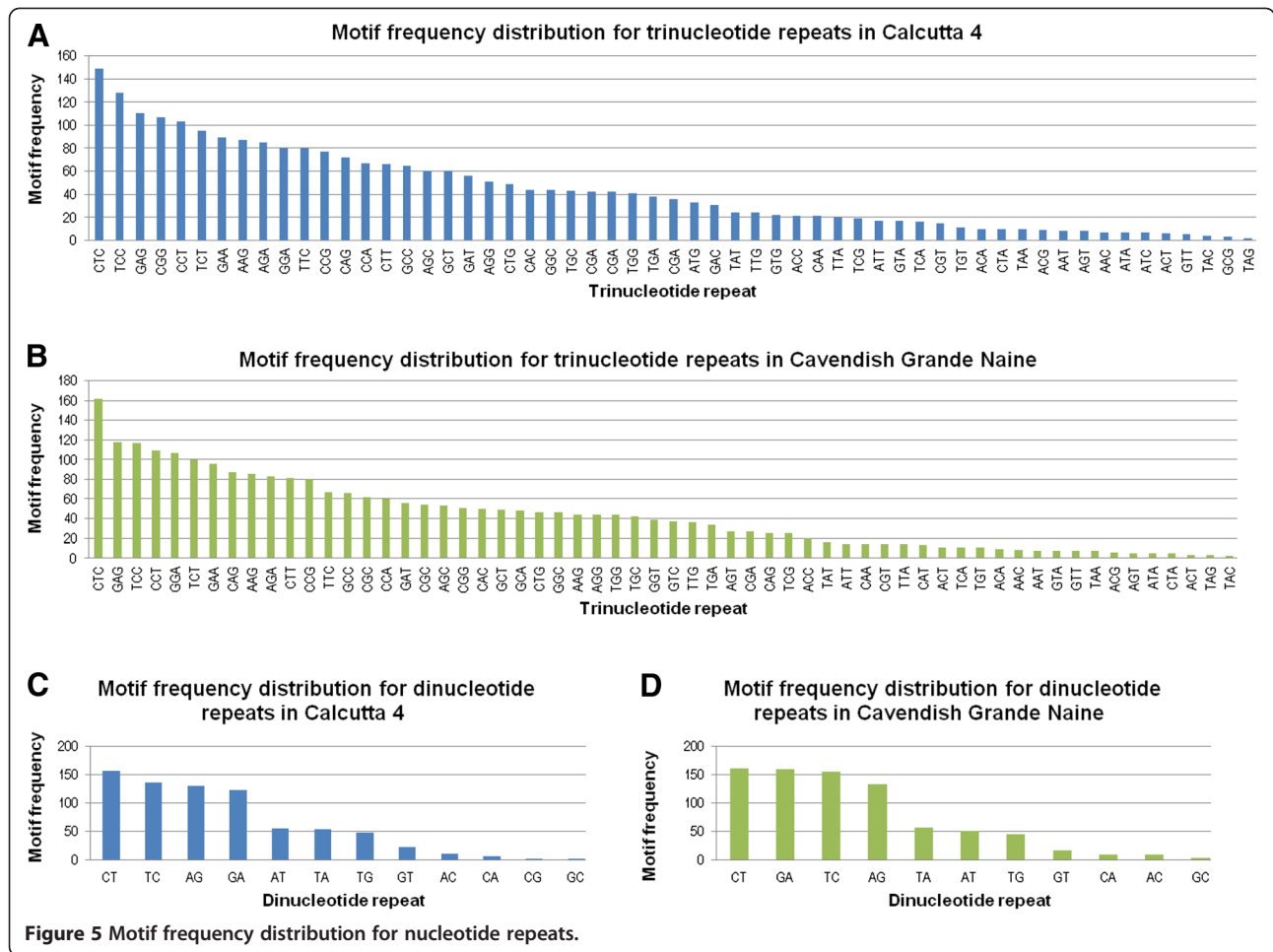


Table 2 Characteristics of polymorphic microsatellite loci isolated from *M. acuminata* Calcutta 4 and Cavendish Grande Naine unigene sets

Locus number	Unigene ID	SSR repeat motif	SSR locus length (bp)	Allelic Amplitude (bp)	Number of alleles	He	Ho	PIC Value
locus821	musa_c4_small_c22955	CGG	14	170-180	3	0.61	0.52	0.53
locus945	musa_c4_small_c25361	TGACA	17	325-330	3	0.62	0.57	0.55
locus1284	musa_c4_small_c7084	AG	11	270-330	6	0.77	0.65	0.74
locus1412	musa_c4_small_c8798	TGC	15	175-190	3	0.39	0.14	0.34
locus1654	musa_c4_small_rep_c12642	CCT	16	170-200	5	0.70	0.71	0.66
locus2108	musa_c4_small_rep_c2099	GA	14	190-220	7	0.81	0.92	0.79
locus2061	musa_c4_small_rep_c1992	AG	19	190-200	5	0.70	0.52	0.66
locus2108	musa_c4_small_rep_c2099	GA	14	205-230	8	0.84	0.91	0.82
locus2112	musa_c4_small_rep_c21184	GTC	18	200-205	3	0.52	0.65	0.46
locus2306	musa_c4_small_rep_c2740	TG	13	170-185	4	0.74	0.38	0.69
locus2984	musa_c4_clusters_pos_c1484_1	TCCT	13	115-130	7	0.80	0.45	0.77
locus371	musa_cav_c18572	AT	12	350-360	3	0.56	0.54	0.48
locus2028	musa_cav_rep_c22610	CTT	17	280-290	4	0.70	0.75	0.65
locus2646	musa_cav_rep_c5186	TTC	18	200-220	5	0.73	0.30	0.68

Polymorphism was evaluated across 21 *M. acuminata* accessions. HE, expected heterozygosity under Hardy-Weinberg expectations; HO, observed heterozygosity; PIC, Polymorphism Information Content.

for 20 minute periods with increasing concentrations of ethanol (10, 20, 30, 50, 70, 80, 90, 95 and 100%). Samples were then dried in a critical point drier (Emitech K850, Kent, UK), mounted on copper stubs and sputter coated with 20 nm gold particles. All samples were observed using a Zeiss DSM 962 Scanning Electron Microscope. Results showed that germ tubes and hyphae were visible at 3 days after inoculation (DAI), with hyphal growth over stomatal cells and penetration in Cavendish Grande Naine occurring at 6DAI onwards (Figure 6). Similar finding were reported by [49], who, investigating the *M. fijiensis*-*M. acuminata* interaction, observed increasing stomatal penetration in Cavendish Grande Naine from 5DAI, up to a total of 11% of stomata penetrated 21DAI, and only a constant 0.95% of stomata infected for Calcutta 4, over the 21 day period. Based on these data, 1 g leaf samples from each genotype were collected at 3, 6 and 9DAI, immediately snap frozen in liquid nitrogen and stored at -80°C until further processing.

RNA preparation

Total RNA from 1 g leaf samples was extracted using Concert[®] RNA Plant Reagent (Invitrogen, Carlsbad, CA, USA) and INVISORB Spin Plant RNA Mini Kit (Invitex, Hayward, CA, USA), according to manufacturer's instructions. Total RNA was treated with DNase-free (Ambion, Austin, TX, USA) using 1.5 units/ μg

of total RNA. Quantification and integrity was assessed by using ethidium bromide-stained 1% agarose gels and Nanodrop ND-1000 spectrophotometry (Thermo Scientific, Waltham, MA, USA), with a cut-off value of 1.8 for the A260 : 280 ratio. A total of 36 RNA samples were extracted from triplicate plant samples for infected leaves and non-inoculated control leaves at 3, 6 and 9 DAI from *M. acuminata* Calcutta 4 (C4I, C4NI) and *M. acuminata* Cavendish Grande Naine (CAVI, CAVNI). Four separate RNA pools were prepared for cDNA library construction, each comprising nine RNA samples (pathogen-challenged or unchallenged *M. acuminata* genotype, pooled timepoints 3, 6 and 9DAI, and triplicate samples).

cDNA library construction and 454-sequencing

Messenger RNA isolation, full-length enriched cDNA library preparation and pyrosequencing was carried out by Eurofins MWG Operon (Ebersberg, Germany) from approximately 50 μg of each total RNA pool. Non-cloned libraries were prepared by random priming of polyA + mRNA and size fractioning. Emulsion PCR and sequencing were conducted according to Roche standard protocols using GS FLX technology and Titanium series chemistry. Each of the four libraries was sequenced in a $\frac{1}{4}$ segment of a picotiter plate.

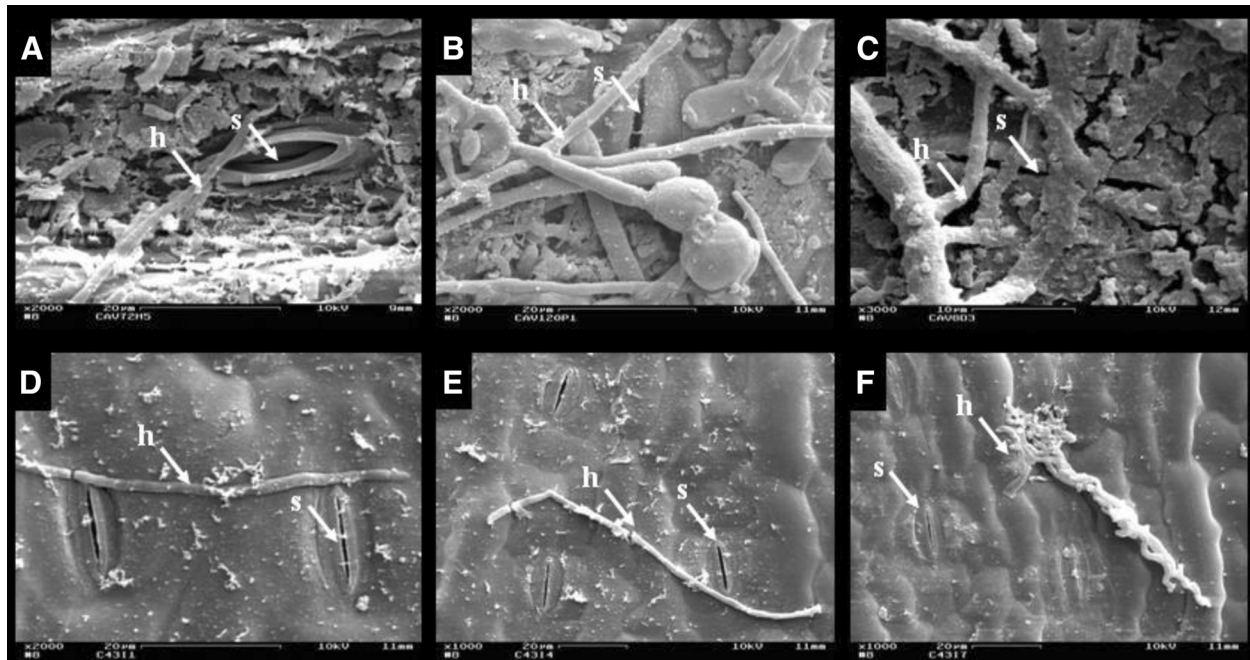


Figure 6 Scanning electronic microscopy (SEM) observation of interaction of *Mycosphaerella musicola* with abaxial leaf surfaces of *Musa acuminata* genotypes. Fungal germ tubes and hyphal growth were observed 3, 6 and 9DAI respectively, for Cavendish Grande Naine (A, B, C) and Calcutta 4 (D, E, F). Hyphal growth and penetration of stomatal openings was observed only in Cavendish Grande Naine 6DAI onwards. Abbreviations: DAI, days after inoculation; h, hypha; s, stoma.

Sequence processing and annotation

Pre-processing

All sequence processing steps were performed iteratively using the software Est2assembly platform [81]. Raw 454 sequence files in SFF Format were converted to FASTA/QUAL-compatible sequences using the SFF Extract software (http://bioinf.comav.upv.es/sff_extract/). For sequence cleaning, adapter-ligated sequences were masked using SSAHA2 [82], and undesirable sequences (such as mitochondrial DNA, rDNA or contaminant-derived) identified for masking using BLAST2 [83]. Repetitive sequences were identified and masked using RepeatMasker version open-3.3.0 (<http://repeatmasker.org>).

Sequence assembly

Frequencies of DNA K-mers in sequence data was obtained using the software Nesoni version 0.89 (Victorian Bioinformatics Consortium), with sequence groups with distinctive K-mer frequencies clustered accordingly using WcdEST version 0.6.3 [84]. Sequence assembly of cluster group members into contigs and singletons was performed using the assembly program MIRA version 3.1 [85], with 454 technology-specific and expected k-mer coverage assembly parameters. Redundancy in assembled sequences originating from -K-mer groups was inspected using BLAST2 (option megablast).

In order to assess accuracy in assembly of the *de novo*-assembled unigene contig sequences for each genotype, alignment of reads plus their associated contig information was conducted against all annotated gene models in the *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang) whole genome sequence (downloaded at <http://banana-genome.cirad.fr>) using the genomic mapping and alignment tool GMAP [86]. Alignment criteria comprised

a minimum of 90% of the query sequence mapped on the genome and the percentage identity of the sequence relative to the mapped segment greater than 95%. Validation of unigene contigs that did not map to gene models on the reference *M. acuminata* genome was conducted through TBLASTX analysis against the NCBI EST (others) database (<http://www.ncbi.nlm.nih.gov>).

Sequence annotation

Functional annotation of known proteins/genes for *de novo*-assembled unique sequences from each K-mer group was conducted against the NCBI non-redundant protein (nr) database (<http://www.ncbi.nlm.nih.gov>) using the BLASTX algorithm, with a typical E-value cut-off of 10^{-5} . Annotation of unigenes with no significant similarity to database sequences were further analyzed for protein domains with specific cell function using the software InterProScan version 4.8 [87]. Functional category assignment for each unique sequence was conducted using the Blast2GO program, classifying according to GO terms within molecular functions, biological processes and cellular components [88,89]. Comparative genomics analysis between Calcutta 4 and Cavendish Grande Naine unigene datasets was conducted through examination of common and distinct gene models in the *Musa* reference genome to which *de novo*-assembled unigenes were mapped.

For identification of potential fungal pathogen transcripts amongst the sequence data derived from infected leaves, pre-processed 454 reads were aligned against the closely related *Mycosphaerella fijiensis* v2.0 All Gene Models (transcripts) database using BWA [90] with default parameters. Counts according to reads mapped to *M. fijiensis* gene models were calculated using GenomicRanges (<http://www.bioconductor.org>).

Table 3 Primer sequences for qRT-PCR analysis of expression of SSR marker-derived genes associated with defense responses and BAC clone-derived LRR receptor-like serine/threonine-protein kinases

Gene / contig abbreviation	KOG Gene description / BAC annotation	Primer Sequence Forward/Reverse	Amplicon Size (bp)	PCR efficiency (%)	Efficiency SD (+/-)
THAU_musa_c4_small_rep_c3564	Pathogenesis related proteins, group 5, Thaumatin	CCGGTGGGACTAATTACAGG/ CAATTCGGATGTCAATGCAG	165	99	0,011
CHIT_musa_c4_small_c3092	Chitinase	CACCATCTCCTGCAAGCATA/ GCAGTCATTCCTCGTTGTCA	123	96	0,008
M09A11_454_Ma4140M09A11	Putative LRR receptor-like serine/threonine-protein kinase	CCAGCAACCACAACCCTAGT/ GATCTTTCAGGCCTCGTTTG	176	99	0,009
B03A51_454_Mac054B03A51	Putative LRR receptor-like serine/threonine-protein kinase	ATTCATGCGCTACGGACAAT/ TTCAGGCCTCGTTTGAATC	192	107	0,005
B03A11_454_Mac054B03A11	Putative LRR receptor-like serine/threonine-protein kinase	CAGCAACCATAACCCCATTC/ AGGATCTTTCAGGCCTCGTT	168	105	0,011
EF	Elongation Factor	AACCCCAAAATATTCCAAGG/ AGATTGGCACGAAAGGAATC	107	102	0,013

Differential gene expression in *M. acuminata* Calcutta 4 inoculated with *M. musicola* was analyzed in relation to non-inoculated controls.

Simple sequence repeat (SSR) identification and marker development

Perfect SSRs in unigene sequences were identified through a computational search with the program Mreps version 2.5 (<http://bioinfo.lifl.fr/mreps/>). Detection required the presence of at least two repeating units (e.g. GC) spanning more than 10 bp. Primer pairs flanking each SSR locus were designed using the program Primer3 [91]. Genic-SSRs potentially related to biotic stress responses were selected based upon KOG classification and description. From a subset of 95 selected primer pairs, marker amplification and allele length polymorphisms were evaluated using 20 diploid (AA) *M. acuminata* accessions contrasting in resistance to Sigatoka diseases, as described previously [60]. A classical 10 bp molecular size marker was added to each gel to enable allele size estimation for PCR products run on denaturing 6% polyacrylamide gels using 7 M urea. Gels were silver stained according to standard protocols. Locus polymorphism was calculated using the Polymorphism information content (PIC) calculator (<http://w3.georgikon.hu/pic/english/default.aspx>).

Quantitative real-time PCR analysis

Expression analysis of a number of selected candidate unigenes was conducted using quantitative real-time PCR (qRT-PCR). Two independent bioassays were performed, as described above, to enable comparison of gene expression in *M. acuminata* genotype cDNA pools for *M. musicola*-inoculated and non-inoculated controls, with three technical replicates for each biological replicate. RNA pools were prepared for cDNA library construction, each comprising nine RNA samples for the pathogen-challenged or unchallenged *M. acuminata* genotype (pooled timepoints of 3, 6 and 9DAI, with triplicate samples). Total RNA were digested with DNase (TURBO DNA-free™, Ambion USA) and converted into cDNA with Super-Script™ II RT and Oligo(dT)20 primer (Invitrogen, Carlsbad, CA, USA). Specific primers were designed using the program Primer3Plus [92], targeting Calcutta 4 defense-related unigene contigs containing SSRs and LRR receptor-like serine/threonine-protein kinase resistance gene analogs (RGAs) in Calcutta 4 and Cavendish Grande Naine. The RGA-targeting primers were designed by mapping all 454 pre-processed transcriptome reads onto annotated Calcutta 4 and Cavendish Grande Naine RGA-containing BAC clones [38], to enable primer design to exon and UTR regions. All RT-qPCR reactions were performed on an ABI 7300 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Each PCR reaction mixture contained 1 µl of template cDNA, primer pairs (175 nM) and Platinum® SYBR® Green qPCR Super Mix-UDG w/ROX kit (Invitrogen, Carlsbad, CA, USA), according to manufacturer's instructions. PCR amplification was conducted with 40 cycles of denaturation at 95°C for 15 s, primer annealing at 60°C for 30 s and

extension at 60°C for 60 s. Gene expression normalization was performed against the internal reference gene coding for an elongation factor for *M. acuminata*. Amplification efficiencies, correlation coefficients, R2 values and relative gene expression (comparative Ct method) were calculated using 7500 v.2.0.4 software (Applied Biosystem, Foster City, CA, USA). All primers used in the study are listed in Table 3.

Additional files

Additional file 1: Mapping data for *M. acuminata* Calcutta 4 and Cavendish Grande Naine unigene contigs aligned to *M. acuminata* DH Pahang all gene models. Data shows all assembled contigs of *M. acuminata* Calcutta 4 and Cavendish Grande Naine generated through *de novo* assembly mapping to *M. acuminata* ssp. *malaccensis* var. Pahang (DH Pahang) whole genome sequence. Specific data is also presented for contigs mapped to cytoplasmic nucleotide binding site-leucine-rich repeat proteins, together with TBLASTX scoring hits obtained by comparison of all unmapped contigs against EST (others) database of GenBank at an E value < e-5.

Additional file 2: Example of *M. acuminata* Calcutta 4 unigene contigs mapping to single *M. acuminata* DH Pahang gene models.

Additional file 3: Venn diagram showing the overlap between transcriptome unigene datasets (contigs and singletons) for the *M. acuminata* genotypes Calcutta 4 and Cavendish Grande Naine. All gene models in the reference *M. acuminata* DH Pahang genome were used as a base for identification of common mapped genes.

Additional file 4: Consensus sequences and BLASTX analysis of assembled unigene contigs of *M. acuminata* Calcutta 4 and Cavendish Grande Naine. Data represent the consensus sequences of assembled contigs of *M. acuminata* Calcutta 4 and Cavendish Grande Naine generated through *de novo* assembly, together with the three best BLASTX scoring hits obtained by comparison against nr database of GenBank at an E value < e-5.

Additional file 5: The most representative InterPro domains (IPR) observed across the assembled unigene contigs of *M. acuminata* Calcutta 4 and Cavendish Grande Naine.

Additional file 6: Analysis of assembled unigene contigs of *M. acuminata* Calcutta 4 and Cavendish Grande Naine for Gene Ontology (GO) terms related to defense response. Assembled contigs of *M. acuminata* Calcutta 4 and Cavendish Grande Naine were assigned to GO terms related to defense.

Additional file 7: Mapping data for *M. acuminata* Calcutta 4 and Cavendish Grande Naine pre-processed 454 sequence reads aligned to *Mycosphaerella fijiensis* v2.0 All Gene Models.

Additional file 8: Sequence information of all the SSR primer pairs designed with Mreps and Primer3 for *M. acuminata* Calcutta 4 and Cavendish Grande Naine unigenes. Data represent all the identified SSR loci, together with supporting information (SSR motif, primer sequences, PCR amplification information, predicted gene function and *in silico* expression data).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MANP participated in bioassays, RNA preparation for cDNA library construction and sequence data analysis. VOC participated in microsatellite marker characterization, validation and data analysis. FLE participated in bioassays, RNA preparation for cDNA library construction and sequence data analysis. CCT conducted bioassays and participated in RNA preparation for cDNA library construction. VCRA participated in microsatellite marker validation and data analysis. AB participated in QRT-PCR and data analysis. EPA participated in bioassays and editing of the manuscript. CFF participated in bioassays and editing

of the manuscript. NFM participated in sequence data analysis and editing of the manuscript. RCT participated in sequence data analysis and editing of the manuscript. GJPJ participated in sequence data analysis and editing of the manuscript. OBSJ participated in sequence data analysis and editing of the manuscript. RINGM conceived the study, participated in bioassays, RNA preparation for cDNA library construction, sequence data analysis, and drafted the manuscript. All authors have contributed to, read and approved the final manuscript.

Acknowledgements

This work was funded by the IAEA (Project 13187), FINEP (Project 0107060900 / 0842/07), CNPq/FAPDF/PRONEX (Project 193.000.5622009) and CNPq Rede EstRESCe (Processo: 564675/2010-5). MANP was supported by a fellowship from the CNPq. VOC and FLE were supported by fellowships from CAPES. We thank anonymous reviewers for their useful comments on the manuscript.

Author details

¹Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Departamento de Biologia Celular, CEP 70.910-900, Brasília, D.F, Brazil. ²Universidade Católica de Brasília, SGAN 916, Módulo B, CEP 70.790-160, Brasília, D.F, Brazil. ³EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, D. F, Brazil. ⁴EMBRAPA Mandioca e Fruticultura Tropical, Rua Embrapa, CEP 44.380-000, Cruz das Almas, BA, Brazil.

Received: 16 July 2012 Accepted: 1 February 2013

Published: 5 February 2013

References

1. FAO/STAT: *Food and agricultural commodities production. Food and Agricultural Organization Statistics*. 2012. <http://faostat.fao.org/>.
2. Jones DR: *Diseases of banana, abacá and enset*. Wallingford, Oxon: CABI Publishing; 1999.
3. Jones DR: *Disease and pest constraints to banana production. Acta Hort* 2009, **828**:21–36.
4. Mouliom-Pefoura A, Lassoudière A, Foko J, Fontem DA: *Comparison of development of Mycosphaerella fijiensis and Mycosphaerella musicola on banana and plantain in the various ecological zones in Cameroon. Plant Dis* 1996, **80**:950–954.
5. James AC, Arzanlou M, Canto-Canche B, Jorge HR, Ferraez CL, Echeverría PS: *Fungal Diseases of Banana*. In *Bananas: Nutrition, Diseases and Trade Issues*. Edited by Cohen AE.: Nova Science Publishers Inc; 2011:65–122.
6. Cordeiro ZJM, Matos AP, Meissner Filho PE: *Doenças e métodos de controle*. In *O Cultivo da Bananeira*, Volume 1. 1st edition. Edited by Borges AL, Souza LS. Cruz das Almas: Nova Civilização: Cruz das Almas: Nova Civilização; 2004:146–182.
7. Cordeiro ZJM, Matos AP, Kimati H: *Doenças Da Bananeira (Musa Spp.)*. In *Manual de Fitopatologia*, Volume 2. 4th edition. Edited by Kimati H, Amorim L, Rezende JAM, Bergamin Filho A, Camargo LEA. São Paulo: Agronômica Ceres; 2005:99–117.
8. Marín DH, Romero RA, Guzmán M, Sutton TB: *Black Sigatoka: an increasing threat to banana cultivation. Plant Dis* 2003, **87**:208–222.
9. Sierotzki H, Parisi S, Steinfeld U, Tenzer I, Poirey S, Gisi U: *Mode of resistance to respiration inhibitors at the cytochrome bc1 enzyme complex of Mycosphaerella fijiensis field isolates. Pest Manag Sci* 2000, **56**:833–841.
10. Heslop-Harrison JS, Schwarzbacher T: *Domestication, genomics and the future for banana. Ann Bot* 2007, **100**:1073–1084.
11. Ortiz R: *Secondary polyploids, heterosis, and evolutionary crop breeding for further improvement of the plantain and banana (Musa spp. L.) genome. Theor Appl Genet* 1997, **94**:1113–1120.
12. Horry JP: *The use of Molecular Markers in the CIRAD Banana Breeding Programme. Acta Hort* 2011, **897**:237–243.
13. Bakry F, Carreel F, Jenny C, Horry JP: *Genetic Improvement of Banana*. In *Breeding Plantain Tree Crops: Tropical Species*. Edited by Jain SM, Priyadarshan PM. New York: Springer Verlag Publisher; 2009:3–46.
14. Lysak MA, Dolezelova M, Horry JP, Swennen R, Dolezel J: *Flow cytometric analysis of nuclear DNA content in Musa. Theor Appl Genet* 1999, **98**:1344–1350.
15. D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, Da Silva C, Jabbari K, Cardic C, Poulain J, Souquet M, Labadie K, Jourda C, Lenglé J, Rodier-Goud M, Alberti A, Bernard M, Correa M, Ayyampalayam S, Mckain MR, Leebens-Mack J, et al: *The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature* 2012, doi:10.1038/nature11241.
16. Roux N, Baurens FC, Dolezel J, Hribova E, Heslop-Harrison P, Town C, Sasaki T, Matsumoto T, Aert R, Remy S, Souza M, Lagoda P: *Genomics of Banana and Plantain (Musa spp.), Major Staple Crops in the Tropics*. In *Genomics of Tropical Crop Plants*. Edited by Moore PH, Ming R. New York: Springer; 2008:83–111.
17. Manrique-Trujillo SM, Ramírez-López AC, Ibarra-Laclette E, Gómez-Lim MA: *Identification of genes differentially expressed during ripening of banana. J Plant Physiol* 2007, **164**:1037–1050.
18. Jin ZQ, Xu BY, Liu JH, Zhang ZB, Jia CH: *Combination of suppression-subtractive hybridization with cDNA microarray, a novel way to identify genes from banana involved in fruit ripening. Acta Hort* 2011, **897**:195–206.
19. Santos CMR, Martins NF, Horberg HM, de Almeida ERP, Coelho MCF, Togawa RC, da Silva FR, Caetano AR, Miller RNG, Souza MT Jr: *Analysis of expressed sequence tags from Musa acuminata ssp. burmannicoides, var. Calcutta 4 (AA) leaves submitted to temperature stresses. Theor Appl Genet* 2005, **110**:1517–1522.
20. Ravishankar KV, Rekha A, Laxman RH, Savitha G, Swarupa V: *Gene Expression Analysis in Leaves of 'Bee Hee Kela' a drought-Tolerant Musa balisiana Genotype from Northeast India. Acta Hort* 2011, **897**:279–280.
21. Nyine M, Nanteza A, Chan A, Town C, Lorenzen J: *Insights from a 454-based Reference Transcriptome for Banana From Drought Stressed and Control Tissues*. In *Proceedings of the XX International Plant and Animal Genome Conference (PAG)*. San Diego; 2012:14–18. W070.
22. Van den Berg N, Berger DK, Hein I, Birch PR, Wingfield MJ, Viljoen A: *Tolerance in banana to Fusarium wilt is associated with early up-regulation of cell wall strengthening genes in the roots. Mol Plant Pathol* 2007, **8**:333–341.
23. Portal O, Izquierdo Y, De Vleeschauwer D, Sánchez-Rodríguez A, Mendoza-Rodríguez M, Acosta-Suárez M, Ocaña B, Jiménez E, Höfte M: *Analysis of expressed sequence tags derived from a compatible Mycosphaerella fijiensis-banana interaction. Plant Cell Rep* 2011, **30**:913–928.
24. Passos MAN, de Oliveira Cruz V, Emediato FL, de Camargo Teixeira C, Souza MT Jr, Matsumoto T, Rennó Azevedo VC, Ferreira CF, Amorim EP, de Alencar Figueiredo LF, Martins NF, de Jesus Barbosa Cavalcante M, Baurens FC, da Bonfim Silva O Jr, Pappas GJ Jr, Pignolet L, Abadie C, Ciampi AY, Piffanelli P, Miller RNG: *Development of expressed sequence tag and EST-SSR marker resources for Musa acuminata. Aob Plants* 2012, doi:10.1093/aobpla/pls030. pls030.
25. Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB: *Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol* 2007, **144**:32–42.
26. Yang H, Hu L, Hurek T, Reinhold-Hurek B: *Global characterization of the root transcriptome of a wild species of rice, Oryza longistaminata, by deep sequencing. BMC Genomics* 2010, **11**:705.
27. Yang SS, Tu ZJ, Cheung F, Xu WW, Lamb JFS, Jung HJG, Vance CP, Gronwald JW: *Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. BMC Genomics* 2011, **12**:199.
28. Liu Z, Crampton M, Todd A, Kalavacharla V: *Identification of expressed resistance gene-like sequences by data mining in 454-derived transcriptomic sequences of common bean (Phaseolus vulgaris L.). BMC Plant Biol* 2012, **12**:42.
29. Mahomed W, Van den Berg N: *EST sequencing and gene expression profiling of defence-related genes from Persea americana infected with Phytophthora cinnamomi. BMC Plant Biol* 2011, **11**:167.
30. Danielsson M, Lundén K, Elfstrand M, Hu J, Zhao T, Amerup J, Ihrmark K, Swedjemark G, Borg-Karlson AK, Stenlid J: *Chemical and transcriptional responses of Norway spruce genotypes with different susceptibility to Heterobasidion spp. infection. BMC Plant Biol* 2011, **11**:154.
31. Narina SS, Buyyarapu R, Kottapalli KR, Sartie AM, Ali MI, Robert A, Hodeba MJ, Sayre BL, Scheffler BE: *Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (Dioscorea alata L.). BMC Genomics* 2011, **12**:100.
32. Fernandez D, Tisserant E, Talhinhas P, Azinheira H, Vieira A, Petitot AS, Loureiro A, Poulain J, da Silva C, Silva MD, Duplessis S: *454-pyrosequencing of Coffea arabica leaves infected by the rust fungus Hemileia vastatrix reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. Mol Plant Pathol* 2011, **13**:17–37.
33. Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE: *Comparison of the transcriptomes of American*

- chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biol* 2009, **9**:11.
34. Uma S, Mustafa MM, Saraswathi MS, Durai P: **Exploitation of diploids in Indian breeding programmes.** *Acta Hort* 2011, **897**:215–223.
35. Cheung F, Town CD: **A BAC end view of the *Musa Accuminata* genome.** *BMC Plant Biol* 2007, **7**:29.
36. Lescot M, Piffanelli P, Ciampi AY, Ruiz M, Blanc G, Leebens-Mack J, da Silva FR, Santos CM, D'Hont A, Garsmeur O, Vilarinhos AD, Kanamori H, Matsumoto T, Ronning CM, Cheung F, Haas BJ, Althoff R, Arbogast T, Hine E, Pappas GJ Jr, Sasaki MT Jr, Souza MT Jr, Miller RNG, Glaszmann JC, Town CD: **Insights into the *Musa* genome: syntenic relationships to rice and between *Musa* species.** *BMC Genomics* 2008, **9**:58.
37. Azhar M, Heslop-Harrison JS: **Genomes, diversity and resistance gene analogues in *Musa* species.** *Cytogenet Genome Res* 2008, **121**:59–66.
38. Miller RNG, Bertoli DJ, Baurens FC, Santos CM, Alves PC, Martins NF, Togawa RC, Souza MT Junior, Pappas GJ Junior: **Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: Isolation, RFLP marker development, and physical mapping.** *BMC Plant Biol* 2008, **8**:15.
39. Beveraggi A, Mourichon X, Salle G: **Etude des interactions hôte-parasite chez des bananiers sensibles et résistants inoculés par *Cercospora fijiensis* (*Mycosphaerella fijiensis*) responsable de la maladie des raies noires.** *Can J Bot* 1995, **73**:1328–1337.
40. Arzanlou M, Abeln ECA, Kema GHJ, Waalwijk C, Carlier J, de Vries I, Guzman M, Crous PW: **Molecular diagnostics for the Sigatoka disease complex of banana.** *Phytopathology* 2007, **97**:1112–1118.
41. Donzelli BGG, Churchill ACL: **A quantitative assay using mycelial fragments to assess virulence of *Mycosphaerella fijiensis*.** *Phytopathology* 2007, **97**:916–929.
42. Churchill ACL: ***Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control.** *Mol Plant Path* 2011, **12**:307–328.
43. Liu G, Kennedy R, Greenfields DL, Peng G, Forseille L, Selvaraj G, Wei Y: **Detached and attached *Arabidopsis* leaf assays reveal distinctive defense responses against hemibiotrophic *Colletotrichum* spp.** *Mol Plant Microbe Interact* 2007, **20**:1308–1319.
44. Nürnberger T, Kemmerling B: **PAMP-triggered basal immunity in plants.** *Adv Bot Res* 2009, **51**:1–38.
45. Jones JDG, Dangl JL: **The plant immune system.** *Nature* 2006, **444**:323–329.
46. Morel JB, Dangl JL: **The hypersensitive response and the induction of cell death in plants.** *Cell Death Differ* 1997, **4**:671–683.
47. Choi D, Kim HM, Yun HK, Park JA, Kim WT, Bok SH: **Molecular cloning of a metallothionein-like gene from *Nicotiana glutinosa* L. and its induction by wounding and tobacco mosaic virus infection.** *Plant Physiol* 1996, **112**:353–359.
48. Hamer DH: **Metallothionein.** *Annu Rev Biochem* 1986, **55**:913–951.
49. Liu P, Goh CJ, Loh CS, Pua EC: **Differential expression and characterization of three metallothionein-like genes in Cavendish banana (*Musa acuminata*).** *Physiol Plant* 2002, **114**:241–250.
50. Cavalcante MJB, Escoute J, Madeira JP, Romero RE, Nicole M, Oliveira LC, Hamelin C, Lartaud M, Verdeil JL: **Reactive oxygen species and cellular interactions between *mycosphaerella fijiensis* and banana.** *Trop Plant Biol* 2011, **4**:134–143.
51. Hammond-Kosack KE, Jones JDG: **Plant disease resistance genes.** *Annu Rev Plant Physiol Plant Mol Biol* 1997, **48**:573–605.
52. McHale L, Tan X, Koehl P, Micheltore RW: **Plant NBS-LRR proteins: adaptable guards.** *Genome Biol* 2006, **7**:212. 1–212. 11.
53. Mohamad A, Heslop-Harrison JS: **Genomes, diversity and resistance gene analogs in *Musa* species.** *Cytogenet Genome Res* 2008, **121**:59–66.
54. Shiu SH, Bleecker AB: **Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases.** Proceedings of the National Academy of Science of the United States of America, Volume 98; 2001:10763–10768.
55. Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C: **A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21.** *Science* 1995, **270**:1804–1806.
56. Gómez-Gómez L, Boller T: **FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*.** *Mol Cell* 2000, **5**:1003–1011.
57. Mohr TJ, Mammarella ND, Hoff T, Woffenden BJ, Jelesko JG, McDowell JM: **The *Arabidopsis* Downy Mildew Resistance Gene *RPP8* Is Induced by Pathogens and Salicylic Acid and Is Regulated by W Box cis Elements.** *MPMI* 2010, **23**:1303–1315.
58. Goodwin SB, Dunkle LD, Zismann VL: **Phylogenetic Analysis of *Cercospora* and *Mycosphaerella* Based on the Internal Transcribed Spacer Region of Ribosomal DNA.** *Phytopathology* 2001, **91**:648–658.
59. Bolton MD, van Esse HP, Vossen JH, de Jonge R, Stergiopoulos I, Stulemeijer IJE, van den Berg GCM, Borrás-Hidalgo O, Dekker HL, de Koster CG, de Wit PJGM, Joosten MHJ, Thomma BPHJ: **The novel *Cladosporium fulvum* lysine motif effector Ecp6 is a virulence factor with orthologs in other fungal species.** *Mol Microbiol* 2008, **69**:119–136.
60. de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MH, Thomma BP: **Conserved fungal LysEffector Ecp6 prevents chitin-triggered immunity in plants.** *Science* 2010, **329**:953–955.
61. Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, Thomma BP, Rudd JJ: **Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat.** *Plant Physiol* 2011, **156**:756–769.
62. Cho Y, Hou S, Zhong S: **Analysis of expressed sequence tags from the fungal banana pathogen *Mycosphaerella fijiensis*.** *Open Mycol J* 2008, **2**:61–73.
63. Lagoda PJJ, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C: **Sequence tagged microsatellite site (STMS) markers in the Musaceae.** *Mol Ecol* 1998, **7**:657–666.
64. Christelová P, Valárik M, Hřibová E, Van den Houwe I, Channelière S, Roux N, Doležal J: **A platform for efficient genotyping in *Musa* using microsatellite markers.** *AoB Plants* 2011, doi:10.1093/aobpla/plr0245. plr024.
65. Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci AM, Jenny C, Perrier X, Carreel F, Argout X, Piffanelli P, Khan I, Miller RNG, Pappas GJ, Mbeguie-a-mbeguie D, Matsumoto T, De Bernardinis V, Huttner E, Kilian A, Baurens FC, D'Hont A, Cote F, Courtois B, Glaszmann JC: **A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas.** *BMC Plant Biol* 2010, **10**:65.
66. Tranbarger TJ, Kluabmongkol W, Sangsraku D, Morcillo F, Tregear JW, Tragoonrun S, Billotte N: **SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of *Elaeis guineensis*.** *BMC Plant Biol* 2012, **12**:1.
67. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS: **Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat.** *Mol Genet Genomics* 2003, **270**:315–323.
68. Kaemmer D, Fisher D, Jarret RL, Baurens FC, Grapin A, Dambier D, Noyer JL, Lanaud C, Kahl G, Lagoda PJJ: **Molecular breeding in genus *Musa*: a strong case for SSR marker technology.** *Euphytica* 1997, **96**:49–62.
69. Bahariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R: **Isolation and characterization of microsatellite markers from *Musa balbisiana*.** *Mol Ecol Notes* 2005, **5**:327–330.
70. Miller RN, Passos MA, Menezes NN, Souza MT Jr, Costa MMC, Azevedo VCR, Amorim EP, Pappas GJ Jr, Ciampi AY: **Characterization of novel microsatellite markers in *Musa acuminata* subsp. burmannicoides, var. Calcutta 4.** *BMC Res Notes* 2010, **3**:148.
71. Ravishankar KV, Vidhya L, Cyriac A, Rekha A, Goel R, Singh NK, Sharma TR: **Development of SSR markers based on a survey of genomic sequences and their molecular analysis in banana (*Musa* spp.).** *J Hort Sci Biotech* 2012, **87**:84–88.
72. Wang JY, Chen YY, Liu WL, Wu YT: **Development and application of EST-derived SSR markers for bananas (*Musa nana* Lour.).** *Hereditas* 2008, **30**:933–940.
73. Amorim EP, Silva PH, Ferreira CF, Amorim VBO, Santos VJ, Vilarinhos AD, Santos CMR, Souza Júnior MT, Miller RNG: **New microsatellite markers for bananas (*Musa* spp.).** *Genet Mol Res* 2012, **11**:1093–1098.
74. Thiel T, Michalek W, Varshney RK, Graner A: **Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.).** *Theor Appl Genet* 2003, **106**:411–422.
75. Varshney RK, Graner A, Sorrells ME: **Genic microsatellite markers in plants: features and applications.** *Trends Biotechnol* 2005, **23**:48–55.
76. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ: **Analysis of SSRs derived from grape ESTs.** *Theor Appl Genet* 2000, **100**:723–726.
77. Raju NL, Gnanesh BN, Lekha P, Jayashree B, Pande S, Hiremath PJ, Byregowda M, Singh NK, Varshney RK: **The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.).** *BMC Plant Biol* 2010, **10**:4.
78. Amorim EP, Lessa LS, da Silva Ledo CA, de Oliveira Amorim VB, Viana dos Reis R, Santos-Serejo JA, de Oliveira e Silva S: **Caracterização agrônômica e**

- molecular de genótipos diplóides melhorados de bananeira. *Rev Bras Frutic* 2009, **31**:154–161.
79. Dochez C, Tenkouano A, Ortiz R, Whyte J, De Waele D: **Host plant resistance to *Radopholus similis* in a diploid banana hybrid population.** *Nematology* 2009, **11**:329–335.
 80. Rekha A, Ravishankar KV, Ambika DS: **Generation of segregating F1 populations for mapping the *Musa* genome.** *Acta Hort* 2011, **897**:277–278.
 81. Papanicolaou A, Stierli R, Ffrench-Constant RH, Heckel DG: **Next generation transcriptomes for next generation genomes using est2assembly.** *BMC Bioinforma* 2009, **10**:447.
 82. Ning Z, Cox AJ, Mullikin JC: **SSAHA: a fast search method for large DNA databases.** *Genome Res* 2001, **11**:1725–1729.
 83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
 84. Hazelhurst S, Hide W, Lipták Z, Nogueira R, Starfield R: **An overview of the wcd EST clustering tool.** *Bioinformatics* 2008, **24**:1542–1546.
 85. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S: **Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs.** *Genome Res* 2004, **14**:1147–1159.
 86. Wu TD, Watanabe CK: **GMAP: a genomic mapping and alignment program for mRNA and EST sequences.** *Bioinformatics* 2005, **21**:1859–1875.
 87. Zdobnov EM, Apweiler R: **InterProScan - an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**:847–848.
 88. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674–3676.
 89. Gene Ontology Consortium: **The Gene Ontology project in 2008.** *Nucleic Acids Res* 2008, **36**:D440–D444.
 90. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754–1760.
 91. Rozen S, Skaletsky HJ: **Primer3 on the WWW for general users and for biologist programmers.** In *Bioinformatics Methods and Protocols; Methods in Molecular Biology*. Edited by Krawetz S, Misener S. New Jersey: Humana Press; 2000:365–386.
 92. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM: **Primer3Plus, an enhanced web interface to Primer3.** *Nucleic Acids Res* 2007, **35**:W71–W74.

doi:10.1186/1471-2164-14-78

Cite this article as: Passos et al.: Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. *BMC Genomics* 2013 **14**:78.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Analysis of Expression of NBS-LRR Resistance Gene Analogs in *Musa acuminata* during Compatible and Incompatible Interactions with *Mycosphaerella musicola*

F.L. Emediato¹, M.A.N. Passos¹, C. de Camargo Teixeira²,
G.J. Pappas Jr.^{2,3} and R.N.G. Miller¹

¹ Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Departamento de Biologia Celular, Asa Norte, CEP 70.910-900, Brasília, DF, Brazil

² Universidade Católica de Brasília, SGAN 916, Módulo B, CEP 70.790-160, Brasília, DF, Brazil

³ EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, DF, Brazil

Keywords: banana improvement, disease resistance mechanisms, gene expression, RGA, Sigatoka leaf spot

Abstract

Sigatoka leaf spot, caused by the fungus *Mycosphaerella musicola*, significantly reduces leaf area and causes premature fruit ripening in banana. It is thus fundamentally important to develop genotypes resistant to fungal pathogens via crop improvement. In this context, transcriptional activity of NBS-LRR Resistance Gene Analogs (RGAs) was compared from whole plant leaf material from *Musa acuminata* ssp. *burmannicoides* ‘Calcutta-4’ (resistant) and Cavendish ‘Grande Naine’ (AAA, susceptible), uninoculated and inoculated in vivo with *M. musicola* pathogen conidiospores. Total RNA was extracted and purified using Concert™ and cDNA prepared using the Kit SuperScript™ II Reverse Transcriptase (Invitrogen). Twenty specific primer pairs for reverse transcriptase (RT) PCR were designed using Primer3plus and targeting NBS-LRR RGA contig exon sequences. Primers were also designed for the constitutively expressed *M. acuminata* actin gene (gi 151413782) and elongation factor (gi 66775523). RT-PCR results revealed examples of constitutive RGA gene expression in both compatible and incompatible interactions across the infection time course, as well as a complete absence of gene expression. Examples of RGAs showing similar expression patterns in compatible and incompatible interactions were also apparent, with susceptible and resistant cultivars, for example, showing low levels of expression in early infection stages and up-regulation in late stages. In contrast, another RGA displayed constitutive early expression in the resistant cultivar, strong down-regulation in the intermediate infection stage, and recovery to higher levels in the final stage of the examined infection time course. This project will contribute to the elucidation of disease resistance mechanisms in *Musa*. Candidate gene application via plant transformation or marker-assisted selection will benefit *Musa* breeding programs for the generation of resistant genotypes.

INTRODUCTION

Evolution of *Musa* via asexual reproduction in the form of vegetative micropropagation or suckers means that today's crop has a restricted genetic base, which lacks resistance to pests and diseases. Sigatoka leaf spot disease, also known as Yellow Sigatoka, caused by the fungal species *Mycosphaerella musicola*, is today one of the principal problems for banana cultivation in Brazil, affecting fruit quality and reducing productivity by up to 50% (Cordeiro et al., 2005; Meredith, 1970). Long-term reliance upon chemical control strategies is inadequate, both socio-economically and environmentally, with integrated strategies required for avoidance of fungicide resistance development in the pathogen.

Non-conventional breeding strategies, such as transformation and marker-assisted

selection, offer an alternative approach for introgression of disease resistance and defense genes in commercial cultivars. However, relatively few such candidate genes have been isolated and characterized to date for banana.

The objectives of this study were to analyze transcriptional activity of NBS-LRR Resistance Gene Analogs (RGAs) in *Musa acuminata* leaf materials, during in vivo interactions with *M. musicola*.

MATERIALS AND METHODS

Bioassays

Musa acuminata ssp. *burmanicoides* 'Calcutta-4' (resistant) and Cavendish 'Grande Naine' (AAA, susceptible) were propagated in vitro and subsequently maintained for 6 months in a greenhouse at 25°C. A conidiospore suspension from a pathogenic isolate of *M. musicola* (source Embrapa CNPMF, Brazil) was prepared at a concentration of 2×10^4 spores/ml, and sprayed onto the abaxial surface of the two youngest leaves of nine plants each of 'Calcutta-4' and 'Grande Naine'. An identical number of plants was included as non-inoculated controls. A histopathological examination of the infection process was performed using scanning electron microscopy. Hyphal colonization and penetration of stomata was observed 6 days after inoculation (Fig. 1). Infected leaf material was collected at 3, 6 and 9 days after inoculation.

RNA Extraction

Total RNA was extracted from all samples of inoculated and uninoculated leaves using Concert Plant RNA Reagent[®] (Invitrogen) according to manufacturer's instructions. RNA quantification and purity analysis was performed using 1% agarose gels and visual inspection of 28S and 18S rRNA and via Nanodrop spectrophotometry. cDNA was prepared using the Kit SuperScript[™] II Reverse Transcriptase (Invitrogen).

Primer Design and PCR Amplification

Twenty specific primer pairs were designed using the program Primer3plus (Untergasser et al., 2007) for expression analysis from *M. acuminata* NBS-LRR RGA contig exon sequences. Control primers for potential constitutively expressed genes were designed for targeting the *M. acuminata* actin gene (gi 151413782) and elongation factor (gi 66775523) in the NCBI database. Successfully tested primer sequences are listed in Table 1. PCR reactions were carried out in 25 µl volumes, containing 20 ng cDNA template, 1.5 mM MgCl₂, 0.2 mM dNTPs, 0.1 µM of each primer, 1.25U of Taq polymerase, and 1X PCR buffer (Invitrogen). Temperature cycling was conducted on a Veriti thermocycler (Applied Biosystems) as follows: initial denaturation at 94°C for 3 min; 40 cycles of 94°C for 1 min, specific primer annealing temperature for 1 min, and extension at 72°C for 1 min; plus a final elongation period of 7 min at 72°C. PCR products were electrophoresed in 1.5% agarose gels.

RESULTS AND DISCUSSION

Gene Expression

Ongoing reverse transcriptase (RT) PCR results have revealed examples of constitutive gene expression for RGAs in both compatible and incompatible interactions across the infection time course, as well as a complete absence of gene expression. Examples of similar expression patterns were observed in compatible and incompatible interactions, with susceptible and resistant cultivars, showing low levels of RGA expression in the early infection stages and up-regulation in late stages. In contrast, other RGAs displayed high expression levels in the resistant cultivar during the early infection stage, strong down-regulation in the intermediate infection stage, and recovery to higher levels in the final stages of the infection time course (Fig. 2).

CONCLUSIONS

This project will contribute to the elucidation of resistance mechanisms in *Musa*. As RGA sequences are characterized, through conserved orthologous sequence (COS) marker-derived amplification from *Musa* genotypes, as well as through transcription profiling approaches during *Musa-Mycosphaerella* interactions, temporal expression profiles in candidate genes potentially involved in resistance or defense responses in *Musa* are being validated via RT-PCR and qRT-PCR along interaction time courses. In addition to increasing our understanding of the molecular processes involved in disease resistance in *Musa*, candidate genes and markers offer potential for the development of effective disease management based upon genetic improvement of the crop via both plant transformation and conventional marker-assisted breeding.

ACKNOWLEDGEMENTS

This work was partially funded by the IAEA (Project 13187), FINEP (Project 0107060900 / 0842/07) and CNPq / FAPDF / PRONEX (Project 193.000.5622009). Flavia Leonel Emediato and Marco Aurelio Ninomia Passos gratefully acknowledge CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) Doctorate degree scholarships from the Brazilian Ministry of Education (CNPq).

Literature Cited

- Cordeiro, Z.J.M., Matos, A.P. and Kimati, H. 2005. Doenças Da Bananeira (*Musa* spp.). p.99-117. In: H. Kimati, L. Amorin, J.A.M. Rezende, A. Bergamin Filho and L.E.A. Camargo (eds.), Manual de Fitopatologia, 4th edition. Agronômica Ceres Ltda., São Paulo.
- Meredith, M.A. 1970. Banana leaf spot disease (Sigatoka) caused by *Mycosphaerella musicola* Leach. Phytopathological Papers 11:7-21.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J.A. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71-W74.

Tables

Table 1. Primer sequences for RT-PCR analysis of expression of a constitutive gene and NBS-LRR contigs in *Musa acuminata*.

Primer name	Primer sequence (5'-3')	PCR product (bp)	T _m (°C) ¹
NBS-LRR_RGA_contig_02	F: GCACGGAGCACTGTATCTGA	233	60.0
	R: GACCGTAAGGTTGCGATTGT		60.0
NBS-LRR_RGA_contig_03	F: TTTCGAGCTTTGTGCATTTG	121	60.0
	R: CAGTGCCTTCCCTGCTAAAG		60.0
NBS-LRR_RGA_contig_07	F: TGGGAGGAGATACCTGCTTG	130	60.2
	R: TTCTTCGATTCTGGTGGTCA		59.2
NBS-LRR_RGA_contig_09	F: CCTTGTGGGAAACTCCGTAA	722	60.0
	R: CAACTGCCACAAGCTGAAAA		60.0
ACTIN1	F: GAATGGTCAAGGCTGGTTTC	109	59.5
	R: TCTTTTTGACCCATCCCAAC		59.8

¹T_m, annealing temperature used.

Figures

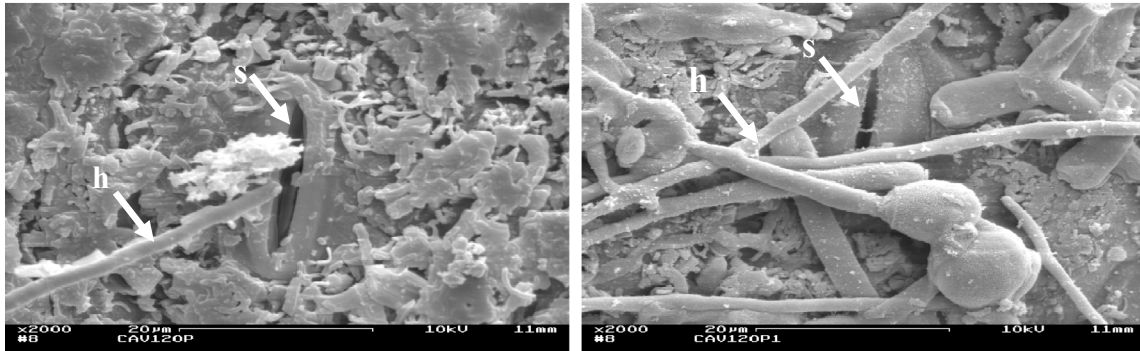


Fig. 1. Scanning electronic microscopy (SEM) images of colonization of cultivar ‘Grande Naine’ (AAA, Cavendish) leaf abaxial surfaces by *Mycosphaerella musicola* hyphae, 6 days after inoculation. Abbreviations: h, hypha; s, stoma.

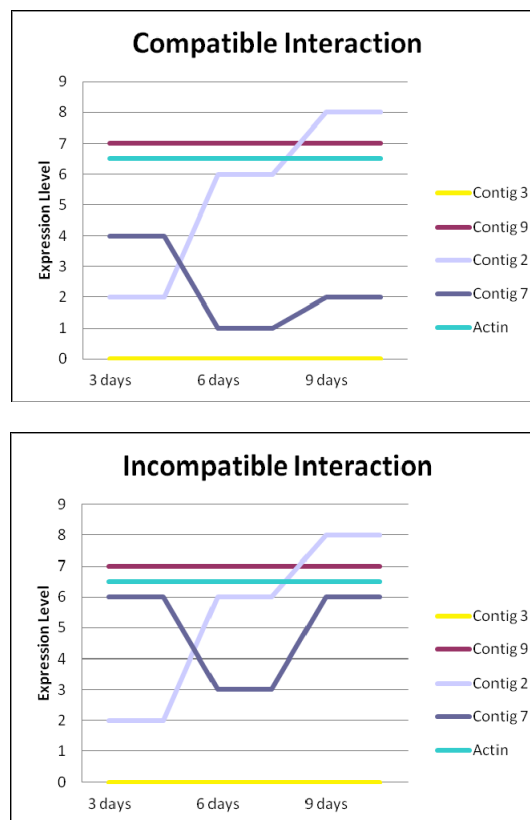


Fig. 2. Relative levels of expression of NBS-LRR Resistance Gene Analogs (RGAs) at different times during the (A) compatible (‘Grande Naine’, AAA, Cavendish) and (B) incompatible (‘Calcutta 4’, *Musa acuminata* ssp. *burmannicoides*) interaction with *Mycosphaerella musicola*.

Understanding Plant Immunity: Transcriptome Profiling in *Musa*-Pathogen Interactions Using Next Generation Sequencing

M.A.N. Passos¹, F.L. Emediato¹, V.O. Cruz¹, C. de Camargo Teixeira²,
L.F. de Alencar Figueiredo¹, N.F. Martins³, R.C. Togawa³, M.M.C. Costa³,
O. Silva Jr.³, G.J. Pappas Jr.^{2,3} and R.N.G. Miller¹

¹ Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Departamento de Biologia Celular, Asa Norte, CEP 70.910-900, Brasília, DF, Brazil

² Universidade Católica de Brasília, SGAN 916, Módulo B, CEP 70.790-160, Brasília, DF, Brazil

³ EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, DF, Brazil

Keywords: *Musa acuminata*, *Mycosphaerella musicola*, 454 transcriptome sequencing

Abstract

The development of novel approaches for crop protection requires continued advances in our understanding of the molecular mechanisms controlling plant immunity. Molecular and genomics tools have advanced our understanding, with two key branches currently recognized. In one, plant innate defense is governed by PAMP-triggered immunity (PTI), following host recognition of pathogen-associated molecular patterns (PAMPs). Successful pathogens can, however, suppress PTI signaling through evolution of specific effector proteins. Plants, in response, have co-evolved cytoplasmic resistance (R) protein receptors which establish effector-triggered immunity (ETI), a second immune system branch which recognizes specific pathogen effectors. Although banana (*Musa* spp.) is one of the world's most important edible crops, contributing towards food security, a comprehensive transcriptomic dataset is not yet available for use in accelerated molecular-based breeding. In order to develop a functional genomics resource for this crop which reveals transcriptional changes during plant immunity responses to biotic stresses, we performed a pyrosequencing study of expressed genes in *Musa acuminata* genotypes during compatible and incompatible reactions with the fungal pathogen *Mycosphaerella musicola*, causal organism of Sigatoka leaf spot disease. Total RNA samples were prepared from whole plant leaf material from 'Calcutta 4' (*M. acuminata* ssp. *burmannicoides*, resistant) and 'Grande Naine' (AAA, Cavendish, susceptible), both uninfected and artificially challenged with pathogen conidiospores. Full-length enriched cDNA libraries were sequenced using a 454 GS-FLX system pyrosequencer with Titanium chemistry, generating 978,133 raw sequencing reads, with an average length of 334 bp and totaling over 460 million bp. Over 35,000 unigenes were assembled for each genotype, with approximately 35% displaying no significant similarity to any sequences in the public databases. In silico analysis identified differentially expressed genes associated with stresses and responses to biotic stimuli. Datasets were exploited for identification of expressed resistance gene analogs and defense genes, as well as large-scale SSR marker development. Together with future functional analysis and genetic map enrichment approaches, the resources generated will contribute to our understanding of plant immunity processes in *Musa*, facilitating long-term disease management based upon genetic improvement.

PLANT IMMUNITY

The development of novel approaches for crop protection requires continued advances in our understanding of the molecular mechanisms governing plant immunity. Over the last two decades, molecular and genomics tools have advanced considerably our understanding of the plant immune system, with two key branches currently recognized,

as shown in simplified form in Figure 1.

PAMP-Triggered Immunity

One branch of the plant immune system is known as PAMP-triggered immunity (PTI), innate defense, basal defense or non-self recognition. This branch is based upon interactions between host surface pattern recognition receptor-like kinases (PRRs) (Shiu and Bleecker, 2001; Altenbach and Robatzek, 2007), in plant extracellular matrix or plasma membrane (Boller and Felix, 2009) which recognize conserved pathogen-associated molecular patterns (PAMPs) (Nürnberg and Kemmerling, 2009). These are microbial epitopes, such as bacterial flagellin, bacterial elongation factor Tu, lipopolysaccharides, peptidoglycan, β -glucans, ergosterol or fungal chitin (Boller and Felix, 2009). Such epitopes, originally termed elicitors, are defined as conserved molecules across a microbial pathogen class and considered essential in function in fitness or survival (Medzhitov and Janeway, 1997). Given that conserved molecules can also occur in non-pathogenic microorganisms, the term microbe-associated molecular patterns (MAMPs) is also employed (Boller and Felix, 2009). Breakdown products of the plant-host cell wall, also serve as endogenous elicitors. They are released as a result of the presence of non-host pathogen hydrolytic enzymes (Vorwerk et al., 2004). Such host-pathogen interactions elicit PTI, which involves the activation of a mitogen-associated protein (MAP) kinase cascade and activation of WRKY transcription factors (TFs), preventing further host colonization, and conferring resistance to the majority of potential pathogens. PTI is considered to be responsible for non-host resistance and can occur at the species level or above. Plants typically show similar PTI responses to recognition of diverse PAMPs.

Effector-Triggered Immunity

In a second branch of the plant immune system, successful pathogens can overcome PTI. Pathogen avirulence (*Avr*) genes have evolved to code for species-, race-, or even strain-specific virulence effectors (previously referred to as avirulence factors). Upon secretion and translocation into the host cell, these genes suppress or interfere with PTI signaling, resulting in plants being susceptible to pathogen infection, with disease as a result of this Effector-Triggered Susceptibility (ETS) (Chisholm et al., 2006; da Cunha et al., 2007; Block et al., 2008; Gohre and Robatzek, 2008). As such, effectors differ from PAMPs, which are pathogen molecules essential for microbial survival, in that they specifically target host defense mechanisms. Bacterial pathogens can transfer such effectors into a susceptible host cell via type III secretion systems. In response, plants have co-evolved specific cytoplasmic resistance R protein receptors which recognize individual pathogen effector molecular signatures (such that the virulence protein is now classed as an avirulence protein). These activate a second line of defense known as Effector-Triggered Immunity (ETI) (Jones and Dangl, 2006), also previously known as gene-for-gene or race-specific resistance. As such, ETI appears to have evolved after PTI, with acquisition of necessary R proteins. In contrast to PTI, which will occur in all members of a particular plant species, ETI operates at the intra-specific level, with resistant genotypes possessing the necessary polymorphism in R gene resistance determinants (Xiao et al., 2008).

A number of distinct plant resistance gene families have now been recognized as involved in ETI, based upon deduced protein domain structure and biochemical function (for a review see Miller et al., 2009). The largest class encodes intracellular receptor proteins within a superfamily containing nucleotide binding site-leucine-rich repeat (NBS-LRR) domains (Baker et al., 1997; Hammond-Kosack and Jones, 1997). Pathogen effector recognition occurs across pathogen kingdoms, including bacteria, oomycetes, fungi and nematodes, with over 60 plant R genes isolated over the last 20 years (Xiao, 2006). Conservation of motifs within R genes, such as those present within nucleotide-binding site leucine-rich repeat domains, have facilitated amplification in diverse plant taxa. Putative R genes or resistance gene analogs (RGAs) are commonly clustered, as a

result of duplication events occurring under diversifying selection. In the rice genome, approximately 400 NBS-LRR genes have been characterized, with around 150 present in the *Arabidopsis* genome (McHale et al., 2006).

Four important models within ETI have emerged, which initiated with the gene-for-gene model, and subsequently the zigzag model, guard model, and most recently the decoy model. These important models can be summarized as follows:

- a. In the gene-for-gene model (Flor, 1942), which was originally described to explain *Melampsora lini*/flax pathosystem, a direct interaction between a plant in possession of a resistance factor and a pathogen with a cognate avirulence (Avr) factor results in a resistance response. If either the plant is lacking the R-gene or the pathogen the cognate Avr factor, infection will occur. Within this model based upon receptor-ligand interaction (Keen, 1990), numerous pathogen *Avr* genes and cognate plant *R* genes have now been characterized, with the complexity of the gene-for-gene relationship continuing to be unraveled.
- b. In the zigzag model, once a pathogen Avr factor or “effector” is recognized, it reduces the fitness of the pathogen. Selective pressure may then result in the loss or alteration of such an effector or conversely the gain of novel effectors (as new virulence factors) through population evolution. Consequently, to avoid plant population extinction, new plant R gene loci must evolve to recognize these new pathogen effectors, in order to regain heritable resistance (Jones and Dangl, 2006). This ‘arms race’ type of co-evolution between pathogen strains that overcome ETI and new plant genotypes that restore ETI, is known as the zigzag model.
- c. In the guard model, an R protein does not interact directly with the Avr protein, but guards or monitors the status of another host protein, known as a guardee, that is the target of the AVR protein (initially proposed to explain the NBS-LRR protein Prf/Pto/AvrPto interaction) (Van der Biezen and Jones, 1998; Dangl and Jones, 2001). Such R proteins therefore recognize host changes brought about as a results of the action of the Avr protein on the target guardee. These pathogen proteins, although traditionally termed avirulence factors, genuinely act as virulence factors and are termed effectors (Bent and Mackey, 2007).
- d. In the decoy model, the R protein once again does not interact directly, but monitors modification of a host decoy protein which, although with no direct defense function itself in this model, mimics a component of plant defense, and as such attracts attention by the Avr protein (van der Hoorn et al., 2008). For example, a *Pseudomonas syringae* effector protein AvrPto infects *Arabidopsis* but causes a resistance response in tomato plants that carry the two resistance proteins Pto (which is a ser/thr kinase) and Prf (which is an NBS-LRR protein). Pto possesses a kinase domain similar to that in PRRs such as FLS2 and EFR (PRRs, as seen earlier, are normally involved in detection of invading pathogens in PTI, but in this specific case are targeted by AvrPto in susceptible plants, resulting in blocking of plant immune response). Pto has been hypothesized to have evolved as a decoy of the normal targets of AvrPto, whereby its interaction with AvrPto, which is monitored by Prf, then triggers resistance signaling responses (Zong et al., 2008).

Overlaps between PTI and ETI

Although PTI and ETI plant immunity branches are generally considered separate, boundaries are not clear (Thomma et al., 2011). For example, PAMPs have been characterized with narrow conservation across microbial classes, such as Pep-13, which is a fragment of a cell wall transglutaminase, conserved only among members of the genus *Phytophthora* and involved in plant defense activation (Brunner et al., 2002). Similarly, some pathogen effectors have been characterized with broad distribution, such as necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins, common across bacterial, oomycete and fungal pathogens (Gijzen and Nürnberger, 2006). Also contradicting current definitions, PAMPs may play roles in virulence. One clear example is flagellin, with *Pseudomonas syringae* flagellin mutants displaying reduced mobility and consequently

reduced virulence in plant hosts (Taguchi et al., 2010).

Different PAMPs, such as the N terminal 22 amino acid sequence in bacterial flagellin flg22, peptidoglycan and chitin, have been reported to induce large transcriptional responses in the plant host, with clear overlaps in the host gene sets showing differential expression at early stages following PAMP treatment (Gust et al., 2007; Wan et al., 2008), but not during late responses (Denoux et al., 2005). This suggests common early downstream signaling, triggered by different PAMPs, but different gene expression at later stages, perhaps reflecting different immune responses necessary for different pathogens. Induction of similar early transcriptome responses by diverse R protein - effector protein interactions has also been reported (Zipfel et al., 2006). Most interestingly, work with flg22 and pathogen effector proteins has also shown overlap in the transcriptional reprogramming induced by both plant immunity branches (Navarro et al., 2004), in line with the proposal that ETI evolved after PTI, with most of its mechanisms originating from pre-existing PTI mechanisms (Tsuda and Katagiri, 2010). Indeed, many downstream signaling components are seen to be shared between PTI and ETI, including an oxidative burst and changes in plant hormone levels. However, plant responses via ETI and PTI typically differ, with ETI response times generally faster than observed in PTI, with longer duration and with more robust effects in the host (Jones and Dangl, 2006; Tsuda and Katagiri, 2010).

Signaling Mechanisms

The hypersensitive response (HR), which is a form of programmed cell death, is a rapid response which restricts pathogen spread (Hammond-Kosack and Jones, 1997; Jones and Dangl, 2006). Although classically associated with ETI, PAMPs such as harpins and flagellins have also been reported to induce HR responses (e.g., Wei et al., 1992; Taguchi et al., 2003; Che et al., 2000), albeit by mediation via different signaling mechanisms. Cellular defense responses in plant immunity often initiate with an oxidative burst via the production of reactive oxygen species (ROS). These can function directly against pathogens or play roles in signaling (Torres et al., 2006). In PTI, ROS production is reported to be rapid and only transient (Zipfel et al., 2006) and dependent upon the NADPH oxidase ArRbohD (Zhang et al., 2007). In contrast, production appears biphasic in ETI, with an initial transient phase (possibly PAMP-induced), followed by a larger-scale and longer-lasting accumulation, with ArRbohD also involved in the second phase (Torres et al., 2006). Mitogen-activated protein kinase (MAPK) signaling cascades occur in both PTI and ETI, although variations in duration may be responsible for differential downstream responses in the two immunity branches (Tsuda and Katagiri, 2010). Systemic acquired resistance (SAR), which can be activated in uninfected parts of the plant, results in a physiologically 'primed' state, whereby an enhanced induction of longer-lasting defense responses to a broad range of biotic stresses is typical. The SAR response is dependent on expression of the regulatory protein NPR1 (Hammerschmidt, 2009). Although reported to enhance resistance when over-expressed in transgenics (e.g., Malnoy et al., 2007), NPR1 can also play negative regulatory functions in plant defense against biotic stresses (Quilis et al., 2008). SAR is dependent on the accumulation of the immunity phytohormone salicylic acid (SA), leading to an increase in acidic pathogenesis-related (PR) gene expression (Ward et al., 1991). Although SA is typically important in ETI against biotrophs and hemibiotrophs, SAR induction has also been reported in *Arabidopsis* with PAMPs such as flagellin and lipopolysaccharides, with typical increased expression of salicylic acid and pathogenesis-related (PR) proteins (Mishina and Zeier, 2007). Jasmonic acid (JA) and ethylene (ET) signalling pathways are also involved in plant defense, generally against necrotrophs, although with exceptions (Glazebrook, 2005). JA is involved in production of basic PR proteins and proteinase inhibitors (Niki et al., 1998). ET has also been reported to be involved in PR protein production (van Loon et al., 2006), synergistic effects with JA signaling (Lorenzo et al., 2004), and enhancing the SA-mediated NPR1 pathway in SAR (Leon-Reyes et al., 2009). Although the three pathways are considered potentially antagonistic (Koornneef and

Pieterse, 2008), evidence of synergy between different components of this network has also been reported (Tsuda et al., 2009). SA, JA and ET have all been reported in both PTI and ETI plant immunity branches, although they do not appear to be always activated. Although signaling mechanisms appear common to PTI and ETI, how they are employed in each branch likely differ (Tsuda and Katagiri, 2010). Definitions of PTI and ETI, however, on the basis of expected signaling and defense responses, especially in terms of scale are not well defined, as responses can be atypical, with limited intensities possible in ETI, depending on effector-R-protein interaction, and robust PTI responses, depending upon the specific host-pathogen interaction. Understanding of all the molecular mechanisms controlling signaling and SAR is far from resolved, with suggestions of involvement of further signaling proteins, post-translational modification, altered expression in priming responsive genes, and upregulation of transcription factors (Van Der Ent et al., 2009).

NEXT-GENERATION SEQUENCING (NGS) TECHNOLOGIES

Since 2005, a number of new technologies have been developed for improved DNA sequencing, which are collectively referred to as next-generation sequencing (NGS) or ‘deep sequencing’ technologies. The principal advantages offered by these approaches lie both in the scale of low-cost sequence data that can be generated, together with a rate which is several orders of magnitude faster than traditional sequencing approaches based upon Sanger dideoxy sequencing (Sanger et al., 1977; Prober et al., 1987). Of the numerous technologies which have been developed, three are currently widely deployed, namely 454 (454 Life Sciences/Roche) (<http://www.454.com/>), Solexa/Illumina (<http://www.Illumina.com>), and AB SOLiD (<http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing.html>). Currently 454 technology offers advantages in terms of sequence read length, whilst Illumina and AB SOLiD offer deeper coverage and lower costs. These technologies offer opportunities to increase our understanding of cell function at the nucleic acid level (Lister et al., 2009). In the case of plant genetic improvement, applications include sequencing genomic DNA of interest, such as: bacterial artificial chromosomes (BACs), reduced representation of a genome, whole genome sequencing, genome re-sequencing, marker development and QTL mapping (SSRs and SNPs), association mapping, mutation mapping, genotyping, cytosine DNA methylation analysis, gene discovery, identification of alternative splicing and small RNA profiling. The transcriptome is defined as the total sum of a cell’s transcripts (including mRNAs, non-coding RNAs and small RNAs), together with gene expression information during particular physiological conditions or developmental stages (Wang et al., 2009). The transcriptome also provides information regarding gene structure and alternative splicing. Whilst experiments based upon microarray technology have served to initiate identification of massive transcriptional changes during plant immunity responses (Tao et al., 2003), NGS technologies offer greater potential in such transcription profiling. These technologies are now superseding microarray analysis and serial analysis of gene expression (SAGE), with rapidly generated deep coverage sequencing data, with unbiased transcript representation, appropriate for gene discovery, including de novo detection of rare transcripts and their expression, as well as alternative splicing and gene sequence polymorphism detection (Wold and Myers, 2008; Wang et al., 2009). Such potential seen in NGS technologies make them well placed for applications in genomic analysis of host-pathogen interactions, as an approach for furthering our understanding of the molecular mechanisms controlling plant immunity. Some recent examples of the application of NGS in transcriptome analysis of plant host-pathogen interactions are given in Table 1.

Ongoing 454 Transcriptome Sequencing in *Musa*

Although banana (*Musa* sp.) is one of the world’s most important edible crops, contributing towards poverty alleviation and food security, a comprehensive transcriptomic dataset is not yet available for use in accelerated molecular-based breeding

in this genus. In order to develop a functional genomics resource for this crop which reveals transcriptional changes during plant immunity responses to biotic stresses, a collaborative functional genomics project is underway between UNB, UCB, Embrapa, and CICY. In this project, we have performed a pyrosequencing study of expressed genes in *Musa acuminata* during compatible and incompatible reactions with the fungal pathogen *Mycosphaerella musicola*, causal organism of Sigatoka leaf spot disease. Pooled total RNA samples were prepared from whole plant leaf material from 6-month-old *M. acuminata* ssp. *burmannicoides* 'Calcutta 4' (resistant) and Cavendish 'Grande Naine' (susceptible) whole plants (*Musa* International Transit Centre accessions ITC0249 and ITC0654), both uninfected (non-inoculated controls) and artificially challenged with pathogen conidiospores. Synthesis of four full-length enriched cDNA libraries and pyrosequencing was carried out by Eurofins MWG Operon (Ebersberg, Germany), using approximately 50 µg of each total RNA pool. Non-cloned libraries were prepared by random priming of polyA+ mRNA and size fractioning. Emulsion PCR and sequencing were conducted according to Roche standard protocols using GS FLX technology and Titanium series chemistry, generating 978,133 raw sequencing reads, with an average length of 334 bp and totalling over 460 million bp. Figure 2 summarizes the in-house bioinformatic analysis applied for sequence processing and assembly. For unigene classification, each unigene was aligned against the NCBI nr protein database using BLASTX, using the criteria were that the alignment length should be >100 amino acids and the expected value (E) <1E10⁻⁵. In order to verify the presence of pathogen transcripts amongst the unigene sets, each unigene was also aligned using BLASTN (megablast option, BLASTN criteria of alignment bitscore >200) against the following databases: *Mycosphaerella fijiensis* v2.0 All Gene Models (transcripts); *Mycosphaerella graminicola* v2.0 All Gene Models (transcripts); and Coding Sequence data from Fungal Genome Research (<http://www.fungalgenome.org>). Gene ontology (GO) terms were selected from GOOSE (<http://go.berkeleybop.org/cgi-bin/goose/goose>) using a database mirror in Berkeley, CA, USA. In silico differential expression analysis was conducted with IDEG6, and general database management using PostgreSQL and Chado.

In excess of 35,000 unigenes were assembled for each *M. acuminata* genotype, with approximately 35% displaying no significant similarity to any sequences in the public databases. Out of a total of 13,513 such assembled sequences for 'Calcutta 4' without matches to the database, interproscan analysis of translated sequences identified functional protein domains in 192 sequences. Similar findings were also observed in 'Grande Naine', with 11,971 sequences with no significant similarity to database sequences, of which 223 contained characterized interpro domains. Over 1900 unigenes were identified in silico as potentially involved in defense and biotic stress responses. In silico expression analysis identified numerous differentially expressed genes associated with resistance to stresses and response to biotic stimuli in both cultivars, based on comparison with non-inoculated controls. These included genes involved in MAP kinase cascades, phenylpropanoid production, flavonoid pathway, plant detoxification, phytohormone signaling, PR proteins and phytoalexin production. EST datasets were also exploited for identification of over 30 expressed NBS-LRR resistance gene analogs.

In addition to gene discovery and expression profiling, the unigene sets were exploited for large-scale simple sequence repeat (SSR) genetic marker development. 'Calcutta 4' and 'Grande Naine' unigene datasets were analyzed for the presence of SSRs using the program mreps (<http://bioinfo.lifl.fr/mreps/>). Flanking primer pairs were designed using Primer3 (Rozen and Skaletsky, 2000). Incidence of SSRs in the consensi to which primers could be designed was approximately 6% for each cultivar, with over 4,000 SSR loci in each cultivar dataset potentially amplifiable via PCR and applicable for genetic marker development. An example of polymorphism obtained with primers for one locus is shown in Figure 3.

CONCLUSIONS AND PROSPECTS

Although our current understanding of the mechanisms involved in plant

immunity has increased considerably with recent advances in molecular biology and genomics technologies, the development of novel and durable disease resistance specificities will depend upon a more complete understanding of plant immunity. The relationship between PTI and ETI, for example, remains blurred, and the role of downstream components of plant defense requires greater elucidation. Additionally, characterization of plant *R* genes and increased understanding of *R* gene motif specificities and their interaction with cognate pathogen effector protein domains is necessary. This would likely then open up opportunities for engineering pathogen recognition, such as via LRR motif reorganization (Lucas, 2010).

High-throughput next generation sequencing of the genomes and transcriptomes of crop species and their wild relatives is undoubtedly increasing our knowledge of the gene pool and diversity of genetic resources for breeding, and facilitating identification of genes and mechanisms involved in plant immunity. Functional analysis of genes is, however, a considerable bottleneck, given the exponential rise in available genome data. Large-scale application of plant transformation approaches and gene silencing strategies, such as RNA interference (RNAi) and virus-induced gene silencing (VIGS), is therefore necessary (Scofield and Nelson, 2009). NGS data is also applicable for large-scale genetic marker development, facilitating genetic map saturation and marker-assisted breeding. As a consequence, genomics-assisted breeding is predicted to continue to gain momentum (Varshney et al., 2009), accelerating both phenotype prediction from genotype as well as genetic improvement via plant transformation.

The unigene datasets developed by our group, exploited for gene identification, transcription profiling, as well as large scale gene-based SSR marker development, will contribute, together with future functional analysis and genetic map enrichment approaches, to both our understanding of genes involved in plant immunity processes in *Musa* and long-term strategies for disease management based upon genetic improvement.

ACKNOWLEDGEMENTS

This work was funded by the IAEA (Project 13187), FINEP (Project 0107060900 / 0842/07) and CNPq / FAPDF / PRONEX (Project 193.000.5622009).

Literature Cited

- Altenbach, D. and Robatzek, S. 2007. Pattern recognition receptors: from the cell surface to intracellular dynamics. *Molecular Plant-Microbe Interactions* 20:1031-1039.
- Baker, B., Zambryski, P., Staskawicz, B. and Dinesh-Kumar, S.P. 1997. Signaling in plant-microbe interactions. *Science* 276:726-733.
- Barakat, A., DiLoreto, D.S., Zhang, Y., Smith, C., Baier, K., Powell, W.A., Wheeler, N., Sederoff, R. and Carlson, J.E. 2009. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biology* 9:11.
- Bent, A.F. and Mackey, D. 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45:399-436.
- Block, A., Li, G., Fu, Z.Q. and Alfano, J.R. 2008. Phytopathogen type III effector weaponry and their plant targets. *Curr. Opin. Plant Biol.* 11:396-403.
- Boller, T. and 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.
- Brunner, F., Rosahl, S., Lee, J., Rudd, J.J., Geiler, C., Kauppinen, S., Rasmussen, G., Scheel, D. and Nürnberger, T. 2002. Pep-13, a plant defense-inducing pathogen-associated pattern from *Phytophthora* transglutaminases. *EMBO J.* 21:6681-6688.
- Che, F.S., Nakajima, Y., Tanaka, N., Iwano, M., Yoshida, T., Takayama, S., Kadota, I. and Isogai, A. 2000. Flagellin from an incompatible strain of *Pseudomonas avenae* induces a resistance response in cultured rice cells. *J. Biol. Chem.* 275:32347-32356.
- Chisholm, S.T., Coaker, G., Day, B. and Staskawicz, B.J. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124:803-814.

- da Cunha, L., Sreerekha, M.V. and Mackey, D. 2007. Defense suppression by virulence effectors of bacterial phytopathogens. *Curr. Opin. Plant Biol.* 10:349-357.
- Dangl, J.L. and Jones, J.D. 2001. Plant pathogens and integrated defence responses to infection. *Nature* 411:826-33.
- Delano-Frier, J.P., Aviles-Arnaut, H., Casarrubias-Castillo, K., Casique-Arroyo, G., Castrillon-Arbelaez, P.A., Herrera-Estrella, L. et al. 2011. Transcriptomic analysis of grain amaranth (*Amaranthus hypochondriacus*) using 454 pyrosequencing: comparison with *A. tuberculatus*, expression profiling in stems and in response to biotic and abiotic stress. *BMC Genomics* 12:363.
- Denoux, C., Galletti, R., Mammarella, N., Gopalan, S., Werck, D., De Lorenzo, G. et al. 2008. Activation of defense response pathways by OGs and Flg22 elicitors in *Arabidopsis* seedlings. *Mol. Plant* 1:423-445.
- Fernandez, D., Tisserant, E., Talhinhos, P., Azinheira, H., Vieira, A., Petitot, A.S. et al. 2011. 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Mol. Plant Pathol.* doi: 10.1111/j.1364-3703.2011.00723.x. Epub 2011 Jun. 1.
- Flor, H.H. 1942. Inheritance of pathogenicity in *Melampsora lini*. *Phytopath.* 32:653-669.
- Gijzen, M. and Nürnberger, T. 2006. Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry* 67:1800-1807.
- Gohre, V. and Robatzek, S. 2008. Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu. Rev. Phytopathol.* 46:189-215.
- Glazebrook, J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 43:205-227.
- Gust, A.A., Biswas, R., Lenz, H.D., Rauhut, T., Ranf, S., Kemmerling, B. et al. 2007. Bacteria-derived peptidoglycans constitute pathogen-associated molecular patterns triggering innate immunity in *Arabidopsis*. *J. Biol. Chem.* 82:32338-32348.
- Hammerschmidt, R. 2009. Systemic acquired resistance. *Advances in Botanical Research* 51:173-222.
- Hammond-Kosack, K.E. and Jones, J.D.G. 1997. Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology* 48:573-605.
- Hsiao, Y.Y., Chen, Y.W., Huang, S.C., Pan, Z.J., Fu, C.H., Chen, W.H., Tsai, W.C. et al. 2011. Gene discovery using next-generation pyrosequencing to develop ESTs for *Phalaenopsis* orchids. *BMC Genomics* 12:360.
- Jones, J.D.G. and Dangl, J.L. 2006. The plant immune system. *Nature* 444:323-329.
- Keen, N.T. 1990. Gene-for-gene complementarity in plant-pathogen interactions. *Annu. Rev. Genet.* 24:447-463.
- Koornneef, A. and Pieterse, C.M.J. 2008. Cross talk in defense signaling. *Plant Physiology* 146:839-844.
- Leon-Reyes, A., Spoel, S.H., De Lange, E.S., Abe, H., Kobayashi, M., Tsuda, S. et al. 2009. Ethylene modulates the role of nonexpressor of pathogenesis-related gene1 in cross talk between salicylate and jasmonate signaling. *Plant Physiology* 149:1797-1809.
- Lister, R., Gregory, B.D. and Ecker, J.R. 2009. Next is now: new technologies for sequencing of genomes, transcriptomes and beyond. *Curr. Opin. Plant Biol.* 12:107-118.
- Lorenzo, O., Piqueras, R., Sánchez-Serrano, J.J. and Solano, R. 2003. Ethylene response factor1 integrates signaling from ethylene and jasmonate pathways in plant defense. *Plant Cell* 15:165-178.
- Lucas, J.A. 2010. Advances in plant disease and pest management. *The Journal of Agricultural Science* 149:91-114.
- Malnoid, M., Jin, Q., Borejsza-Wysocka, E.E., He, S.Y. and Aldwinckle, H.S. 2007. Overexpression of the apple *MpNPR1* gene confers increased disease resistance in *Malus domestica*. *Molecular Plant-Microbe Interactions* 20:1568-1580.

- McHale, L., Tan, X., Koehl, P. and Michelmore, R.W. 2006. Plant NBS-LRR proteins: adaptable guards. *Genome Biology* 7:212-1-212-11.
- Medzhitov, R. and Janeway, C.A., Jr. 1997. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91:295-298.
- Miller, R.N.G., Bertioli, D.J., Baurens, F.C., Quirino, B.F., Ciampi, A.Y., Santos, C.M.R. et al. 2009. Understanding plant responses to biotic stress: ongoing research in *Musa*. *Acta Hort.* 828:255-272.
- Mishina, T.E. and Zeier, J. 2007. Pathogen-associated molecular pattern recognition rather than development of tissue necrosis contributes to bacterial induction of systemic acquired resistance in *Arabidopsis*. *Plant J.* 50:500-513.
- Narina, S.S., Buyyarapu, R., Kottapalli, K.R., Sartie, A.M., Ali, M.I., Robert, A. et al. 2011. Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (*Dioscorea alata* L.). *BMC Genomics* 12:100.
- Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T. and Jones, J.D. 2004. The transcriptional innate immune response to flg22, Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. *Plant Physiol.* 135:1113-1128.
- Niki, T., Mitsuhashi, I., Seo, S., Ohtsubo, N. and Ohashi, Y. 1998. Antagonistic effect of salicylic acid and jasmonic acid on the expression of pathogenesis-related (PR) protein genes in wounded mature tobacco leaves. *Plant and Cell Physiology* 39:500-507.
- Nürnberg, T. and Kemmerling, B. 2009. PAMP-triggered basal immunity in plants. *Advances in Botanical Research* 51:1-38.
- Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J. et al. 1987. A system for rapid DNA sequencing with fluorescent chain terminating dideoxynucleotides. *Science* 238:336-341.
- Quilis, J., Peñas, G., Messeguer, J., Brugidou, C. and San Segundo, B. 2008. The *Arabidopsis* AtNPR1 inversely modulates defense responses against fungal, bacterial, or viral pathogens while conferring hypersensitivity to abiotic stresses in transgenic rice. *Molecular Plant-Microbe Interactions* 21:1215-1231.
- Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* 132:365-386.
- Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Scofield, S.R. and Nelson, R.S. 2009. Resources for virus-induced gene silencing in the grasses. *Plant Physiology* 149:152-158.
- Shiu, S.H. and Blecker, 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.
- Taguchi, F., Yamamoto, M., Ohnishi-Kameyama, M., Iwaki, M., Yoshida, M., Ishii et al. 2010. Defects in flagellin glycosylation affect the virulence of *Pseudomonas syringae* pv. *tabaci* 6605. *Microbiology* 156:72-80.
- Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.S., Han, B., Zhu, T., Zou, G. and Katagiri, F. 2003. Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell* 15:317-330.
- Thomma, B.P.H.J., Nürnberg, T. and Joosten, M.H.A.J. 2011. Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *The Plant Cell* 23:4-15.
- Torres, M.A., Jones, J.D. and Dangl, J.L. 2006. Reactive oxygen species signaling in response to pathogens. *Plant Physiol.* 141:373-378.
- Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J. and Katagiri, F. 2009. Network properties of robust immunity in plants. *PLoS Genetics* 5:12 e1000772.
- Tsuda, K. and Katagiri, F. 2010. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr. Opin. Plant Biol.* 13:459-465.
- Van der Biezen, E. and Jones, J.G. 1998. Plant disease-resistance proteins and the gene-

- for-gene concept. *Trends in Biochemical Sciences* 23:454-456.
- Van Der Ent, S., Van Hulten, M., Pozo, M.J., Czechowski, T., Udvardi, M.K., Pieterse, C.M.J. and Ton, J. 2009. Priming of plant innate immunity by rhizobacteria and beta-aminobutyric acid: differences and similarities in regulation. *New Phytologist* 183:419-431.
- van der Hoorn, R.A. and Kamoun, S. 2008. From guard to decoy: a new model for perception of plant pathogen effectors. *Plant Cell* 20:2009-2017.
- van Loon, L.C., Rep, M. and Pieterse, C.M.J. 2006. Significance of inducible defense-related proteins in infected plants. *Annual Review of Phytopathology* 44:135-162.
- Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotech.* 27(9):522-530.
- Venu, R.C., Zhang, Y., Weaver, B., Carswell, P., Mitchell, T.K. and Meyers, B.C. 2011. Large scale identification of genes in plant-fungal interactions sequencing-by-synthesis technology. *Methods in Molecular Biology* 722:167-178.
- Vorwerk, S., Somerville, S. and Somerville, C. 2004. The role of plant cell wall polysaccharide composition in disease resistance. *Trends Plant Sci.* 9:203-209.
- Wan, J., Zhang, X.C., Neece, D., Ramonell, K.M., Clough, S., Kim, S.Y., Stacey, M.G. and Stacey, G. 2008. A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in *Arabidopsis*. *Plant Cell* 20:471-481.
- Wang, Z., Gerstein, M. and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* 10:57-63.
- Ward, E.R., Uknes, S.J., Williams, S.C., Dincher, S.S., Wiederhold, D.L., Alexander, D.C. et al. 1991. Coordinate gene activity in response to agents that induce systemic acquired resistance. *Plant Cell* 3:1085-1094.
- Wei, Z.M., Laby, R.J., Zumoff, C.H., Bauer, D.W., He, S.Y., Collmer, A. and Beer, S.V. 1992. Harpin, elicitor of the hypersensitive response produced by the plant pathogen *Erwinia amylovora*. *Science* 257:85-88.
- Wold, B. and Myers, R.M. 2008. Sequence census methods for functional genomics. *Nature Methods* 5:19-21.
- Wu, J., Zhang, Y., Zhang, H., Huang, H., Folta, K. and Lu, J. 2010. Whole genome wide expression profiles of *Vitis amurensis* grape responding to downy mildew by using Solexa sequencing technology. *BMC Plant Biology* 10:234.
- Xiao, S. 2006 Current perspectives on molecular mechanisms of plant disease resistance, p.317-333 in *Floriculture, Ornamental and Plant Biotechnology: Advances and Topical Issues, Vol. III*, edited by J. Teixeira da Silva. Global Science Books, London.
- Xiao, S., Wang, W. and Yang, X. 2008. Evolution of resistance genes in plants. *Innate immunity of plants, animals, and humans. Nucleic Acids and Molecular Biology* 21:1-25.
- Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T. and Felix, G. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell* 125:749-760.
- Zhang, J., Shao, F., Li, Y., Cui, H., Chen, L., Li, H. et al. 2007. A *Pseudomonas syringae* effector inactivates MAPKs to suppress PAMP-induced immunity in plants. *Cell Host Microbe* 1:175-185.
- Zong, N., Xiang, T., Zou, Y., Chai, J. and Zhou, J.-M. 2008. Blocking and triggering of plant immunity by *Pseudomonas syringae* effector AvrPto. *Plant Signal Behav.* 3:583-585.

Tables

Table 1. NGS applications in understanding plant immunity mechanisms.

Plant-pathogen interaction	NGS technology	Applications	References
Interaction between <i>Dioscorea alata</i> genotypes and <i>Colletotrichum gloeosporioides</i>	454 transcriptome analysis	Transcriptome analysis of young leaf material in resistant and susceptible genotypes during interaction with <i>Colletotrichum gloeosporioides</i> ; 44,757 unique ESTs; multiple pathogenesis-related genes and host defense related genes	Narina et al., 2011
Interaction between <i>Phalaenopsis equestris</i> and <i>Erwinia chrysanthemi</i>	454 transcriptome analysis	Transcriptome analysis of pooled RNA from distinct tissues, biotic stress (leaf tissues during interaction with bacterial pathogen <i>Erwinia chrysanthemi</i>) and abiotic stress conditions across three species of the orchid <i>Phalaenopsis</i> ; 22,234 unigenes identified	Hsiao et al., 2011
Interaction between <i>Amaranthus hypochondriacus</i> and <i>Pseudomonas argentinensis</i>	454 transcriptome analysis	Transcriptome analysis of <i>A. hypochondriacus</i> plant leaf tissues during interaction with bacterial amaranth pathogen <i>Pseudomonas argentinensis</i> ; 20,408 isotigs and 799 contigs assembled; 1% of transcripts with defense role assigned (41% associated with bacterial infection, 24% with jasmonic acid (JA)-regulation)	Délano-Frier et al., 2011
Interaction between <i>Coffea arabica</i> and <i>Hemileia vastatrix</i>	454 transcriptome analysis	Transcriptomes from <i>C. arabica</i> leaf tissues during compatible interaction with rust fungal pathogen <i>H. vastatrix</i> ; 22,774 unigenes identified; 30% of contigs assigned to <i>H. vastatrix</i> , 61% to <i>C. arabica</i> ; in silico prediction of pathogen effector homologues of flax rust haustorially expressed secreted proteins (HESPs) and bean rust transferred protein 1 (RTP1); Q-RTPCR profiling of rust genes during infection time course	Fernandez et al., 2011
Interaction between <i>Castanea dentata</i> and <i>Castanea mollissima</i> with <i>Cryphonectria parasitica</i>	454 transcriptome analysis	Transcriptomes from stem tissues during interaction with chestnut blight pathogen <i>C. parasitica</i> , 28,890 unigenes generated in <i>C. dentata</i> (susceptible), 40,039 unigenes in <i>C. mollissima</i> (resistant); identification of differentially expressed genes involved in resistance and defense mechanisms such as hypersensitivity, jasmonic acid pathway, lignin biosynthesis, phenylpropanoid pathway, signaling, and resistance gene expression regulation	Barakat et al., 2009
Interaction between <i>Vitis amurensis</i> and <i>Plasmopara viticola</i>	Illumina/Solexa transcriptome analysis	Transcriptomes from leaf tissues during interaction with downy mildew pathogen <i>P. viticola</i> ; 14549 unigenes identified in uninoculated control, 15249 in infection library; identification of differentially expressed stress-related genes	Wu et al., 2010
Interaction between creeping bentgrass and <i>Sclerotinia homoeocarpa</i>	Illumina/Solexa transcriptome analysis	Transcriptome data from interaction between creeping bentgrass and <i>S. homoeocarpa</i> ; adapted methodology appropriate for interaction cDNA library construction and data analysis	Venu et al., 2011

Figures

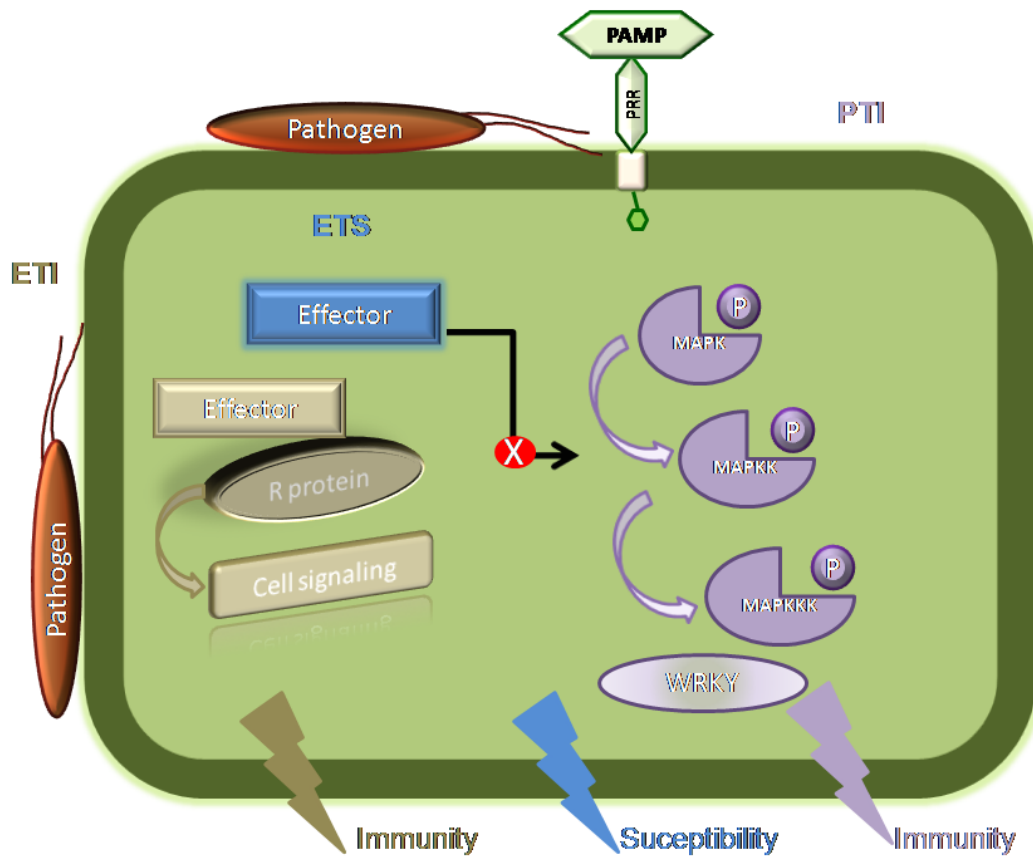


Fig. 1. Schematic representation of plant immune system activation by pathogen: PAMP-Triggered Immunity (PTI) and Effector-Triggered Immunity (ETI). PTI, depicted in lilac, follows pathogen-associated molecular pattern (PAMP) recognition by host pattern recognition receptors (PRRs), and results in activation of a MAP kinase cascade and WRKY transcription factors. The result is a blocking of further host colonization and provision of resistance to the majority of potential pathogens. Evolution of pathogen virulence effectors, depicted in blue, results in suppression of PTI signaling and a breakdown of resistance, known as Effector-Triggered Susceptibility (ETS). Co-evolving host resistance R proteins can then recognize pathogen effectors and activate Effector-Triggered Immunity (ETI), depicted in gold.

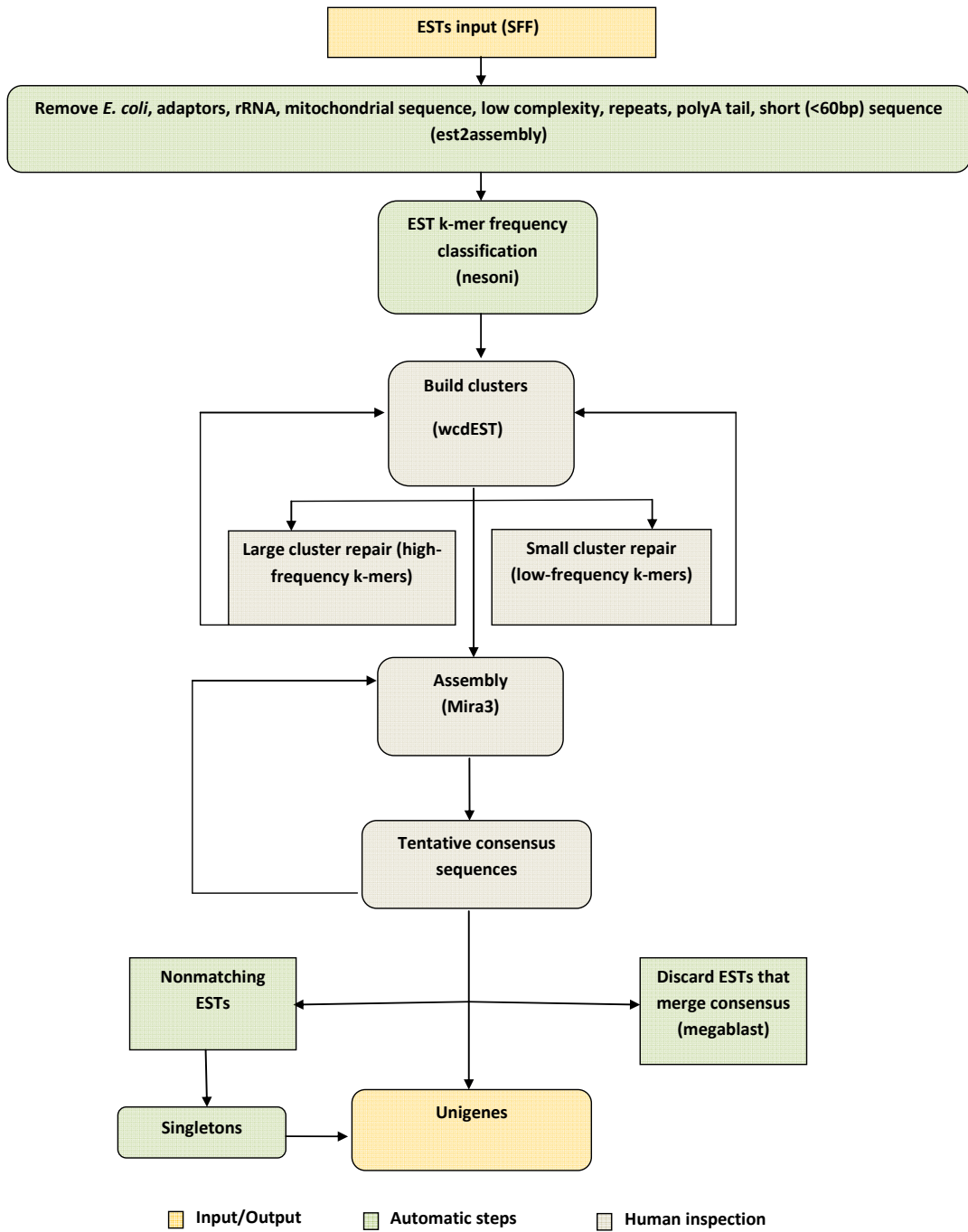


Fig. 2. Schematic representation of the bioinformatic analysis methods: 454 transcriptome sequence processing and assembly.

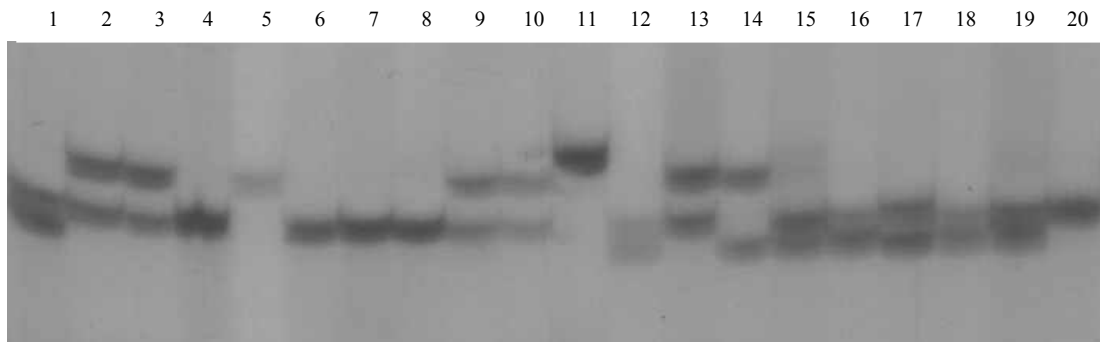


Fig. 3. Silver nitrate-stained 4% polyacrylamide gel showing amplification profiles using primers for SSR locus 2028. Polymorphism was tested across 22 *M. acuminata* diploid accessions. 1: 1741-01; 2: SH32-63; 3: 4279-06; 4: 1318-01; 5: 0323-03; 6: 1304-06; 7: 0116-01; 8: 9179-03; 9: Burmanica; 10: Calcutta 4; 11: Microcarpa; 12: Pisang Berlin; 13: Lidi; 14: Khai Nai On; 15: Niyarma Yik; 16: Jaribuaya; 17: Raja Uter; 18: Sowmuk; 19: Tjau Lagada; 20: F2P2.



Open access – Research article

THIS ARTICLE IS PART OF A SPECIAL ISSUE ENTITLED
'MOLECULAR TECHNOLOGIES TO IMPROVE
VEGETATIVELY PROPAGATED BANANA AND CASSAVA'

Development of expressed sequence tag and expressed sequence tag–simple sequence repeat marker resources for *Musa acuminata*

Marco A. N. Passos¹, Viviane de Oliveira Cruz¹, Flavia L. Emediato¹, Cristiane de Camargo Teixeira², Manoel T. Souza Jr³, Takashi Matsumoto⁴, Vânia C. Rennó Azevedo³, Claudia F. Ferreira⁵, Edson P. Amorim⁵, Lucio Flavio de Alencar Figueiredo¹, Natalia F. Martins³, Maria de Jesus Barbosa Cavalcante⁶, Franc-Christophe Baurens⁷, Orzenil Bonfim da Silva Jr³, Georgios J. Pappas Jr^{1,3}, Luc Pignolet⁸, Catherine Abadie⁸, Ana Y. Ciampi³, Pietro Piffanelli^{7,9} and Robert N. G. Miller^{1*}

¹ Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Asa Norte, CEP 70910-900, Brasília, DF, Brazil

² Postgraduate Program in Genomic Science and Biotechnology, Universidade Católica de Brasília, SGAN 916, Módulo B, CEP 70.790-160, Brasília, DF, Brazil

³ EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, DF, Brazil

⁴ National Institute of Agrobiological Resources, Tsukuba 305-8602, Japan

⁵ EMBRAPA Mandioca e Fruticultura Tropical, Rua Embrapa, CEP 44380-000, Cruz das Almas, BA, Brazil

⁶ EMBRAPA Acre, CP 321 BR 364 Km 14, Rio Branco, AC, Brazil

⁷ CIRAD/UMR DAP 1098, TA A 96/03 Avenue Agropolis, 34098 Montpellier Cedex 5, France

⁸ CIRAD/UMR BGPI, TA A 54/K Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

⁹ Present address: Genomics Platform at Parco Tecnologico Padano, Via Einstein, Località Cascina Codazza, 26900 Lodi, Italy

Received: 4 March 2012; **Returned for revision:** 13 August 2012; **Accepted:** 14 September 2012; **Published:** 26 September 2012

Citation details: Passos MAN, de Oliveira Cruz V, Emediato FL, de Camargo Teixeira C, Souza MT Jr, Matsumoto T, Rennó Azevedo VC, Ferreira CF, Amorim EP, de Alencar Figueiredo LF, Martins NF, de Jesus Barbosa Cavalcante M, Baurens F-C, da Silva OB Jr, Pappas GJ Jr, Pignolet L, Abadie C, Ciampi AY, Piffanelli P, Miller RNG. 2012. Development of expressed sequence tag and expressed sequence tag–simple sequence repeat marker resources for *Musa acuminata*. *AoB PLANTS* 2012: pls030; doi:10.1093/aobpla/pls030

Abstract

Background and aims

Banana (*Musa acuminata*) is a crop contributing to global food security. Many varieties lack resistance to biotic stresses, due to sterility and narrow genetic background. The objective of this study was to develop an expressed sequence tag (EST) database of transcripts expressed during compatible and incompatible banana–*Mycosphaerella fijiensis* (*Mf*) interactions. Black leaf streak disease (BLSD), caused by *Mf*, is a destructive disease of banana. Microsatellite markers were developed as a resource for crop improvement.

Methodology

cDNA libraries were constructed from *in vitro*-infected leaves from BLSD-resistant *M. acuminata* ssp. *burmanicoides* Calcutta 4 (MAC4) and susceptible *M. acuminata* cv. Cavendish Grande

* Corresponding author's e-mail address: robertmiller@unb.br

Published by Oxford University Press on behalf of the Annals of Botany Company. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Naine (MACV). Clones were 5'-end Sanger sequenced, ESTs assembled with TGICL and unigenes annotated using BLAST, Blast2GO and InterProScan. Mreps was used to screen for simple sequence repeats (SSRs), with markers evaluated for polymorphism using 20 diploid (AA) *M. acuminata* accessions contrasting in resistance to *Mycosphaerella* leaf spot diseases.

Principal results

A total of 9333 high-quality ESTs were obtained for MAC4 and 3964 for MACV, which assembled into 3995 unigenes. Of these, 2592 displayed homology to genes encoding proteins with known or putative function, and 266 to genes encoding proteins with unknown function. Gene ontology (GO) classification identified 543 GO terms, 2300 unigenes were assigned to EuKaryotic orthologous group categories and 312 mapped to Kyoto Encyclopedia of Genes and Genomes pathways. A total of 624 SSR loci were identified, with trinucleotide repeat motifs the most abundant in MAC4 (54.1 %) and MACV (57.6 %). Polymorphism across *M. acuminata* accessions was observed with 75 markers. Alleles per polymorphic locus ranged from 2 to 8, totalling 289. The polymorphism information content ranged from 0.08 to 0.81.

Conclusions

This EST collection offers a resource for studying functional genes, including transcripts expressed in banana–*Mf* interactions. Markers are applicable for genetic mapping, diversity characterization and marker-assisted breeding.

Introduction

Commercially cultivated varieties of banana and plantains are derived from the progenitors *Musa acuminata* Colla (AA) and *Musa balbisiana* Colla (BB). These crops are of extreme importance across the world's tropical and sub-tropical regions, contributing to both food security and export commodity revenue, with a global annual production in excess of 97 million tonnes (FAOSTAT 2009).

Cultivated bananas have evolved from hybridization of wild species of *M. acuminata* (A genome) and *M. balbisiana* (B genome). In contrast to fertility in wild species, many of today's commercial cultivars are sterile triploids or diploids, with fruit development via parthenocarpy. Together with female sterility, this results in either seedless fruits or non-viable seeds. Consequential asexually driven evolution has resulted in a narrow genetic base, with the crop often lacking resistance to pests and diseases. For this reason, the industry has witnessed numerous pathogen and pest outbreaks. Of the >40 fungal diseases affecting banana (Jones 1999), the foliar pathogen *Mycosphaerella fijiensis* (*Mf*) is today one of the most threatening. Responsible for black leaf streak disease (BLS) in banana, commonly known as black Sigatoka, yield losses range from 20 to 80 % (Churchill 2011), with premature fruit ripening also affecting export markets. Although cultural practices contribute to disease control, without the integrated use of chemicals their impact is insufficient. Commercial banana plantations are therefore dependent upon long-term use of agrochemicals, which

implies a constant threat for the emergence of fungicide-tolerant or -resistant *Mf* strains. The development of disease-resistant genotypes is today therefore regarded as the most cost-effective long-term control strategy available for the *Musa* industry.

Current breeding strategies for *Musa* rely upon sexually active wild or improved fertile *M. acuminata* diploids, which, in contrast to most commercial *Musa* varieties, where genetic diversity is fixed by vegetative propagation, serve as sources of resistance to biotic and abiotic stresses for transfer across varieties. Programmes for the development of tetraploid hybrids, for example, are typically generated via crosses between semi-fertile established triploids and wild or improved fertile diploid parents with agronomic traits of interest (Ortiz 1997; Amorim et al. 2011). Such breeding strategies can, however, have only limited success, given low numbers or absence of seeds. Complementary strategies for resolving these constraints for perennial crop breeding are therefore required.

Isolation of candidate genes of agronomic interest and development of specific molecular markers for application in molecular genotyping and marker-assisted selection (MAS) allow for both accelerated conventional breeding and gene-transfer programmes as strategies for genetic improvement. Expressed sequence tags (ESTs) are 5'- or 3'-end single-pass-sequenced portions of randomly isolated cDNA clones, which as such represent part of the transcribed region of the genome in given conditions. As a rapid approach for gene discovery and analysis of gene expression and regulation, data can also be exploited for the development of functional genetic markers. For *Musa*, a total of only 15 464 ESTs

in *M. acuminata* and 5289 in *M. balbisiana* are currently publically available in GenBank (accessed March 2012). These datasets have been generated from a number of cultivars, plant tissues (Roux et al. 2008), during abiotic stress responses (Santos et al. 2005) and post-harvest ripening (Manrique-Trujillo et al. 2007). Only limited analysis of gene expression in response to biotic stresses has been reported (e.g. Van den Berg et al. 2004; Portal et al. 2011).

Highly variable microsatellites or simple sequence repeats (SSRs) are abundant in eukaryotic genomes, and may occur in both coding and non-coding regions (e.g. Tamana and Khan 2005). Typically they are reproducible, somatically stable, highly polymorphic, co-dominant, multi-allelic markers, with application in population genetics, genetic mapping and molecular breeding. Locus-by-locus *de novo* development is costly and time consuming, in contrast to mining from EST sequence databases. As EST-SSR markers originate from transcribed genes, they offer potential for analysis of functional diversity in populations and application in MAS, through utilization of markers that either originate from a gene responsible for a desirable phenotypic trait, or that co-localize with a particular quantitative trait locus (QTL) (Varshney et al. 2005). Applications of SSR markers in *Musa* have focused on evolution and taxonomy (e.g. Lagoda et al. 1998), genotyping (e.g. Crouch et al. 2001; Crete et al. 2003; Christelová et al. 2011), and, more recently, linkage map saturation (e.g. Hippolyte et al. 2010). In comparison with other important crops, however, still relatively few SSR markers have been developed for *M. acuminata* and *M. balbisiana* material (e.g. Kaemmer et al. 1997; Lagoda et al. 1998; Crouch et al. 2001; Buhariwalla et al. 2005; Crete et al. 2006; Cheung and Town 2007; Miller et al. 2010). Considering that alleles can be monomorphic or even absent when applied across cultivars, the number of useful SSR loci available remains limited.

This work describes the generation of an EST resource for *M. acuminata* and its mining for gene-derived SSR markers. The annotated ESTs were generated from two cDNA libraries constructed from BLSD-resistant *M. acuminata* ssp. *burmannicoides* var. Calcutta 4 (MAC4) and BLSD-susceptible *M. acuminata* subgroup Cavendish cv. Grande Naine (MACV) leaves *in vitro* infected with *Mf*. The wild diploid cultivar Calcutta 4 is widely employed in breeding programmes as a source of resistance to fungal pathogens and nematodes. It has also been used as a model for comparative genomics (Cheung and Town 2007; Lescot et al. 2008), functional genomics (e.g. Santos et al. 2005) and candidate resistance gene discovery (e.g. Azhar and Heslop-Harrison 2008; Miller et al. 2008). A subset of the EST-SSR marker loci was screened for polymorphism across *M. acuminata* accessions contrasting in resistance to *Mycosphaerella* leaf spot diseases.

Materials and methods

Bioassays

In vitro-derived, 6-month-old whole plants of *M. acuminata* Calcutta 4 (BLSD resistant) and Cavendish Grande Naine (BLSD susceptible) (*Musa* International Transit Centre accessions ITC0249 and ITC0654, respectively) were maintained in a greenhouse under a 12-h light/12-h dark photoperiod at 25 °C and 85 % relative humidity. Leaf disc materials (squares of 36 cm²) for the two contrasting *M. acuminata* cultivars were collected from the two youngest leaves and spray inoculated on the adaxial surface using conidiospore suspensions (3×10^3 mL⁻¹) of the *Mf* strain CIRAD89. Inoculated leaf discs were incubated in a climatic chamber at 25 °C, again under a 12-h light/12-h dark photoperiod. Calcutta 4 was shown to be highly resistant, with a typical incompatible response, whereas Cavendish was found to be highly susceptible, displaying symptoms of a compatible interaction. Seven replicate leaf discs were prepared to ensure sufficient material for RNA purification and microscopic examination following infection. The *in vitro*-infected leaf disc tissues were maintained for extended periods in a green, non-senescent state, according to Abadie et al. (2008).

cDNA library construction

Two cDNA libraries were constructed, the first from a pool of RNA samples isolated from infected leaf discs at early time points in the incompatible interaction [4, 6, 7, 10, 12, 14 days after inoculation (DAI)] [*M. acuminata* ssp. *burmannicoides* Calcutta 4 (MAC4)] and the second from pooled late time points in the compatible interaction (19, 25, 31, 39 DAI) [*M. acuminata* cv. Cavendish Grande Naine (MACV)]. This approach was adopted not only to generate EST resources, but also to potentially enrich the unigene set for genes involved in defence responses during this host-pathogen interaction. Collected leaf material was flash frozen in liquid nitrogen to prevent RNA degradation and stored at -80 °C. Total RNA was extracted from leaf tissue using the Trizol kit (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's instructions. Total RNA quantification and quality analyses were conducted on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Poly A⁺ RNA was isolated from total RNA using a Micro-Poly(A)PuristTM mRNA Isolation Kit (Ambion, Austin, TX, USA), according to the manufacturer's instructions. Full-length cDNA libraries were constructed using the Creator SMART cDNA Library Construction kit (Clontech, Palo Alto, CA, USA). Poly A⁺ RNA quality was compared with an in-house control, and cDNA synthesized by reverse transcriptase, via long-distance polymerase chain reaction

(PCR). High-quality cDNA was isolated via fractioning, digested with Sfi and ligated to the plasmid cloning vector pDRN-LIB (Clontech). Transformation into *Escherichia coli* and recombinant selection on selection medium followed the manufacturer's protocols. Library qualities were examined by colony PCR and PCR amplification of plasmid inserts from randomly selected cDNA clones, with over 90 % showing inserts >400 bp. A total of 27 648 clones were prepared for each cDNA library and preserved as glycerol cultures.

Sequence analysis

Randomly selected clones from each cDNA library were 5'-end single-pass di-deoxy-based Sanger sequenced in Brazil at the Universidade Católica de Brasília, Embrapa Recursos Genéticos e Biotecnologia and in Japan at the National Institute of Agrobiological Resources using BigDye chemistry (Applied Biosystems, Foster City, CA, USA). A total of 14 272 sequences were generated from the MAC4 library and 7623 from the MACV library. Sequence analysis began with base calling and quality assignment using the program Phred and a $Q < 16$ quality score (Q) threshold (Ewing and Green 1998). Low-quality sequences were removed using the program Lucy (Chou and Holmes 2001) and vectors were masked using Cross_Match (Ewing and Green 1998). Sequences were screened for contaminant *E. coli*, chloroplast and mitochondrial DNAs utilizing the SSAHA package (Website 1, <http://www.sanger.ac.uk/Software/analysis/SSAHA/>). The processed sequences were assembled into sequence consensi with the program TGICL (Pertea et al. 2003).

To annotate unique transcripts (unigenes) and identify putative functions, similarity searches were performed on assembled sequences using the Basic Local Alignment Search Tool (BLAST) suite of programs, version 2.2.24+ (Altschul et al. 1997), against distinct databases to identify protein functional categories [NCBI non-redundant sequence database (Website 2, <http://www.ncbi.nlm.nih.gov/COG/>); The Swiss-Prot Database (Website 3, <http://www.uniprot.org/downloads>, uniprot_sprot_release of 2010 04 23); The TAIR Database: The Arabidopsis Information Resource (Website 4, <http://www.arabidopsis.org/>, Tair_9_pep_release 2009 06 19); KOG (clusters of eukaryotic orthologous proteins from complete eukaryotic genomes); LSE (lineage specific expansions); and TWOG (clusters for two species)]. BLASTX criteria accounting for identity significance were that the alignment length should be >100 amino acids and the expected value (E) $\leq 1E^{-10}$. Species distribution for *Musa* unigenes was calculated via homology searches against all plant proteins in the NCBI NR database, based upon best hit for each analysed sequence. An E -value $\leq 1E^{-3}$ was set as the threshold to consider a

BLAST hit significant. Unigene annotation based on protein domain comparisons with InterPro, Pfam and COG databases was conducted using InterProScan (version 4.5, <ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/>), HMMER3 (<http://hmm.janelia.org>) and BLAST analyses. Gene placement prediction was performed using Metabolic pathway annotation against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2004). Functional classification of annotated unigenes according to the categories of molecular function, biological process and cellular component was conducted using Blast2GO (Conesa and Götz 2008), following the gene ontology (GO) scheme (Consortium 2008).

Transposable elements (TEs) were identified during EST pre-processing steps using RepeatMasker Open-3.0 (<http://www.repeatmasker.org>) with the MIPS Repeat Element Database (Spannagl et al. 2007). Repeats were classified into superfamily, family and class according to version 4.3 of mips-REdat.

Candidate gene expression at different time points

The isolated RNA samples used for cDNA library construction were normalized and 10 μ g of each size separated via agarose gel electrophoresis (1.2 %) under denaturing conditions. Northern blot analyses of candidate gene expression at different time points during *Musa*-*Mf* interactions in the contrasting cultivars were carried out using Nylon Hybond N+ membranes according to the manufacturer's instructions. Polymerase chain reaction fragments of three selected cDNA clones of interest (GenBank accession numbers JK533438, JK545622 and JK535529) were labelled with α -³²P dCTP via random hexanucleotide-primed DNA synthesis using the Megaprime™ DNA Labelling System RPN 1607 (Amersham Biosciences, Piscataway, NJ, USA). Membrane hybridization signals were observed after exposure on an autoradiography Storm 820 imaging system (Amersham Biosciences, Piscataway, NJ, USA).

In silico SSR identification and marker development

A computational search using the program Mreps (Website 5, <http://bioinfo.lifl.fr/mreps/>) was used to locate perfect SSRs across EST subsets (2186 ESTs from the MAC4 library and 2363 from the MACV library). Microsatellite detection required the presence of at least two repeating units (e.g. GC) spanning >10 bp. Flanking forward and reverse primers were designed using the program Primer 3 (Rozen and Skaletsky 2000).

In order to assess amplification and allele length polymorphisms, markers were evaluated using 20 diploid (AA) *M. acuminata* accessions belonging to the Embrapa Cassava and Fruits breeding programme collection,

Table 1 Diploid (AA) *M. acuminata* accessions contrasting in resistance to *Mycosphaerella* leaf spot diseases selected for use in SSR marker validation.

<i>M. acuminata</i> accession	Resistance/susceptibility to BLSD	Resistance/susceptibility to yellow Sigatoka
Calcutta 4	Resistant	Resistant
Lidi	Resistant	Resistant
0323-03	Resistant	Resistant
SH32-63	Resistant	Susceptible
1304-06	Resistant	Resistant
0116-01	Resistant	Resistant
Burmanica	Resistant	Resistant
Microcarpa	Resistant	Resistant
1741-01	Nd ^a	Resistant
9179-03	Nd	Resistant
1318-01	Nd	Resistant
4279-06	Nd	Resistant
Pisang Berlin	Partially resistant	Susceptible
Niyarma Yik	Susceptible	Susceptible
Raja Uter	Nd	Susceptible
Tjau Lagada	Susceptible	Susceptible
F2P2	Nd	Susceptible
Khai Nai On	Susceptible	Susceptible
Sowmuk	Resistant	Susceptible
Jaribuaya	Resistant	Susceptible

^aNot defined.

contrasting in resistance to Sigatoka diseases, and potential parents for genetic map construction (Table 1). Genomic DNA was extracted from leaves of each accession using a modified mixed alkyl trimethyl ammonium bromide procedure (Gawel and Jarret 1991). Polymerase chain reactions were carried out in 13- μ L volumes, using 3 ng of genomic DNA, 2.5 mM MgCl₂, 0.2 mM dNTP mix, 0.5 μ M primer, 1.25 U of *Taq* polymerase (Invitrogen) and 1 \times buffer. Polymerase chain reaction amplification was conducted with the following temperature cycling: denaturation at 94 °C for 5 min; 29 cycles of denaturation at 94 °C for 1 min, specific primer annealing temperature for 1 min and product extension at 72 °C for 1 min; plus a final elongation period of 7 min at 72 °C. Polymerase chain reaction products were initially checked for amplicon size and PCR specificity on 3.5 % agarose gels in 1 \times TBE buffer. Allele sizes were determined for products run against 10-bp molecular size markers (Invitrogen) on denaturing

6 % polyacrylamide gels using 7 M urea. Polymerase chain reaction products were visualized by silver staining according to the standard protocols (Creste et al. 2001). Polymorphism per locus was calculated via the polymorphism information content (PIC) calculator (Website 6, <http://www.liv.ac.uk/~kempsj/pic.html>).

Results

Bioassay

A highly reproducible *in vitro* infection procedure was developed to assess the level of resistance to *Mf* in *M. acuminata*. Two *Musa* genotypes were selected for their contrasting resistance responses to the fungal pathogen, with Fig. 1 showing significant phenotypic differences at the macroscopic level. Following inoculation with *Mf* conidiospore suspensions, early cellular responses (19 DAI) were observed in Calcutta 4, leading to the activation of apoptotic events that blocked fungal growth after ingress via the stomata. Apoptosis was limited to sub-stomatal cells, with no further cell death progression observed between 19 and 31 DAI. These observations are indicative of a complete arrest of fungal growth in Calcutta 4. In this early biotrophic infection phase, such rapid induction of sub-stomatal cell death would deprive the fungus of nutrients required for survival. By contrast, the infection time course in leaves of the genotype Cavendish Grande Naine revealed fungal penetration of the host, with infection of sub-stomatal cells advancing in the mesophyll, resulting in extensive cell death during later necrotrophic stages (Fig. 1, magnified image, DAI31).

Unigenes

For the development of an EST dataset for *M. acuminata*, two full-length cDNA libraries were constructed, from MAC4 and MACV leaf tissue samples, both *in vitro* infected with *Mf*. The estimation of insert size via both restriction digestion with *Sfi*I and PCR amplification revealed averages in excess of 400 bp, showing that both cDNA libraries were of high quality.

From a total of 10 995 single-pass 5'-sequenced clones in the MAC4 cDNA library, vector trimming and quality analyses resulted in 9333 high-quality reads. In the case of the MACV cDNA library, from an initial 4157 clones, a total of 3962 high-quality reads were generated. Size distribution analysis revealed a mean length of ESTs following quality filtering and vector trimming of 370 bp for MAC4-derived ESTs and 494 bp for MACV-derived ESTs. The most common length distribution categories were between 201 and 500 bp for MAC4 ESTs, and between 401 and 500 bp in the case of MACV ESTs. All high-quality sequences were deposited in

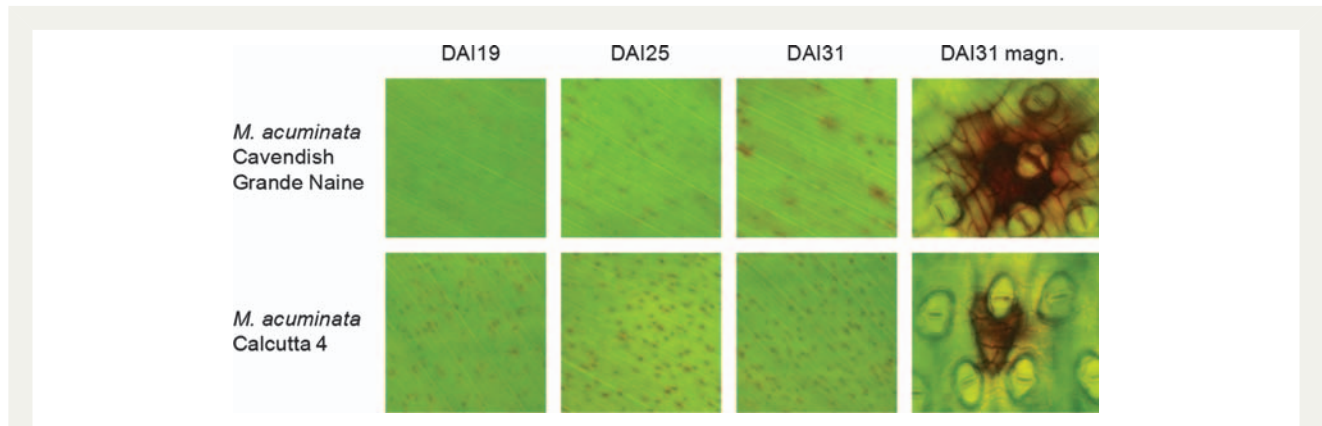


Fig. 1 Macroscopic and microscopic observation of infected tissues during the time course interaction with *Mf*. Symptoms of apoptosis were apparent only in Calcutta 4 (19–31 DAI), while cell death attributed to necrotic disease stages was restricted to Cavendish Grande Naine (DAI31). DAI, days after inoculation; magn, magnified.

NCBI with GenBank accession numbers JK531581–JK540913 (MAC4) and JK542313–JK546274 (MACV).

Assembly of high-quality *M. acuminata* ESTs from the two libraries generated 3995 non-redundant unigene clusters, consisting of 1368 contigs and 2627 singletons (1908 from MAC4 and 719 from MACV). Clustering resulted in an average of 16 EST sequences. As expected, contigs with fewer EST members were more represented than those composed of more ESTs (Fig. 2).

Functional annotation and classification

Expressed sequence tag annotation was conducted via the BLASTX algorithm-based alignment against the NCBI non-redundant sequence database, SwissProt, MIPS-*Arabidopsis*, GO and KOG. Conserved protein domains were also identified using InterproScan. A total of 2592 unigene sequences displayed significant homology to genes encoding proteins with known or putative function, 266 to genes encoding proteins with unknown function, and 1137 showed no significant homology to any sequences in the database. A total of 486 (12%) matched genes in rice (*Oryza sativa*), 182 (5%) matched genes in maize (*Zea mays*) and 247 (6%) matched genes in sorghum (*Sorghum bicolor*) (Fig. 3). Only 4.1% of BLAST hits (165 unigenes) originated from *Musa* NR database proteins, indicating considerable gene discovery for the genus. Gene ontology is employed to provide an organized vocabulary for describing unigenes according to categories (Ashburner et al. 2000). Functional annotation of the 3995 unigenes with InterproScan analysis identified a total of 543 GO terms. Unigenes were annotated with GO identifier into three principal categories: molecular functions (46.43%), cellular components (19.21%) and biological processes (34.34%). Two

unigenes (0.04%) remained unclassified, possibly reflecting limited sequence length or that they are novel proteins. Details of assigned high-level GO terms are shown in Fig. 4. As unigenes could occasionally be assigned to more than one category, the combined total number of assigned GO mappings exceeded the number of unigenes analysed. In the molecular function category, the four most represented unigene functional classes were: other enzyme activity (468), other binding (262), nucleotide binding (236) and structural molecule activity (197). The principal functional classes observed in the biological function category belonged to metabolic process (272), translation (271), protein metabolic process (204) and transport (175). In the cellular component category, most unigenes coded for intracellular cell part (242), ribosome (199), membranes (123) and macromolecular complex (89).

KOG categories A total of 2300 unigenes were assigned to KOG categories amino acid transport and metabolism (72/1.8%), carbohydrate transport and metabolism (110/2.8%), cell cycle control, cell division, chromosome partitioning (28/0.7%), cell motility, lipid transport and metabolism (1/0.02%), cell wall/membrane/envelope biogenesis (28/0.7%), chromatin structure and dynamics (23/0.5%), coenzyme transport and metabolism (19/0.4%), cytoskeleton (49/1.2%), defence mechanisms (26/0.6%), energy production and conversion (127/3%), extracellular structures (8/0.2%), function unknown (160/4%), general function prediction (319/7.9%), inorganic ion transport and metabolism (67/1.6%), intracellular trafficking, secretion and vesicular transport (119/2.9%), lipid transport and metabolism (56/1.4%), nucleotide transport and metabolism (21/0.5%), post-translational modification, protein turnover, chaperones (320/8%),

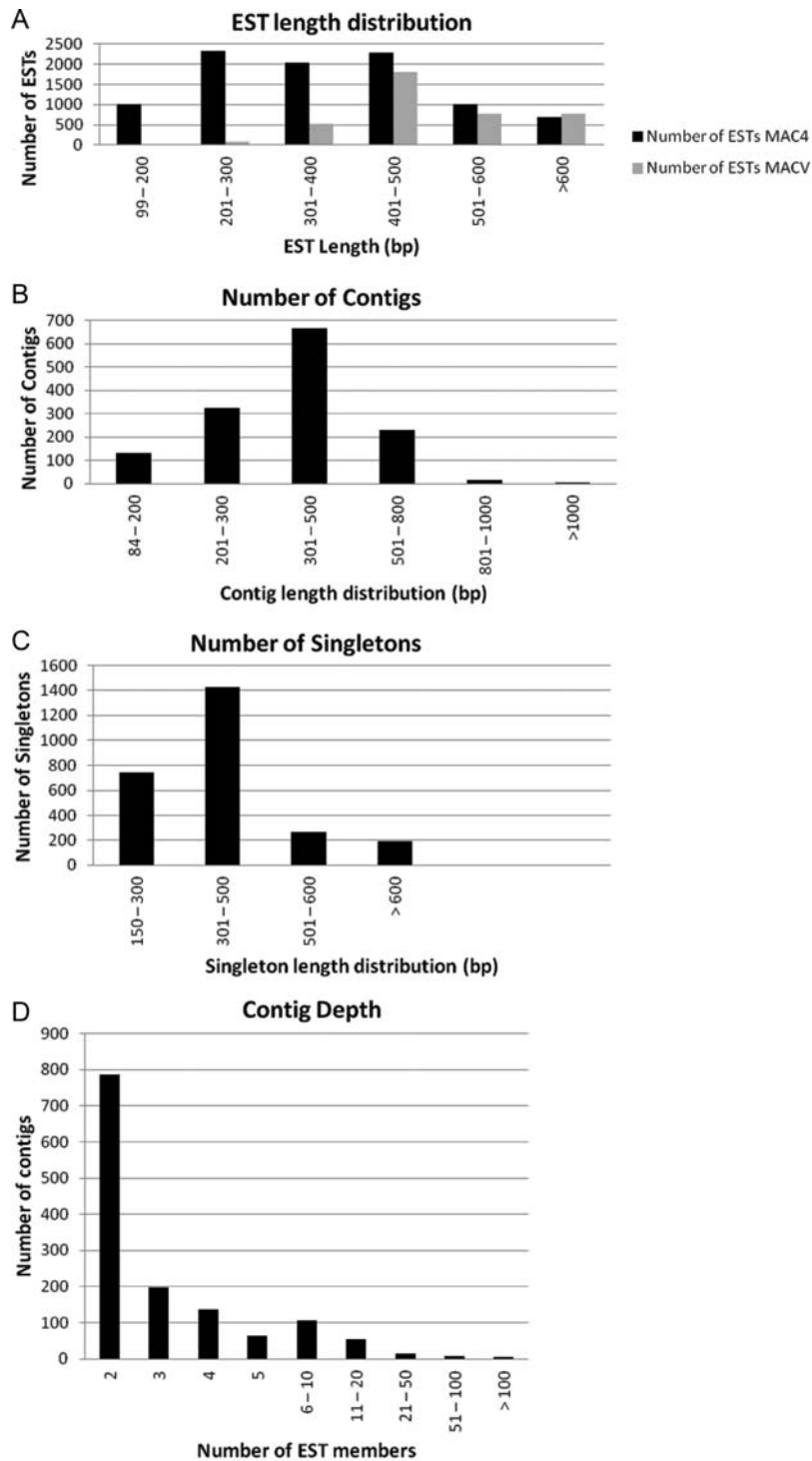
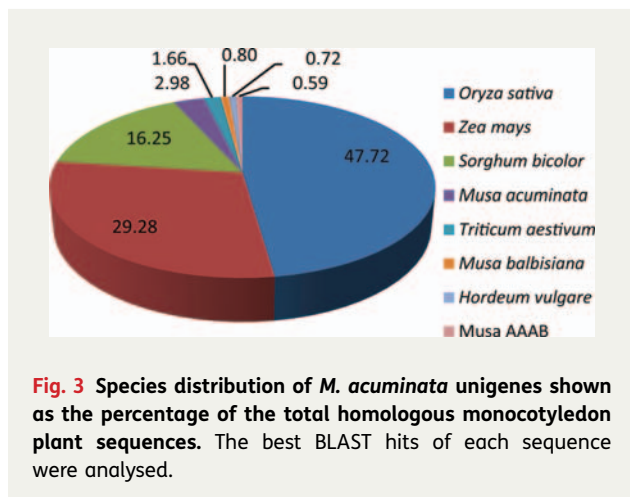


Fig. 2 Summary of EST quality and sequence assembly from combined MAC4 and MACV datasets. (A) Length distribution of the *M. acuminata* ESTs; (B) length distribution of the assembled *M. acuminata* unigene contigs; (C) length distribution of the *M. acuminata* unigene singletons; (D) frequency and distribution of ESTs in assembled *M. acuminata* unigene contigs.



replication, recombination and repair (15/0.3 %), RNA processing and modification (89/2.2 %), secondary metabolite biosynthesis, transport and catabolism (53/1.3 %), signal transduction mechanisms (172/4.3 %), transcription (109/2.7 %), translation, ribosomal structure and biogenesis (289/7 %), and no hits (1696/42.5 %) [see ADDITIONAL INFORMATION 1 and 2].

Defence Defence-related transcript functions identified according to the KOG classification across the unigenes comprised nine germin/oxalate oxidases (OXOs), three regulators of pathogen resistance responses of RPS2 and RPM1 genes, two dual-specificity phosphatases, two flavonol reductase/cinnamoyl-CoA reductases, two HVA22/DP1 gene product-related proteins, two macrophage migration inhibitory factors, a drought-induced protein, a bax-mediated apoptosis inhibitor TEGT/BI-1, a BPI/LBP/CETP family protein, a mercaptopyruvate sulfurtransferase/thiosulfate sulfurtransferase, a predicted protein tyrosine phosphatase, and a protein involved in the control of unknown local host defence mechanisms to pathogens.

Signal transduction From a total of 172 predicted unigenes classified to the KOG category of signal transduction, a number are typically associated with plant immunity mechanisms. These included 20 unigenes characterized as ‘Receptor protein kinase containing LRR repeats’. Other significant findings were 20 serine/threonine protein kinases, two mitogen-activated protein kinase kinase (MAP2K) and five WRKY superfamily transcription factors. Further manual mining also revealed predicted unigenes typically associated with defence responses in plants: five isoflavone reductase/pinoresinol-lariciresinol reductase/phenylcoumaran benzylic ether reductases (phenylpropanoid/flavonoid pathway); 14 glutathione S-transferases (GSTs), two metallothioneins,

three superoxide dismutases (SOD) (plant detoxification); one 1-aminocyclopropane-1-carboxylate synthase (ethylene biosynthesis); three β -1,3 glucanases (PR 2 proteins); one transcription factor containing NAC and TS-N domains (plant defence); one cysteine proteinase inhibitor B (cystatin B) (plant defence); and six glycolate oxidases [production of reactive oxygen species (ROS)].

Functional validation of defence-related gene expression

By employing RNA samples used for cDNA library construction, a northern blot time course was conducted to assess differential induction of a set of identified defence-related genes following infection of each *M. acuminata* genotype. Selected candidate genes comprised an OXO (clone accession number JK533438), one representative of the metallothionein type 2 gene family (JK545622) and one peroxidase (JK535529). The time course for the analysis of gene expression during the interaction analysis covered early (4 DAI) until late time points (31 DAI in the Calcutta 4–*Mf* interaction, 39 DAI in the Cavendish Grande Naine–*Mf* interaction). Analysis revealed differences in the pattern of expression induction of the selected genes between the incompatible and compatible *Musa*–*Mf* interaction. Early induction of the OXO (4 DAI, 6 DAI), the metallothionein (6 DAI) and the peroxidase (6 DAI) in Calcutta 4 cells correlated with the observed apoptotic events (Fig. 5), suggesting their involvement in a rapid activation of defence responses. By contrast, no significant early induction of the three genes was observed in Cavendish Grande Naine, with an increased expression of both OXO and metallothionein only 31 and 39 DAI, and a relatively constant expression of peroxidase throughout the time course.

KEGG pathways To identify biological pathways represented among the unigenes, enzyme commission numbers derived from BLASTX alignments were mapped against the KEGG database. A total of 312 unigenes were identified in the pathway maps, with the categories genetic information processing and metabolism accounting for 86 % of the unigenes (Table 2). The five most represented pathway subcategories were: translation; energy metabolism; folding, sorting and degradation; carbohydrate metabolism; and amino acid metabolism.

Transposable elements

To analyse the abundance and diversity of expressed TEs, EST pre-processing employed RepeatMasker (version open-3.2.8), with classification to type level according to the database mips_REdat_4.3. Differences in the proportion of retrotransposons (85 %) and transposons (16 %)

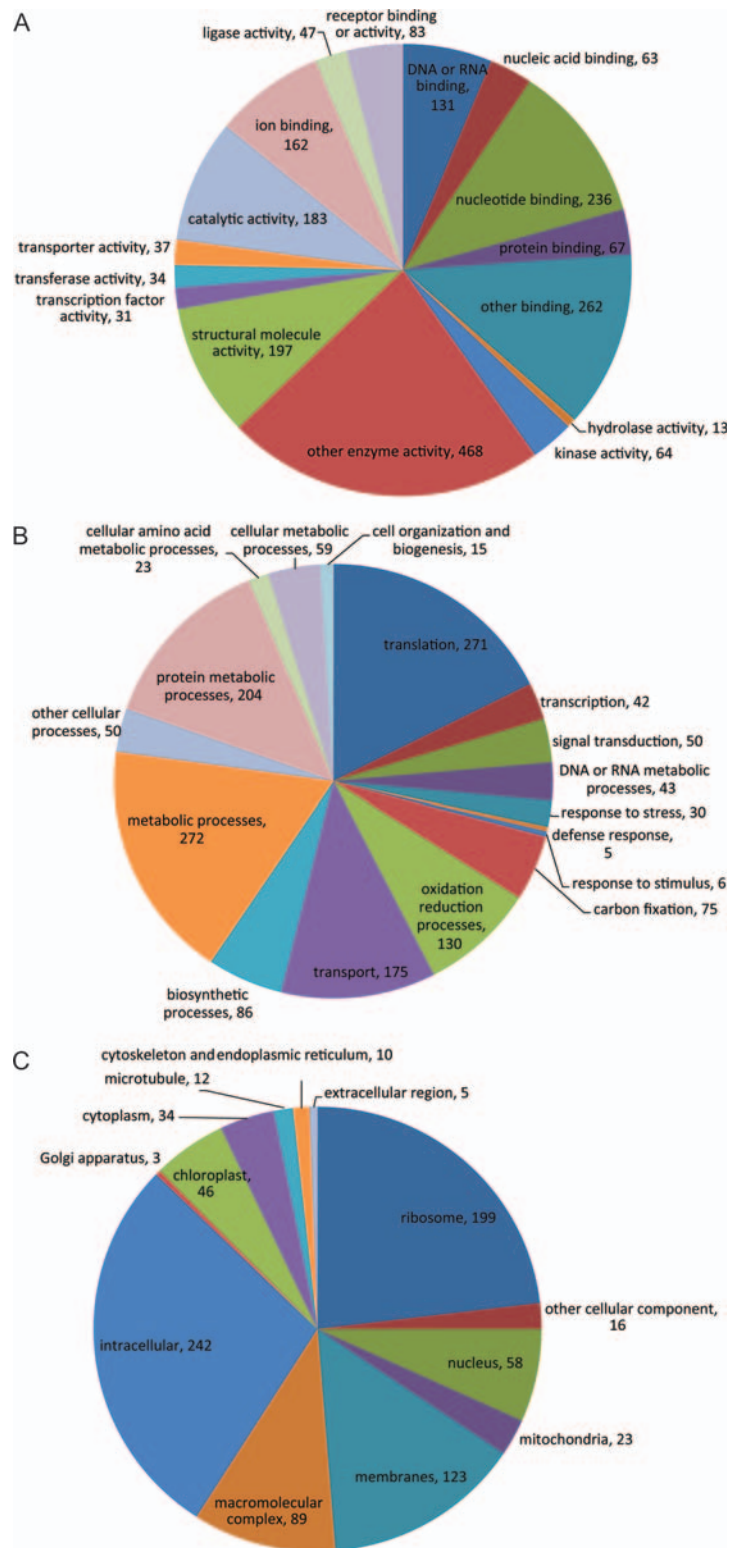


Fig. 4 Representation of *M. acuminata* unigenes classified in functional groups. The GO hits were assigned to categories (A) molecular function, (B) biological process and (C) cellular component.

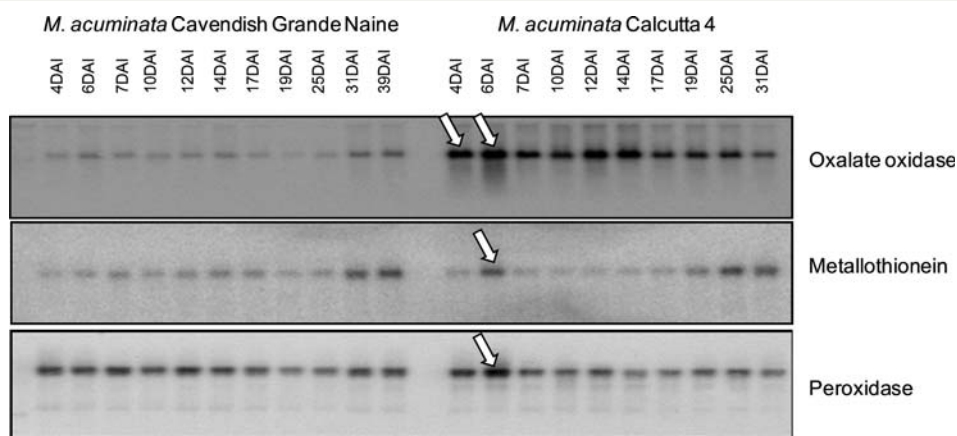


Fig. 5 Northern blot analysis of expression of selected defence-related *M. acuminata* unigenes during the time course interaction with *Mf*. Arrows show differential gene expression between the tested cultivars for each gene probe.

Table 2 Summary of *M. acuminata* unigenes mapped in KEGG pathways.

Pathway categories	Unigenes
Genetic information processing	143
Metabolism	128
Cellular processes	18
Organism systems	14
Environmental information processing	9
Unigene total	312

Table 3 Abundance and diversity of expressed TEs in *M. acuminata* EST datasets.

Class	Type	Number of ESTs
I	LTR	220
I	Non-LTR	21
I	Retrotransposon	16
II	Transposon	48
Total		305

were observed. Class I TEs were classified into long terminal repeat (LTR), non-LTR and retrotransposon type, while Class II TEs were classified only to transposon type. Table 3 summarizes the number of EST sequences containing each TE type.

Genic-SSR marker development

Computational mining of *M. acuminata* ESTs (2186 from the MAC4 library and 2363 from the MACV library, with a total size of 2104 Mbp) identified SSRs across 13.7 % of sequences. For 303 out of 624 SSR-positive sequences, PCR primers could be successfully designed, for potential use as molecular markers based on repeat length polymorphisms (Table 4). A total of 12.5 % of analysed MAC4 ESTs contained SSRs, with five classes identified. The trinucleotide repeats appeared the most abundant (54.1 %), followed by di- (31.6 %), tetra- (6.7 %), hexa- (5.3 %) and penta-nucleotide repeats (2.3 %). The most abundant trinucleotide repeat motifs were GAA, CTC, AAG, AGA, CCT,

CAG, GAG, GAT, CAC and AGG, accounting for 68 % of such repeats. Of the dinucleotide repeat motifs, GA, AG, TC and CT accounted for 78.5 % of repeats. Tetranucleotide repeats were less abundant, with the majority of motifs in equal abundance (11.1 % each), with the exception of the more frequent GAGG motif (22.2 %). Penta- and hexa-nucleotide repeats represented the least abundant in Calcutta 4, with equal abundance observed for each motif. Analysis of MACV ESTs revealed 14.8 % containing SSRs. In contrast to Calcutta 4, a greater array of repeat classes was observed, from di- through to hendecanucleotide repeats. As in the case of Calcutta 4, trinucleotide repeats were the most abundant (57.6 %). These were followed, in decreasing frequency, by di- (25.3 %), tetra- (7.1 %), hexa- (4.7 %), penta- (3.5 %), hepta- (0.6 %), octa- (0.6 %) and hendeca-nucleotide repeats (0.6 %). Trinucleotide repeat motifs included, in decreasing prevalence, CTC, AGA, TTC, AAG, GAA, CCT, GGA and TCT, representing 50.0 % of tri-repeats. The most common dinucleotide repeat motifs GA, TC, AG and CT,

Table 4 Overview of SSR repeat abundance in *M. acuminata* ESTs and primer design statistics.

	<i>M. acuminata</i> Calcutta 4	<i>M. acuminata</i> Cavendish Grande Naine
Sequences analysed	2186	2363
Sequences with SSR repeats	273	351
Bases analysed	1 021 899	1 082 286
Bases with repeats	5478	7171
Primer pairs designed	133	170
Failed primers	51	67

also common in Calcutta 4, accounted for 88.4 % of repeats. Tetranucleotide repeat motifs were all present in equal abundance (8.3 % each). Penta- and hexanucleotide repeat motif types were also each present in equal abundance per class, at 16.7 and 12.5 %, respectively. In the case of hepta-, octa- and hendeca-nucleotide classes, only one motif type per class was observed. In general, the shorter the nucleotide core sequence, the greater were the number of repeats observed. In the case of Calcutta 4 there were an average of 9.4 repeats for di-nucleotide motifs, 5.2 for tri-, 3.5 for tetra-, 3.3 for penta- and 3.4 for hexa-motifs. Similarly, for Cavendish Grande Naine there were an average of 9.6 repeats for di-, 5.4 for tri-, 3.9 for tetra-, 3.1 for penta-, 3.8 for hexa-, 3.1 for hepta-, 3 for octa- and 3.1 for hendeca-motifs.

Of the 303 EST-derived SSR markers for which primers could be designed, 149 yielded reproducible PCR amplicons [see ADDITIONAL INFORMATION 3]. A total of 75 (24.7 %) were identified with consistent amplification and as polymorphic loci when tested, initially on agarose gels and subsequently on polyacrylamide gels, across the contrasting *M. acuminata* accessions (Table 5). A total of 289 alleles were scored across these polymorphic loci. Fourteen polymorphic loci possessed two alleles across the tested accessions; 21 loci showed three alleles; 17 loci showed four alleles; 13 loci showed five alleles; six loci showed six alleles; three loci showed seven alleles; and one locus displayed eight alleles. The PIC values ranged from 0.08 to 0.81, with an average value of 0.50.

Discussion

The objectives of this work were to generate an EST resource for studying functional genes in *M. acuminata*, which also included transcripts expressed in banana-*Mf* interactions during compatible and incompatible

reactions. We also pursued the development of gene-based microsatellite markers as a resource for genetic mapping, diversity characterization and MAS of specific traits in conventional breeding populations.

Unigenes

In total, 9333 high-quality ESTs were generated from MAC4 and 3964 from MACV. At the time of analysis in December 2011, only 15 464 ESTs were publically available for *M. acuminata*. This study therefore contributes almost a two-fold increase in EST resources for this species. BLASTX homology searches of the 3995 *M. acuminata* unigenes against monocotyledonous plant proteins in the NCBI NR database revealed 28.4 % of unigenes as potentially novel and exclusive to *M. acuminata*, with only 4.1 % of BLAST hits to existing *Musa* NR database proteins. This dataset therefore provided a significant contribution of value for gene discovery and validation of function for the genus.

Functional categorization assigned a large number of unigenes to involvement in intracellular cell components, membranes, organelles, metabolic processes, translation, transport, oxidation and reduction processes, enzyme activity, binding, structural molecule activity and catalytic activity. Given the still limited characterization of gene expression during banana-*Mf* interactions (e.g. Portal et al. 2011), a strategy for potential enrichment of *Musa* EST resources to also include genes involved in defence responses was employed. Given that defence responses typically occur earlier in incompatible rather than compatible interactions, distinct time points for cDNA library preparation were chosen to reflect such expected differences. Although the sequences encoding activities related to response to stress, defence response and signal transduction were less represented, numerous unigene sequences potentially involved in plant effector-triggered immunity (ETI) and pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) were characterized. Pathogen-associated molecular pattern-triggered immunity is considered to be based upon interactions between host pattern recognition receptor-like kinases and conserved PAMPs (Nürnberg and Kemmerling 2009), conserved across a microbial class and essential in fitness. Pathogen-associated molecular pattern-triggered immunity involves activation of a mitogen-associated protein kinase cascade and WRKY transcription factors (TFs), conferring resistance to the majority of potential pathogens. Effector-triggered immunity (Jones and Dangl 2006) is based upon co-evolution of plant resistance R-protein receptors and specific pathogen effector molecules, conferring resistance at the intra-specific level. Many downstream signal

Table 5 Characteristics of polymorphic microsatellite loci isolated from *M. acuminata* Calcutta 4 and Cavendish Grande Naine EST data.

Locus name	SSR repeat motif	SSR locus length (bp)	Obtained allele size range (bp)	Allele number	He	Ho	PIC value
MA41	AG	27	350–390	8	0.83	0.63	0.81
MA45	TCTT	16	110–120	3	0.6	0.88	0.53
MA47	CTC	16	140–160	2	0.48	0.81	0.37
MA412	GAA	25	350–360	3	0.67	0.57	0.59
MA413	CTC	16	240–250	3	0.55	0.42	0.48
MA418	GAG	13	320–330	4	0.7	0.61	0.64
MA425	AG	21	220–250	7	0.79	0.79	0.76
MA426	TAAT	14	150–160	3	0.27	0.22	0.25
MA428	AAG	36	200–220	4	0.64	0.33	0.57
MA432	CAC	13	300–320	5	0.67	0.76	0.61
MA435	GAA	19	330–350	4	0.71	0.83	0.66
MA440	TTC	19	170–175	2	0.19	0.13	0.17
MA441	CT	13	100–110	5	0.63	0.52	0.58
MA443	TTC	18	310–340	5	0.71	0.58	0.67
MA444	CT	33	260–280	3	0.29	0.35	0.26
MA446	TC	14	140–150	3	0.61	0.27	0.53
MA452	AGA	21	170–180	5	0.7	0.73	0.66
MA453	CGC	12	330–335	2	0.50	0.21	0.37
MA455	GAA	20	240–250	3	0.54	0.23	0.45
MA466	CAG	16	170–180	4	0.75	0.67	0.7
MA472	GAGAG	15	300	4	0.60	0.35	0.54
MA476	AC	13	200–210	4	0.72	0.58	0.67
MA479	CCT	13	340–350	4	0.3	0.38	0.26
MA489	AT	14	260–300	3	0.44	0.23	0.37
MA490	CAC	16	190–200	3	0.57	0.47	0.49
MA492	AT	23	175–180	5	0.69	0.75	0.65
MA493	CTC	12	310–320	2	0.5	0.5	0.37
MA495	GAA	16	180–220	2	0.42	0.00	0.33
MA4100	AG	14	320–335	3	0.38	0.48	0.33
MA4104	GA	47	300–320	4	0.61	0.38	0.55
MA4110	TCT	30	280–300	6	0.74	0.25	0.70
MA4111	AGC	26	375–290	4	0.73	0.52	0.68
MA4116	CT	23	380–410	5	0.72	0.65	0.68
MA4128	CTT	25	210–220	3	0.47	0.47	0.38
MACV11	CAG	20	210–220	4	0.5	0.5	0.46
MACV15	GA	22	270–290	6	0.81	0.33	0.79
MACV20	TC	12	340–360	5	0.72	0.83	0.66
MACV21	TCA	36	340–360	4	0.47	0.57	0.43
MACV27	TC	36	290–310	5	0.73	0.63	0.68

Continued

Table 5 Continued

Locus name	SSR repeat motif	SSR locus length (bp)	Obtained allele size range (bp)	Allele number	He	Ho	PIC value
MACV29	TAAT	14	140–150	2	0.39	0.13	0.32
MACV36	CT	33	240–270	5	0.78	0.77	0.74
MACV37	AGC	21	100–110	3	0.1	0.1	0.09
MACV42	AGA	24	320–330	3	0.56	0.25	0.48
MACV47	GA	32	320–320	3	0.64	0.61	0.56
MACV49	AGGCGA	21	270–275	2	0.39	0.33	0.31
MACV50	TC	13	130–140	3	0.1	0.1	0.09
MACV51	CT	19	180–200	4	0.4	0.48	0.37
MACV54	CAG	12	160–180	4	0.68	0.7	0.62
MACV55	AGA	18	170–190	6	0.73	0.83	0.69
MACV62	TTTTTA	28	250–280	3	0.53	0.1	0.43
MACV63	TGC	25	120–140	6	0.72	0.67	0.67
MACV73	CTCTC	16	220–225	2	0.41	0.42	0.33
MACV77	GCC	12	90–115	4	0.58	0.54	0.52
MACV81	GAA	15	200–205	2	0.28	0.05	0.24
MACV87	TCC	15	110–130	4	0.69	0.83	0.64
MACV88	GAA	27	250–300	5	0.76	0.71	0.72
MACV96	TCT	15	180–210	5	0.75	0.57	0.70
MACV99	GA	13	240–260	2	0.08	0.08	0.08
MACV104	AG	14	290–330	7	0.81	0.65	0.79
MACV109	GGA	23	385–390	2	0.08	0.00	0.08
MACV111	CTACA	16	130–150	2	0.63	0.55	0.60
MACV112	CTC	12	330–360	3	0.78	0.74	0.74
MACV115	AAG	12	150–165	6	0.59	0.55	0.51
MACV128	ATGCTC	20	420–450	4	0.43	0.27	0.39
MACV132	CTC	31	280–290	4	0.58	0.65	0.53
MACV134	CCT	13	320–335	3	0.51	0.78	0.42
MACV139	GA	17	90–120	7	0.75	0.95	0.71
MACV148	CTC	19	150–170	5	0.69	0.57	0.63
MACV151	TTC	21	270–290	5	0.75	0.87	0.71
MACV154	AAG	12	130–150	3	0.12	0.13	0.12
MACV155	TC	21	330–350	3	0.63	0.26	0.56
MACV157	GGA	13	140–155	2	0.29	0.35	0.25
MACV161	GCA	12	110–120	3	0.61	0.42	0.53
MACV162	ATCTG	15	190–200	2	0.15	0.17	0.14
MACV169	TC	37	150–170	6	0.79	0.75	0.76

Polymorphism was evaluated across 20 *M. acuminata* accessions.

HE, expected heterozygosity under Hardy–Weinberg expectations; HO, observed heterozygosity; PIC, polymorphism information content.

transduction components are shared between PTI and ETI, including an oxidative burst via the production of ROS and changes in plant hormone levels. Mitogen-activated protein kinase signalling cascades also occur in both PTI and ETI, with variations in duration probably responsible for differential downstream responses in the two immunity branches (Tsuda and Katagiri 2010). Overall, a number of expressed genes potentially involved in different pathways in PTI or ETI were identified, the most abundant of which included host receptor genes involved in PAMP or pathogen effector recognition, unigenes involved in signalling mechanisms, phenylpropanoid/flavonoid pathway genes, phytohormone biosynthesis genes, pathogenesis-related protein coding unigenes and genes involved in plant detoxification.

Host receptor genes and signal transduction Distinct plant *R*-gene families are recognized as involved in ETI, based upon protein domain structure and biochemical function. The most abundant class encodes cytoplasmic receptor proteins containing nucleotide binding site-leucine-rich repeat (NBS-LRR) domains (Hammond-Kosack and Jones 1997). In rice, ~400 NBS-LRR genes have been characterized, with around 150 present in the *Arabidopsis* genome (McHale et al. 2006). Conservation of motifs within *R*-genes, such as those present within NBS-LRR domains, have facilitated amplification in diverse plant taxa. Such work has been reported in *Musa*, with large-scale analyses of NBS-LRR *R*-gene family RGA diversity across the genus (e.g. Miller et al. 2008; Mohamad and Heslop-Harrison 2008). In the current study, three regulators of pathogen resistance responses of NBS-LRR *R*-genes RPS2 and RPM1 genes were identified in the transcribed unigene dataset. The RPM1 protein is known to be associated with the host plasma membrane (Boyes et al. 1998), as is RPS2 (Mackey et al. 2003), where they recognize modification in the *Arabidopsis thaliana* negative regulator RPM1 interacting protein 4 (RIN4), target of *Pseudomonas syringae* type III bacterial effector proteins (Mackey et al. 2002), triggering the hypersensitive response (HR) or programmed cell death of infected cells, characterized by the appearance of small necrotic lesions at infection sites. Other known plant *R*-gene classes include extracellular LRRs anchored by transmembrane domains (receptor-like proteins), extracellular LRRs linked to cytoplasmic serine–threonine kinase domains, intracellular serine–threonine kinases and proteins with a coiled-coil domain anchored to the cell membrane. EuKaryotic orthologous group-based analysis predicted a total of 20 unigenes with function assigned as ‘Receptor protein kinase containing LRR repeats’. Over 600 such receptor-like kinases (RLKs) have

been characterized in *Arabidopsis* (Shiu and Bleeker 2001), with the disease ETI resistance gene *Xa21* being one of the earliest examples from this class, conferring durable resistance to *Xanthomonas oryzae* pv. *oryzae* (Song et al. 1995). In a previous study in *M. acuminata* (Miller et al. 2011), sequence similarity analysis of amplification products generated using degenerate primers for RLKs identified numerous sequences with significant similarity to *R*-gene and RGA sequences for this class. A total of 20 serine/threonine protein kinases were also identified on the basis of KOG function assignment. Examples of such kinases include the intracellular cytoplasmic *R*-gene *Pto*, which was the first *R*-gene in tomato (*Solanum lycopersicum*) proved to confer resistance to *Pseudomonas syringae* pv. *tomato* strains that express the *AvrPto* gene (Martin et al. 1993). Defence reactions associated with HR and programmed cell death are considered to be induced following *AvrPto* recognition in the presence of an NBS-LRR protein known as *Prf*, which is present in the *Pto* kinase gene cluster. Other significant findings in relation to unigenes typically involved in signal transduction from pathogen recognition to defence gene expression included two MAP2K and five WRKY superfamily transcription factors.

Phenylpropanoid/flavonoid pathway Phenylpropanoids in plants are involved in a number of defence responses, acting as antimicrobial compounds (phytoanticipins and phytoalexins) and molecules involved in signalling (Dixon et al. 2002; Naoumkina et al. 2010). EuKaryotic orthologous group classification revealed five isoflavone reductase/pinoresinol-lariciresinol reductase/phenylcoumaran benzylic ether reductases. Isoflavone reductase is an enzyme required for biosynthesis of the phytoalexin pterocarpan. Monolignols serve as precursors of plant lignins and lignans, which are composed of phenolic compounds and are involved in physical and chemical plant defence mechanisms. Cinnamoyl-CoA reductase is the first enzyme specific for monolignol synthesis. EuKaryotic orthologous group data identified two unigenes encoding this enzyme.

Pathogenesis-related proteins Pathogenesis-related (PR) proteins were initially observed in tobacco (*Nicotiana tabacum*) and are now known to accumulate in diverse plant hosts when under pathogen attack. These structurally and functionally diverse proteins have been classified into 17 families (van Loon et al. 2006). Given that both HR observed in incompatible plant–pathogen interactions and subsequent systemic acquired resistance (SAR) to diverse pathogens are associated with accumulation of PR proteins in local and systemic tissues, such proteins are believed to contribute to

resistance. Our unigene set included three β -1,3 glucanases, which are recognized as PR-2 family members. This widely studied family has been reported to limit activity in diverse fungal pathogens, through degradation of the cell wall component β -1,3 glucan. Up-regulation of β -1,3 glucanases in incompatible interactions has been reported (Elvira et al. 2008), and over-expression analyses have confirmed involvement in resistance. For example, PR2 genes from soybean (*Glycine max*) have been shown to confer resistance in potato (*Solanum tuberosum*) to *Phytophthora infestans* (Borkowska et al. 1998) and in kiwi (*Actinidea deliciosa*) to *Botrytis cinerea* (Grover and Gowthaman 2003). Similarly, a PR2 gene from potato increased resistance to both *Fusarium oxysporum* and *Fusarium culmorum* in flax (*Linum usitatissimum*) (Wróbel-Kwiatkowska et al. 2004).

Germin OXOs Numerous germin OXOs were encountered in the unigene sets. Within the germin protein family, OXOs have been reported to play roles in calcium regulation, oxalate metabolism and response to pathogenesis (Davidson et al. 2009). Evidence for the latter includes up-regulation in cereals in response to powdery mildew (Zhou et al. 1998) and co-segregation of markers for OXO genes and with rice blast resistance QTLs (Wu et al. 2004). Oxalate oxidases can catalyse the conversion of ROS to H_2O_2 (Requena and Bornemann 1999), important components of HR in plants. Hydrogen peroxide is involved in cell wall cross-linking and messenger activity for activation of defence genes, triggering SAR. Also reported as a molecule necessary for phytoalexin biosynthesis, H_2O_2 has been shown to have direct antimicrobial activity, causing oxidation of invading pathogens (Wei et al. 1998). Our northern blot data revealed early increased expression (4–6 DAI) of OXO in resistant Calcutta 4 only, suggesting a possible involvement in ROS and associated HR components.

Plant detoxification Eukaryotic orthologous group-derived mining revealed 14 GSTs in the unigene sets. Glutathione S-transferases appear to be ubiquitous in plants, with a function in endogenous and xenobiotic compound detoxification, such as herbicides (Hayes and McLellan 1999). Up-regulation has been shown in individual GSTs during pathogen attack in numerous plant species (e.g. Mauch and Dudler 1993; Alvarez et al. 1998), with likely involvement in detoxification of products of oxidative stress during HR, thus limiting both cell damage and the extent of cell death. Expression in *Mf*-*M. acuminata* compatible interactions has recently been reported (Portal et al. 2011). The

potential role of GSTs in cell signalling pathways has also been suggested, with a GST from parsley involved in UV-dependent signal transduction (Loyall et al. 2000).

Metallothioneins are low-molecular-weight polypeptides rich in cysteine residues. Present across prokaryotes and eukaryotes, they play a role in detoxification and homeostasis, sequestering metal ions such as Cu^{2+} , Zn^{2+} and Cd^{2+} , and preventing mutations (Hamer 1986; Robinson et al. 1993). Up-regulation has been observed in plants in response to increased metal concentrations (Hsieh et al. 1995). Since plants experience oxidative stresses following pathogen infection, it has also been argued that this protein family might be associated with regulation of intracellular redox potential and oxygen detoxification (Hamer 1986; Choi et al. 1996), protecting cells from damaging effects of ROS. Previous reports also indicate differential regulation of metallothioneins after viral infection in tobacco (Choi et al. 1996), temperature stress (Hsieh et al. 1995) and foliar senescence. Four distinct types (MT1–MT4) have been described in plants, according to distribution of cysteine residues (Robinson et al. 1993). Expression of MT1 is generally more associated with vascular tissues and roots, MT2 with shoots and leaves, MT3 with leaves and mature fruits, and MT4 with seed tissues. Liu et al. (2002) reported isolation of MT2 and MT3 in banana, with expression influenced in response to ethylene and metals. More recent examination of transcripts has reported that metallothionein-like genes are abundant in *M. acuminata* Calcutta 4 (Santos et al. 2005). Our study confirmed this, with isolation of MT2 and MT3 unigene sequences derived from contigs with considerable numbers of EST members. Northern blot data showing early expression of type 2 metallothionein-like proteins only in Calcutta 4 suggest involvement in cell protection ROS-scavenging during HR responses. By contrast, it is possible that late expression in Cavendish Grande Naine may indirectly reflect increased ROS during the fungal necrotrophic disease phase. Necrotrophs have been reported to induce ROS accumulation in plant hosts as a mechanism for promoting pathogen access to nutrients through triggering host programmed cell death (Govrin and Levine 2000).

Peroxidases A number of peroxidases were also observed in the unigene sets. In addition to auxin metabolism, cell wall reinforcement and phytoalexin synthesis, such enzymes are also typically involved in ROS metabolism during defence responses (Almagro et al. 2009). Recent histochemical analysis of peroxidase and H_2O_2 accumulation during *Mf*-*M. acuminata* interactions reported a peaked accumulation of both in resistant *M. acuminata* Calcutta 4 at 10 days after conidial

inoculation of detached leaf material, with no accumulation in susceptible Cavendish Grande Naine and partially susceptible Pisang Madu (Cavalcante et al. 2011). The marked early peroxidase gene expression induction 6 DAI observed only in resistant Calcutta 4 is in agreement with these previous findings, and again correlates with HR-like responses observed in this genotype.

Transposable elements

Transposable elements are known to occur in all living organisms, and can occupy over 50 % of nuclear DNA. Given that these elements display mobility, they are important in plant evolution, through creation of novel genes or modifying gene function (Bennetzen 2000). In the case of vegetatively propagated crops such as banana, it is therefore likely that some somaclonal variation events can be due to such TE activity. Classification of eukaryotic TEs is based on the mode of transposition, with RNA-mediated TEs (Class I) and DNA TEs (Class II). Class I TEs can be divided into subclasses: long terminal repeats (LTRs), retroelements without LTRs (the long interspersed nuclear elements (LINEs) and the small interspersed nuclear elements (SINEs)) and TRIMs (Terminal-repeat Retrotransposons In Miniature). Class II TEs include the MITEs (Miniature Inverted-repeat Transposable Elements) (Feschotte et al. 2002). Our results revealed a predominance of retrotransposons to transposons. Similar distributions of DNA repeats have recently been reported in *M. acuminata* Calcutta 4, based on low-depth 454 sequencing of genomic DNA (Hribová et al. 2010).

Markers

The development of genomic libraries enriched for SSRs is typically expensive and labour intensive, in contrast to data mining in ESTs. Expressed sequence tag-derived SSR markers enable enrichment of genetic maps with gene-based markers (Kota et al. 2001), as opposed to anonymous genomic DNA-derived SSRs which are predominantly derived from intergenic regions. Given that markers are isolated from coding regions, conservation is expected to be high, such that these EST-SSR markers are generally also transferable to related species (e.g. Gupta et al. 2003). The gene-based marker tools developed in this study for *Musa* also serve as a resource for diversity characterization and downstream marker-assisted breeding using markers for traits. Work is ongoing in the research community for the development of suitable segregant populations for traits of interest (Amorim et al. 2009; Dochez et al. 2009; Lorenzen et al. 2011). Linkage disequilibrium mapping is a potential alternative route for identifying

genes for traits of interest in *Musa* (Heslop-Harrison and Schwarzacher 2007), which, while not dependent upon crosses and progeny maintenance, requires hundreds of plant accessions and thousands of genetic markers. The SSR markers designed in our work are also applicable for such a study. In general, the frequency and distribution of SSRs in ESTs and in genomic sequences differ, with dinucleotides typically more abundant in genomic survey sequences and trinucleotides more common in ESTs (e.g. La Rota et al. 2005; Varshney et al. 2005; Miller et al. 2010). In our study, trinucleotide repeat motifs (an average of 55.8 % across both EST datasets) were indeed more abundant than dinucleotide motifs (average of 28.4 %). All other motifs, from tetra- to hendeca-repeats, were only poorly represented. Such a predominance of trinucleotides probably reflects the fact that such motifs in gene regions will avoid frameshift mutations which would cause changes at the protein level. Simple sequence repeat mining criteria in software may also distort real differences in motif abundance (Varshney et al. 2005). A total of 75 out of 303 tested SSR marker primer pairs were reproducibly polymorphic when tested across *M. acuminata* accessions contrasting in resistance to Sigatoka diseases, complementing the previous work by our group (Miller et al. 2010). Similar polymorphism rates have been observed in other crop species such as wheat and cotton (Eujayl et al. 2002; Han et al. 2006). Polymerase chain reaction amplification failed, however, for 106 primer pairs. Possible reasons include SSR extension across splice sites, poor sequence quality or chimeric DNA (Varshney et al. 2005). It has been reported that EST-derived SSR markers show less polymorphism than genomic sequence-derived SSRs, as a result of conservation in gene regions (Raju et al. 2010). Indeed, from a total of 75 loci, only 289 alleles were observed, with an average of 3.8 alleles per locus and an average PIC of 0.5. Considering that a total of 303 potentially functional SSR markers were identified from a subset of 4549 ESTs in the present study, it is possible to estimate approximately a further 1000 markers that could be derived from the 15 464 publically available *M. acuminata* ESTs. Given the advent of next-generation sequencing-derived gene expression sequence data, however, this number looks set to increase considerably.

Conclusions and forward look

This study contributes considerably to publically available EST resources for *M. acuminata*, providing a unigene set of 3995 sequences derived from accessions Calcutta 4 and Cavendish Grande Naine during

incompatible and compatible interactions with *Mf*. Genes was characterized according to the KOG-based classification, Interpro-based domain identification and GO category assignment. A large set of genic-SSR markers was developed, with polymorphic markers applicable for genetic map enrichment, diversity characterization and downstream marker-assisted breeding. In summary, it is anticipated that this dataset contributes to genomic resources for *Musa*, with downstream application in genetic improvement. Ongoing next-generation sequencing-based investigation of gene expression (including transcription profiling) in *Musa*–pathogen interactions by our group will offer potential for further elucidation of gene expression during plant immune responses, and will contribute to validating annotated gene models in the *Musa* whole-genome sequencing project.

Additional information

The following additional information is available in the online version of this article –

File 1: *Musa acuminata* unigene assignment to KOG categories.

File 2: EuKaryotic orthologous group category abundance of *M. acuminata* unigenes.

File 3: Details of 303 *M. acuminata* genic-SSR markers, validated for polymorphism across 20 diploid accessions.

Accession numbers

High-quality 5' single-pass ESTs for 9333 cDNA clones from the MAC4 library and 3962 from the MACV library have been deposited in the GenBank database (<http://www.ncbi.nlm.nih.gov/dbEST/>) [accession numbers: JK531581–JK540913 (MAC4); JK542313–JK546274 (MACV)].

Sources of funding

This work was funded by the IAEA, Austria (Project 13187), FINEP, Brazil (Project 0107060900/0842/07), CNPq, Brazil (Projects 680.398/01-5 and 506165/2004-3). M.A.N.P. was supported by a fellowship from the CNPq.

Contributions by the authors

M.A.N.P. participated in EST sequencing, microsatellite marker validation and data analysis. V.O.C. participated in microsatellite marker validation and data analysis. F.L.E. participated in EST sequencing and data analysis. C.C.T. participated in EST sequencing. M.T.S.J. conceived the study and participated in EST sequencing. T.M. participated in EST sequencing. V.C.R.A. participated in microsatellite marker validation and data analysis.

C.F.F. participated in microsatellite marker validation and data analysis. E.P.A. participated in microsatellite marker validation and data analysis. L.F.A.F. participated in EST sequence data analysis. N.F.M. participated in EST sequence data analysis. M.J.B.C. participated in northern blot analyses. F.C.B. conceived the study and participated in northern blot analyses. O.S. supervised and participated in EST sequence data analysis and computational searches for microsatellite identification and primer design. G.J.P.J. supervised and participated in EST sequence data analysis and computational searches for microsatellite identification and primer design, and editing of the manuscript. L.P. conducted *in vitro* bioassays. C.A. conceived the study and conducted *in vitro* bioassays. A.Y.C. conceived the study, prepared cDNA libraries and supervised microsatellite marker validation and data analysis. P.P. conceived the study, prepared cDNA libraries, participated in northern blot analysis and edited the manuscript. R.N.G.M. conceived the study, participated in EST sequencing and data analysis, and drafted the manuscript. All authors have contributed to, read and approved the final manuscript.

Acknowledgements

R.N.G.M. acknowledges the financial support from IAEA (13187). The work from all authors was in the context of IAEA Coordinated Research Project 'Molecular Tools for Quality Improvement in Vegetatively Propagated Crops Including Banana and Cassava (D23027)' and the FAO/IAEA Joint Programme. We thank anonymous reviewers for their useful comments on the manuscript.

Conflict of interest statement

None declared.

Literature cited

- Abadie C, Zapater MF, Pignolet L, Carlier J, Mourichon X. 2008. Artificial inoculation on plants and banana leaf pieces with *Mycosphaerella* spp., responsible for Sigatoka leaf spot diseases. *Fruits* **63**: 319–323.
- Almagro L, Gomez Ros LV, Belchi-Navarro S, Bru R, Ros Barcelo A, Pedreno MA. 2009. Class III peroxidases in plant defence reactions. *Journal of Experimental Botany* **60**: 377–390.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.
- Alvarez ME, Pennell RI, Meijer PJ, Ishikawa A, Dixon RA, Lamb C. 1998. Reactive oxygen species intermediates mediate a systemic signal network in the establishment of plant immunity. *Cell* **92**: 773–784.
- Amorim EP, Lessa LS, da Silva Ledo CA, de Oliveira Amorim VB, Viana dos Reis R, Santos-Serejo JA, de Oliveira e Silva S.

2009. Caracterização agrônômica e molecular de genótipos diplóides melhorados de bananeira. *Revista Brasileira de Fruticultura* **31**: 154–161.
- Amorim EP, Amorim VBO, Silva SO, Pillay M. 2011.** Quality improvement of cultivated *Musa*. In: Pillay M, Tenkouano A, eds. *Banana breeding: progress and challenges*. New York: CRC Press, 252–280.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000.** Gene ontology: tool for the unification of biology. *Nature Genetics* **25**: 25–29.
- Azhar M, Heslop-Harrison JS. 2008.** Genomes, diversity and resistance gene analogues in *Musa* species. *Cytogenetic and Genome Research* **121**: 59–66.
- Bennetzen JL. 2000.** Transposable element contributions to plant genome evolution. *Plant Molecular Biology* **42**: 251–269.
- Borkowska M, Krzymowska M, Talarczyk A, Awan MFM, Yakovleva L, Kleczkowski K, Wielgat B. 1998.** Transgenic potato plants expressing soybean β -1,3-endoglucanase gene exhibit an increased resistance to *Phytophthora infestans*. *Zeitschrift für Naturforschung* **53**: 1012–1016.
- Boyce DC, Nam J, Dangl JL. 1998.** The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proceedings of the National Academy of Science of the USA* **95**: 15849–15854.
- Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R. 2005.** Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Molecular Ecology Notes* **5**: 327–330.
- Cavalcante M de JB, Escoute J, Madeira JP, Romero RE, Nicole MR, Oliveira LC, Hamelin C, Lartaud M, Verdeil JL. 2011.** Reactive oxygen species and cellular interactions between *Mycosphaerella fijiensis* and banana. *Tropical Plant Biology* **4**: 134–143.
- Cheung F, Town CD. 2007.** A BAC end view of the *Musa acuminata* genome. *BMC Plant Biology* **7**: 29.
- Choi D, Kim HM, Yun HK, Park JA, Kim WT, Bok SH. 1996.** Molecular cloning of a metallothionein-like gene from *Nicotiana glutinosa* L. and its induction by wounding and tobacco mosaic virus infection. *Plant Physiology* **112**: 353–359.
- Chou HH, Holmes MH. 2001.** DNA sequence quality trimming and vector removal. *Bioinformatics* **17**: 1093–1104.
- Christelová P, Valárik M, Hříbová E, Van den houwe I, Channelière S, Roux N, Doležel J. 2011.** A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB PLANTS* **2011**: plr024; doi:10.1093/aobpla/plr024.
- Churchill ACL. 2011.** *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Molecular Plant Pathology* **12**: 307–328.
- Conesa A, Götz S. 2008.** Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* **2008**: 619832.
- Creste S, Tulmann-Neto A, Figuéira A. 2001.** Detection of single sequence repeats polymorphisms in denaturing polyacrylamide sequencing gel by silver staining. *Plant Molecular Biology Reporter* **19**: 299–306.
- Creste S, Tulmann-Neto A, Silva SO, Figuéira A. 2003.** Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* **132**: 259–268.
- Creste S, Benatti TR, Orsi MR, Risterucci AM, Figuéira A. 2006.** Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. *Molecular Ecology Notes* **6**: 303–306.
- Crouch JH, Ortiz R, Crouch HK, Ford-Lloyd BV, Howell EC, Newbury HJ, Jarret RL. 2001.** Utilization of molecular genetic techniques in support of plantain and banana improvement. *Acta Horticulturae* **540**: 185–191.
- Davidson RM, Reeves PA, Manosalva PM, Leach JE. 2009.** Germins: a diverse protein family important for crop improvement. *Plant Science* **177**: 499–510.
- Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MSS, Wang L. 2002.** The phenylpropanoid pathway and plant defence—a genomics perspective. *Molecular Plant Pathology* **3**: 371–390.
- Dochez C, Tenkouano A, Ortiz R, Whyte J, De Waele D. 2009.** Host plant resistance to *Radopholus similis* in a diploid banana hybrid population. *Nematology* **11**: 329–335.
- Elvira MI, Galdeano MM, Gilardi P, García-Luque I, Serra MT. 2008.** Proteomic analysis of pathogenesis-related proteins (PRs) induced by compatible and incompatible interactions of pepper mild mottle virus (PMMoV) in *Capsicum chinense* L3 plants. *Journal of Experimental Botany* **59**: 1253–1265.
- Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W. 2002.** Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics* **104**: 399–407.
- Ewing B, Green P. 1998.** Base-calling of automated sequencer traces using phred II. Error probabilities. *Genome Research* **8**: 186–194.
- FAOSTAT. 2009.** *Food and agricultural commodities production*. Food and Agricultural Organization Statistics. <http://faostat.fao.org>.
- Feschotte C, Jiang N, Wessler SR. 2002.** Plant transposable elements: where genetics meets genomics. *Nature Genetics* **3**: 329–341.
- Gawel NJ, Jarret RL. 1991.** A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Plant Molecular Biology Reporter* **9**: 262–266.
- The Gene Ontology Consortium. 2008.** The Gene Ontology project in 2008. *Nucleic Acids Research* **36**: D440–444.
- Govrin E, Levine A. 2000.** The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. *Current Biology* **10**: 751–757.
- Grover A, Gowthaman R. 2003.** Strategies for development of fungus-resistant transgenic plants. *Current Science* **84**: 330–240.
- Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. 2003.** Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* **270**: 315–323.
- Hamer DH. 1986.** Metallothionein. *Annual Reviews of Biochemistry* **55**: 913–951.
- Hammond-Kosack KE, Jones JDG. 1997.** Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**: 573–605.
- Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T. 2006.** Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theoretical and Applied Genetics* **112**: 430–439.

- Hayes JD, McLellan LI. 1999. Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. *Free Radical Research* 31: 273–300.
- Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. *Annals of Botany* 100: 1073–1084.
- Hippolyte I, Bakry F, Seguin M, Gardes L, Rivallan R, Risterucci AM, Jenny C, Perrier X, Carreel F, Argout X, Piffanelli P, Khan I, Miller RNG, Pappas GJ, Mbeguie-a-mbeguie D, Matsumoto T, De Bernardinis V, Huttner E, Kilian A, Baurens FC, D'hont A, Cote F, Courtois B, Glaszmann JC. 2010. A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biology* 10: 65.
- Hribová E, Neumann P, Matsumoto T, Roux N, Macas J, Dolezel J. 2010. Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biology* 10: 204.
- Hsieh HM, Liu WK, Huang PC. 1995. A novel stress-inducible metallothionein-like gene from rice. *Plant Molecular Biology* 28: 381–389.
- Jones DR. 1999. *Diseases of banana, abacá and enset*, 1st edn. Wallingford, Oxon: CABI Publishing.
- Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444: 323–329.
- Kaemmer D, Fisher D, Jarret RL, Baurens FC, Grapin A, Dambier D, Noyer JL, Lanaud C, Kahl G, Lagoda PJJ. 1997. Molecular breeding in genus *Musa*: a strong case for SSR marker technology. *Euphytica* 96: 49–62.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research* 32: D277–D280.
- Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A. 2001. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas* 135: 145–151.
- La Rota M, Kantety RV, Yu JK, Sorrells ME. 2005. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* 6: 23.
- Lagoda PJJ, Noyer JL, Dambier D, Baurens FC, Grapin A, Lanaud C. 1998. Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molecular Ecology* 7: 657–666.
- Lescot M, Piffanelli P, Ciampi AY, Ruiz M, Blanc G, Leebens-Mack J, da Silva FR, Santos CM, D'Hont A, Garsmeur O, Vilarinhos AD, Kanamori H, Matsumoto T, Ronning CM, Cheung F, Haas BJ, Althoff R, Arbogast T, Hine E, Pappas GJ Jr, Sasaki T, Souza MT Jr, Miller RNG, Glaszmann JC, Town CD. 2008. Insights into the *Musa* genome: syntenic relationships to rice and between *Musa* species. *BMC Genomics* 9: 58.
- Liu P, Goh C, Loh C, Pua E. 2002. Differential expression and characterization of three metallothionein-like genes in Cavendish banana (*Musa acuminata*). *Physiologia Plantarum* 114: 241–250.
- Lorenzen J, Hearne S, Mbahjo G, Nyine M, Close T. 2011. Use of molecular markers in banana and plantain improvement. *Acta Horticulturae* 897: 231–236.
- Loyall L, Uchida K, Braun S, Furuya M, Frohnmeyer H. 2000. Glutathione and a UV light-induced glutathione S transferase are involved in signaling to chalcone synthase in cell cultures. *The Plant Cell* 12: 1939–1950.
- Mackey D, Holt BF III, Wiig A, Dangl JL. 2002. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108: 743–754.
- Mackey D, Belkhadir Y, Alonso JM, Ecker JR, Dangl JL. 2003. *Arabidopsis* RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. *Cell* 112: 379–389.
- Manrique-Trujillo SM, Ramírez-López AC, Ibarra-Laclette E, Gómez-Lim MA. 2007. Identification of genes differentially expressed during ripening of banana. *Journal of Plant Physiology* 164: 1037–1050.
- Martin GB, Brommonshenkel SH, Chungwongse J, Frary A, Ganai MW. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* 262: 1432–1436.
- Mauch F, Dudler R. 1993. Differential induction of distinct glutathione S-transferases of wheat by xenobiotics and by pathogen attack. *Plant Physiology* 102: 1193–1201.
- McHale L, Tan X, Koehl P, Michelmore RW. 2006. Plant NBS-LRR proteins: adaptable guards. *Genome Biology* 7: 212.1–212.11.
- Miller RNG, Bertoli DJ, Baurens FC, Santos CM, Alves PC, Martins NF, Togawa RC, Souza MT Jr, Pappas GJ Jr. 2008. Analysis of non-TIR NBS-LRR resistance gene analogs in *Musa acuminata* Colla: isolation, RFLP marker development, and physical mapping. *BMC Plant Biology* 8: 15.
- Miller RNG, Passos MA, Menezes NN, Souza MT Jr, do Carmo Costa MM, Rennó Azevedo VC, Amorim EP, Pappas GJ Jr, Ciampi AY. 2010. Characterization of novel microsatellite markers in *Musa acuminata* subsp. *burmannicoides*, var. *Calcutta 4*. *BMC Research Notes* 3: 148.
- Miller RNG, Passos MAN, Emediato FL, de Camargo Teixeira C, Pappas Júnior GJ. 2011. Candidate resistance gene discovery: resistance gene analog characterization and differential gene expression analysis in *Musa-Mycosphaerella* host-pathogen interactions. *Acta Horticulturae* 897: 179–186.
- Mohamad A, Heslop-Harrison JS. 2008. Genomes, diversity and resistance gene analogs in *Musa* species. *Cytogenetic and Genome Research*. 121: 59–66.
- Naoumkina MA, Zhao Q, Gallego-Giraldo L, Dai X, Zhao PX, Dixon RA. 2010. Genome-wide analysis of phenylpropanoid defence pathways. *Molecular Plant Pathology* 11: 829–846.
- Nürnberg T, Kemmerling B. 2009. PAMP-triggered basal immunity in plants. *Advances in Botanical Research* 51: 1–38.
- Ortiz R. 1997. Secondary polyploids, heterosis, and evolutionary crop breeding for further improvement of the plantain and banana (*Musa* spp. L.) genome. *Theoretical and Applied Genetics* 94: 1113–1120.
- Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. 2003. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19: 651–652.
- Portal O, Izquierdo Y, De Vleeschauwer D, Sánchez-Rodríguez A, Mendoza-Rodríguez M, Acosta-Suárez M, Ocaña B, Jiménez E, Höfte M. 2011. Analysis of expressed sequence tags derived from a compatible *Mycosphaerella fijiensis*-banana interaction. *Plant Cell Reports* 30: 913–928.
- Raju NL, Gnanesh BN, Lekha P, Jayashree B, Pande S, Hiremath PJ, Byregowda M, Singh NK, Varshney RK. 2010. The first set of EST resource for gene discovery and marker development in pigeon-pea (*Cajanus cajan* L.). *BMC Plant Biology* 10: 45.

- Requena L, Bornemann S. 1999.** Barley (*Hordeum vulgare*) oxalate oxidase is a manganese-containing enzyme. *Biochemical Journal* **343**: 185–190.
- Robinson NJ, Tommey AM, Kuske C, Jackson PJ. 1993.** Plant metallothioneins. *Biochemical Journal* **295**: 1–10.
- Roux N, Baurens FC, Dolezel J, Hribova E, Heslop-Harrison P, Town C, Sasaki T, Matsumoto T, Aert R, Remy S, Souza M, Lagoda P. 2008.** Genomics of banana and plantain (*Musa* spp.), major staple crops in the tropics. In: Moore PH, Ming R, eds. *Genomics of tropical crop plants*. New York: Springer, 83–111.
- Rozen S, Skaletsky HJ. 2000.** PRIMER 3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, eds. *Bioinformatics methods and protocols; methods in molecular biology*. Totowa, NJ: Humana Press, 365–386.
- Santos CMR, Martins NF, Horberg HM, de Almeida ERP, Coelho MCF, Togawa RC, da Silva FR, Caetano AR, Miller RNG, Souza MT Jr. 2005.** Analysis of expressed sequence tags from *Musa acuminata* ssp. *burmannicoides*, var. Calcutta 4 (AA) leaves submitted to temperature stresses. *Theoretical and Applied Genetics* **110**: 1517–1522.
- Shiu SH, Bleecker AB. 2001.** Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Science of the USA* **98**: 10763–10768.
- Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C. 1995.** A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* **270**: 1804–1806.
- Spannagl M, Noubibou O, Haase D, Yang L, Gundlach H, Hindemitt T, Klee K, Haberer G, Schoof H, Mayer KF. 2007.** MIPSPlantsDB—plant database resource for integrative and comparative plant genome research. *Nucleic Acids Research* **35**: D834–D840.
- Tamana A, Khan AU. 2005.** Mapping and analysis of simple sequence repeats in the *Arabidopsis thaliana* genome. *Bioinformation* **1**: 64–68.
- Tsuda K, Katagiri F. 2010.** Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Current Opinion in Plant Biology* **13**: 459–465.
- Website 1.** The SSAHA package. <http://www.sanger.ac.uk/Software/analysis/SSAHA/> (1 June 2011).
- Website 2.** NCBI Non-Redundant sequence database. <http://www.ncbi.nlm.nih.gov/COG/> (1 June 2011).
- Website 3.** The Swiss-Prot Database. <http://www.uniprot.org/downloads>, uniprot_sprot_release of 2010 04 23 (1 June 2011).
- Website 4.** The TAIR Database: The Arabidopsis Information Resource. <http://www.arabidopsis.org/>, Tair_9_pep_release 2009 06 19 (1 June 2011).
- Website 5.** Mreps. <http://bioinfo.lifl.fr/mreps/> (1 June 2011).
- Website 6.** Polymorphism information content (PIC) calculator. <http://www.liv.ac.uk/~kempsj/pic.html> (1 June 2011).
- Van den Berg N, Crampton BG, Hein I, Birch PRJ, Berger DK. 2004.** High-throughput screening of suppression subtractive hybridisation cDNA libraries using DNA microarray analysis. *Biotechniques* **37**: 818–824.
- Van Loon LC, Rep M, Pieterse CMJ. 2006.** Significance of inducible defense related proteins in infected plants. *Annual Review of Phytopathology* **44**: 135–162.
- Varshney RK, Graner A, Sorrells ME. 2005.** Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48–55.
- Wei Y, Zhang Z, Andersen CH, Schmelzer E, Gregersen P, Collinge DB, Smedegaard-Petersen V, Thordal-Christensen H. 1998.** An epidermis/papilla-specific oxalate oxidase-like protein in the defence response of barley attacked by the powdery mildew fungus. *Plant Molecular Biology* **36**: 101–112.
- Wróbel-Kwiatkowska M, Lorenc-Kukula K, Starzycki M, Oszmianski J, Kepczynska E, Szopa J. 2004.** Expression of β -1,3-glucanase in flax causes increased resistance to fungi. *Physiological and Molecular Plant Pathology* **65**: 245–256.
- Wu JL, Sinha PK, Variar M, Zheng KL, Leach JE, Courtois B, Leung H. 2004.** Association between molecular markers and blast resistance in an advanced backcross population of rice. *Theoretical and Applied Genetics* **108**: 1024–1032.
- Zhou F, Zhang Z, Gregersen L, Mikkelsen J, de Neergaard E, Collinge D, Thordal-Christensen H. 1998.** Molecular characterization of the oxalate oxidase involved in the response of barley to the powdery mildew fungus. *Plant Physiology* **117**: 33–41.

Candidate Resistance Gene Discovery: Resistance Gene Analog Characterisation and Differential Gene Expression Analysis in *Musa-Mycosphaerella* Host-Pathogen Interactions

R.N.G. Miller^{1,2}, M.A.N. Passos¹, F.L. Emediato¹, C. de Camargo Teixeira¹ and G.J. Pappas Júnior^{1,3}

¹Universidade Católica de Brasília, Brasília, DF, Brazil

²Universidade de Brasília, Departamento de Biologia Celular, Brasília, DF, Brazil

³EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil

Keywords: black leaf streak, black Sigatoka, EST, plant disease resistance genes, RGA

Abstract

Many banana cultivars (*Musa* spp.) are sterile triploids or diploids, evolving only via somatic mutation. As a consequence, this crop can lack resistance to pests and diseases. Numerous disease resistance genes (*R*-genes) have been characterised in plants, conferring resistance to bacteria, viruses, fungi and nematodes. Identification and cloning of *R*-genes in *Musa* will provide new opportunities for genetic improvement. Our group has identified over 50 distinct NBS-LRR-type resistance gene analogs (RGAs) in the resistant wild diploid *Musa acuminata* ssp. *burmannicoides* 'Calcutta 4'. Characterisation is ongoing in *M. acuminata* cultivars contrasting in resistance to *Mycosphaerella* leaf spot diseases, focusing on both the NBS-LRR *R*-gene family and cytoplasmic receptor-like kinases (RLKs) with extracellular LRRs. NBS-LRR class RGA probes applied to 'Calcutta 4', 'Grande Naine' (AAA, Cavendish subgroup), and *M. balbisiana* 'Pisang Klutuk Wulung' BAC libraries have revealed many putative resistance loci. Sequence data for such selected clones will provide insight into organisation and evolution of this *R*-gene class in *Musa*. Candidate gene discovery is also ongoing via analysis of differential gene expression from infected leaf cDNA during *Musa-Mycosphaerella* interactions. Candidate *R*-genes will be applicable for banana genetic improvement via both plant transformation and conventional breeding using marker-assisted selection.

INTRODUCTION

Commercial banana cultivars are grown across 130 countries in different tropical and subtropical environments. Conventional breeding strategies for this crop are however hindered by the fact that many cultivars are sterile and do not produce seeds. Asexually-driven evolution via vegetative micropropagation or suckers has consequently resulted in a restricted genetic base, lacking resistance to pests and disease. The industry has witnessed numerous pathogen and pest outbreaks as a result. Nonconventional breeding strategies, such as transformation and marker-assisted selection, offer an alternative approach for introgression of resistance genes in commercial cultivars. However, given the current limited number of studies conducted at the molecular level for *Musa*-pathogen interactions, only very few banana genes have been isolated and characterised to date.

Gene-for-gene race-specific disease resistance in many plant-pathogen interactions appears to result from the activation of dominant, allele-specific constitutive disease resistance (*R*) gene protein products, or receptors, by dominant allele-specific pathogen avirulence (*Avr*) gene products, or elicitors. Such resistance is often associated with rapid, localised cell death or the hypersensitive response (HR). Believed to be abundant across plant species (Michelmore and Meyers, 1998), numerous *R*-genes from model plants and crop species have now been characterised (Martin et al., 2003). Despite the broad range of recognised pathogens, *R*-genes share significant homologies in amino-acid sequences and structural protein motifs, suggesting common mechanisms controlling protein-protein interactions. A number of distinct *R*-gene structural classes have been identified based upon conserved protein domains, the most prevalent class of which encode proteins with cytoplasmic nucleotide binding (NBS) and leucine-rich repeat (LRR) domains. Other

common *R*-gene classes include extracellular LRRs anchored by transmembrane domains (receptor-like proteins), extracellular LRRs linked to cytoplasmic serine-threonine kinase domains, intracellular serine-threonine kinases, and proteins with a coiled-coil domain anchored to the cell membrane. Given the conservation across *R*-gene-coding protein motifs, PCR-based conserved orthologous sequence (COS) marker strategies with degenerate primers can facilitate isolation of resistance gene analogs (RGAs). Such work has recently been reported in *Musa*, with large-scale analyses of NBS-LRR *R*-gene family RGA diversity across the genus (e.g. Miller et al., 2008; Mohamad and Heslop-Harrison, 2008).

Expressed sequence tags (ESTs) are sequenced portions of complementary DNA copies of mRNA, representing part of the transcribed region of the genome under given conditions. EST characterisation has revealed that model plants such as *Arabidopsis* can express hundreds of potential *R*-genes (Botella et al., 1997). In *Musa*, however, only a limited number of such studies have analysed gene expression during host-pathogen interactions (e.g. van den Berg et al., 2002; Lim et al., 2004).

This study describes ongoing work for candidate resistance gene discovery in the *Musa-Mycosphaerella* interaction. Comparative transcriptome analysis is presented for resistant and susceptible *Musa acuminata* cultivars, from leaf material in vitro-infected with *Mycosphaerella fijiensis*. Continued RGA characterisation is also presented, targeting two common *R*-gene structural classes, across *M. acuminata* clones contrasting in resistance to black leaf streak disease (BLSD), also known as black Sigatoka.

MATERIALS AND METHODS

RGA Characterisation

Genomic DNA was extracted from *M. acuminata* ssp. *burmannicoides* ‘Calcutta 4’ and AA cultivar ‘Lidi’ (resistant to BLSD), and AA cultivars ‘Pisang Berlin’ and ‘Niyarma Yik’ (susceptible to BLSD), using a standard CTAB approach (Rogers and Bendich, 1988).

Degenerate primers were designed according to Miller et al. (2008), via a selection of monocot-specific Genbank-derived sequences containing characteristic *R*-gene class conserved domains. Protein sequences were aligned using the program MUSCLE (Edgar, 2004), and subgroups identified using CLUSS (Abdellali et al., 2007). Degenerate primers were designed using CODEHOP (Rose et al., 1998), targeting *R*-gene cytoplasmic serine-threonine (Ser/Thr) receptor-like protein kinases (RLKs) with extracellular LRRs.

Together with primers previously designed for NBS-LRR RGAs (Penuela et al., 2002; Bertioli et al., 2003; Miller et al., 2008), PCR products were amplified, purified using a Qiagen QIAquick PCR purification kit, cloned using pGEM-T-Easy, DH5a *Escherichia coli* cells transformed by electroporation, and plasmid DNA extracted by a standard alkaline-lysis procedure. Single-pass sequencing of clones was conducted on an ABI 377 sequencer. Sequences were cleaned of vectors using the Staden package (Staden, 1996), and contig assembly performed using CAP3 (Huang and Madan, 1999). RGAs were identified on the basis of sequence similarity via the program BLAST (Altschul et al., 1997), together with protein domain analysis using the program GENEWISE (Birney et al., 2004).

Transcriptome Analysis

Analysis of differential gene expression was conducted on contrasting ‘Calcutta 4’ and AAA Cavendish-type ‘Grande Naine’ leaf material inoculated in vitro with *M. fijiensis*. Unidirectional 5’ sequencing was conducted on approximately 7700 randomly selected clones per library, with data deposited in the *Musa* genome database DataMusa (http://genoma.embrapa.br/musa/index.html/DATA_musa.html). Following EST processing for sequence assembly and annotation, differential gene expression between the contrasting cultivars was analysed using three statistical tests to provide *P*-values for

the null hypothesis that there is no differential expression based upon counts between the two libraries, namely Stekel test (Stekel et al., 2000), Fisher exact test and Audic-Claverie test (Claverie, 1999). A computational search for simple sequence repeats (SSRs) was performed on EST consensi using the program mreps (Kolparov et al., 2003).

RESULTS AND DISCUSSION

RGA Characterisation

A total of 756 novel sequences with significant sequence similarity to RGAs were amplified across the genotypes contrasting in resistance to BLSD. Primers targeting NBS-LRR *R*-genes generated 194 sequences with similarity to RGAs from this class (Table 1). Clustered into 31 distinct contigs, a total of 19 displayed best Blast hits with members of the genus *Musa*. The overall percentage of clones displaying similarity to NBS-LRR RGAs varied between the three primer combinations, with greater numbers observed using primer combinations exclusively targeting NBS motifs, rather than a combination spanning NBS and LRR motifs. Differentiation of RGA contigs on the basis of genotype origin was not observed, with positive contig-member clones originating from both BLSD-resistant and -susceptible genotypes. Sequence similarity analysis of amplification products generated using degenerate primers for protein-kinase *R*-genes (RLKs) identified 562 sequences with significant similarity to *R*-gene and RGA sequences for this class (Table 2). Out of a total of 66 contigs, 26 showed best Blast hits to *Musa* protein kinase-like proteins. High percentages of sequenced clones displaying similarity to this *R*-gene class were observed across all primer combination products tested. As with NBS-LRR targeting primers, no differentiation of protein-kinase RGA contigs could be made on the basis of genotype and resistance to BLSD.

Analysis of RGA-Containing BAC Clones

Using two NBS-LRR RGA probes, a total of 86 unique NBS-LRR class RGA-containing BAC clones have been identified to date across BAC libraries representing the 'Calcutta 4' and 'Grande Naine' genomes, together with the *Musa balbisiana* 'Pisang Klutuk Wulung' genome (Miller et al., 2008). Clustering of multi-copy *R*-genes together with RGAs is commonly observed across plant genomes. Our work also reflected this observation, with multiple copies frequent in positive BACs. Considered to be a result of tandem duplications of paralogs (Meyers et al., 2003), within such clusters different *R*-genes have been hypothesised to confer resistance to different strains of a pathogen or to diverse pathogens (van der Vossen et al., 2000). Such clustered RGAs may also facilitate genetic variation for evolution of new *R*-genes (Michelmore and Meyers, 1998). In relation to *R*-gene cloning, although BAC clone isolation is facilitated by the presence of multiple RGAs at an *R*-gene locus, distinction between a candidate *R*-gene and non-functional sequences is complex, with implications in terms of specific genetic marker development and *R*-gene mapping (Xiao et al., 2007).

Given that 33 contiguous *Musa* NBS-LRR RGAs were identified in our first study (Miller et al., 2008), with further NBS-LRR and RLKs characterised in this current work, continued application in BAC identification will provide a useful approach for general identification and characterisation of putative resistance loci across *Musa* genomes.

Transcriptome Analysis

To better understand the molecular basis of the resistance and defence response against the BLSD pathogen, an EST approach was used to identify genes differentially expressed during infection stages in 'Calcutta 4' and 'Grande Naine' contrasting in resistance. In-silico analysis revealed statistically significant differential expression in 220 genes. Amongst these, a total of 24 genes were identified as related to resistance or defence, including type-3 metallothioneins, germin-like proteins, ferredoxin and glutathione S-transferase, pathogenesis-related proteins, thioredoxin, glycolate oxidase, putative ethylene-responsive small GTP-binding protein, superoxide dismutase (SOD),

putative stress enhanced protein, and an auxin-repressed protein-like protein (arp1).

Next Generation Sequencing approaches are ongoing to increase the resolution of differential gene expression analysis in this host-pathogen interaction and to enable large-scale genetic marker development.

Confirming Function of Candidate Genes and Prospects for Durable Resistance

Disease-resistance genes in plants share significant homologies in amino-acid sequences and structural motifs, suggesting common protein-protein interactions as components of receptor systems and common roles in signal transduction signalling events in plant defence systems. Although numerous *R*-genes and RGAs have now been cloned, determination of activity and specificity against a given pathogen remains a bottleneck for development of durable resistance in important crop species (Michelmore, 2003). Recent advances, however, in map-based cloning approaches that involve RGA mapping and BAC sequencing, indicate that together with optimisation of transformation protocols, transfer of *R*-gene candidates to elite cultivars is now becoming an available technology to complement breeding programmes.

Gene Pyramids

Crop monocultures with single *R*-genes typically provide strong selection pressure for mutation of corresponding *Avr* genes, where a single loss-of-function mutation results in only ephemeral interactions of *R*-genes with *Avr* elicitors, with subsequent loss of resistance in the plant (Pink, 2002). Such mutation in *Avr* genes is likely to occur with no fitness cost incurred in the pathogen because single *Avr* genes probably make small contributions to virulence individually. Incorporation of numerous *R*-genes (or pyramids) into plant cultivars is likely to result in a more durable resistance, as mutation in multiple *Avr* genes would likely result in a cumulative negative effect on pathogen virulence (McDowell and Woffenden, 2003). Success of such an approach has been observed in rice, with the introduction of four *R*-genes conferring resistance to *Xanthomonas oryzae* pv. *oryzae* (Li et al., 2001). However, suitable promoters are critical, as ectopic expression of pyramids of *R*-genes has been reported to activate defence in the absence of pathogen elicitors, which can reduce fitness and ultimately crop yield (Mindrinis et al., 1994). Planting of multiple plant lines, each with a different *R*-gene, as well as simultaneous limited planting of a susceptible line, can also reduce selection pressure for *Avr* gene mutation (Mundt, 2002).

Restricted Taxonomic Functionality

In any attempt to utilise *R*-genes between distant plant taxa, restricted taxonomic functionality (RTF) must be considered (Tai et al., 1999). Altered or loss of function of a particular gene when expressed in a different plant host can occur (e.g. Tai et al., 1999), perhaps reflecting inabilities in interaction with signal transduction mechanisms present in each host. If we consider the guard hypothesis for *R*-protein and *Avr* protein interactions, RTF may reflect absence or incompatible guard proteins. As our understanding of signalling mechanisms in resistance and defence response components continues to be broadened, such incompatibility may be overcome.

CONCLUSIONS

Candidate resistance and defence gene discovery is ongoing via transcriptome analysis during *Musa-Mycosphaerella* interactions, COS marker approaches for RGA discovery, and whole BAC shotgun sequencing. In addition to increasing our understanding of the molecular processes involved in disease resistance in *Musa*, candidate genes offer potential for the development of effective disease management based upon genetic improvement via both plant transformation and conventional breeding using marker-assisted selection.

ACKNOWLEDGEMENTS

This work was funded by the IAEA (Project 13187), FINEP (Project 0107060900/0842/07), CNPq (Projects 680.398/01-5 and 506165/2004-3), the Generation Challenge Program (GCP-SP2-project #15), Embrapa and the Universidade Católica de Brasília. MANP was supported by a fellowship from the CNPq.

Literature Cited

- Abdellali, K., Wang, S., Brzezinski, R. and Fleury, A. 2007. CLUSS: Clustering of protein sequences based on a new similarity measure. *BMC Bioinformatics* 8:286.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
- Bertioli, D.J., Leal-Bertioli, S.C., Lion, M.B., Santos, V.L., Pappas, G., Jr., Cannon, S.B. and Guimaraes, P.M. 2003. A large scale analysis of resistance gene homologues in *Arachis*. *Mol. Genet. Genomics* 270:34-45.
- Birney, E., Clamp, M. and Durbin, R. 2004. GeneWise and Genomewise. *Genome Research* 14:988-995.
- Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T., Jones, J.D. and Somerville, S.C. 1997. Map positions of 47 Arabidopsis sequences with sequence similarity to disease resistance genes. *Plant J.* 12:1197-1211.
- Claverie, J.M. 1999. Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.* 8:1821-1832.
- Huang, X. and Madan, A. 1999. CAP3: a DNA sequence assembly program. *Genome Res.* 9:868-877.
- Kolparov, R., Bana, G. and Kucherov, G. 2003. mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res.* 31:3672-3678.
- Li, Z.K., Sanchez, A., Angeles, E., Singh, S., Domingo, J., Huang, N. and Khush, G.S. 2001. Are the dominant and recessive plant disease resistance genes similar? A case study of rice R genes and *Xanthomonas oryzae* pv. *oryzae* races. *Genetics* 159:757-765.
- Lim, K.J., Tan, S.H., Othman, Y.R., Suhami, N., Raha, A.R., Gurmit, S. and Ho, Y.W. 2004. Microarray analysis of gene expression using plants infected with *Fusarium oxysporum* f.sp. *cubense*. p.31. In: 1st International Congress on *Musa*. Harnessing Research to Improve Livelihoods, Penang, Malaysia, 6-9 July 2004.
- Martin, G.B., Bogdanove, A.J. and Sessa, G. 2003. Understanding the functions of plant disease resistance proteins. *Annu. Rev. Plant Biol.* 54:23-61.
- McDowell, J.M. and Woffenden, B.J. 2003. Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol.* 21:178-183.
- Meyers, B.C., Kozik, A., Griego, A., Kuang, H. and Michelmore, R.W. 2003. Genome-wide analysis of NBS-LRR encoding genes in Arabidopsis. *Plant Cell* 15:809-834.
- Michelmore, R.W. 2003. The impact zone: genomics and breeding for durable disease resistance. *Curr. Opin. Plant Biol.* 6:397-404.
- Michelmore, R.W. and Meyers, B.C. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8:1113-30.
- Miller, R.N.G., Bertioli, D.J., Baurens, F.C., Santos, C.M., Alves, P.C., Martins, N.F., Togawa, R.C., Souza, M.T. Junior and Pappas, G.J. Junior. 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.
- Mindrinos, M., Katagiri, F., Yu, G.L. and Ausubel, F.M. 1994. The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* 78:1089-1099.
- Mohamad, A. and Heslop-Harrison, J.S. 2008. Genomes, diversity and resistance gene analogs in *Musa* species. *Cytogenet. Genome Res.* 121:59-66.
- Mundt, C.C. 2002. Use of multiline cultivars and cultivar mixtures for disease management. *Annu. Rev. Phytopathol.* 40:381-410.

- Penuela, S., Danesh, D. and Young, N.D. 2002. Targeted isolation, sequence analysis, and physical mapping of nonTIR NBS-LRR genes in soybean. *Theor. Appl. Genet.* 104:261-272.
- Pink, D.A.C. 2002. Strategies using genes for non-durable resistance. *Euphytica* 1:227-236.
- Rogers, S.O. and Bendich, A.J. 1988. Extraction of DNA from plant tissues. p.A6:1-10. In: S. Gelvin and R.A. Schilperoot (eds.), *Plant Molecular Biology Manual*. Kluwer Academic Press, Boston, MA.
- Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M. and Henikoff, S. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences, *Nucleic Acids Res.* 26:1628-1635.
- Staden, R. 1996. The Staden sequence analysis package. *Mol. Biotechnol.* 5:233-241.
- Stekel, D.J., Git, Y. and Falciani, F. 2000. The comparison of gene expression from multiple cDNA libraries. *Genome Res.* 10:2055-2061.
- Tai, T.H., Dahlbeck, D., Clark, E.T., Gajiwala, P., Pasion, R., Whalen, M.C., Stall, R.E. and Staskawicz, B.J. 1999. Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. *P. Natl. Acad. Sci. USA.* 96:14153-14158.
- van den Berg, N., Birch, P.R.J., Viljoen, A., Wingfield, M.J., Visser, M. and Berger, D.K. 2002. Suppression subtraction hybridization used to identify genes associated with resistance to *Fusarium oxysporum* f.sp. *cubense* in Cavendish bananas. In: 3rd International Symposium on Molecular and Cellular Biology of Bananas, Leuven, Belgium, 9-11 September 2002.
- 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.
- Xiao, W., Zhao, J., Fan, S., Li, L., Dai, J. and Xu, M. 2007. Mapping of genome-wide resistance gene analogs (RGAs) in maize (*Zea mays* L.). *Theor. Appl. Genet.* 115:501-508.

Tables

Table 1. Primer combinations, target motifs and NBS-LRR RGA isolation in *Musa acuminata*.

Degenerate primer combinations	Target conserved motifs	Target domains	Total number of insert-containing plasmids-producing high-quality sequences	Total number of contigs	Total number of contigs with homology to <i>R</i> -genes or RGAs	Total number and percentage of sequences with homology to <i>R</i> -genes or RGAs
1. P1B- RNBSD	P-loop and RNBS-D non-TIR	TIR and non-TIR NBS	92	11	6	83 (90.2%)
2. 1F-P3B	P-loop and GLPL	non-TIR NBS	109	32	9	53 (48.6%)
3. 3F2-13R1	Kinase 2 and LRR	non-TIR NBS-LRR	213	41	16	58 (27.2%)

Table 2. Primer combinations, target motifs and protein-kinase RGA isolation in *Musa acuminata*.

Degenerate primer combinations	Target domains	Total number of insert-containing plasmids-producing high-quality sequences	Total number of contigs	Total number of contigs with homology to <i>R</i> -genes or RGAs	Total number and percentage of sequences with homology to <i>R</i> -genes or RGAs
RLK_S1_K_F-RLK_S1_ID_R	Protein kinase and inter-domain	184	27	16	137 (74.5%)
RLK_S3_K_F-RLK_S3_ID_R	Protein kinase and inter-domain	178	41	19	119 (66.9%)
RLK_S4_K1_F-RLK_S4_ID_R	Protein kinase and inter-domain	140	17	6	124 (88.6%)
RLK_S4_K2_F-RLK_S4_ID2_R (Ia)	Protein kinase and inter-domain	150	37	19	125 (83.3%)
RLK_S4_K2_F-RLK_S4_ID2_R (Ib)	Protein kinase and inter-domain	100	16	6	57 (57.0%)

1: MAN; Cruz, VO; Emediato, FL; Teixeira CC, Azevedo, VCR; Brasileiro, ACM; Amorim, EP; Ferreira, CF; Martins, NF; Togawa, RC; Pappas, GJ; Silva, OB; and Miller, RNG. 2013. Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. BMC Genomics 14: 78. doi:10.1186/1471-2164-14-78

Passos et al. *BMC Genomics* 2013, **14**:78
<http://www.biomedcentral.com/1471-2164/14/78>



RESEARCH ARTICLE

Open Access

Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development

Marco A N Passos¹, Viviane Oliveira de Cruz¹, Flavia L Emediato¹, Cristiane Camargo de Teixeira², Vânia C Rennó Azevedo³, Ana C M Brasileiro³, Edson P Amorim⁴, Claudia F Ferreira⁴, Natalia F Martins³, Roberto C Togawa³, Georgios J Pappas Júnior¹, Orzenil Bonfim da Silva Jr³ and Robert NG Miller^{1*}

2: Emediato, F.L., Passos, M.A.N., de Camargo Teixeira, C., Pappas Jr., G.J. and Miller, R.N.G. 2013. Analysis of expression of NBS-LRR resistance gene analogs in *Musa acuminata* during compatible and incompatible interactions with *Mycosphaerella musicola*. Acta Hort. (ISHS) 986:255-258

Analysis of Expression of NBS-LRR Resistance Gene Analogs in *Musa acuminata* during Compatible and Incompatible Interactions with *Mycosphaerella musicola*

F.L. Emediato¹, M.A.N. Passos¹, C. de Camargo Teixeira²,
G.J. Pappas Jr.^{2,3} and R.N.G. Miller¹

¹ Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Departamento de Biologia Celular, Asa Norte, CEP 70.910-900, Brasília, DF, Brazil

² Universidade Católica de Brasília, SGAN 916, Módulo B, CEP 70.790-160, Brasília, DF, Brazil

³ EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, DF, Brazil

Keywords: banana improvement, disease resistance mechanisms, gene expression, RGA, Sigatoka leaf spot

3: M.A.N; Emediato, F. L; Cruz, V. O; Teixeira, CC; Figueiredo, L. F. de A; Martins, N. F; Togawa, R. C; Costa, M. M. C; Silva Jr, O; Pappas Jr. G. J.; Miller, R. N. G. 2013. Understanding Plant Immunity: Transcriptome Profiling in *Musa*-Pathogen Interactions using Next Generation Sequencing. Acta Hort. 986: 227-240..

Understanding Plant Immunity: Transcriptome Profiling in *Musa*-Pathogen Interactions Using Next Generation Sequencing

M.A.N. Passos¹, F.L. Emediato¹, V.O. Cruz¹, C. de Camargo Teixeira², L.F. de Alencar Figueiredo¹, N.F. Martins³, R.C. Togawa³, M.M.C. Costa³, O. Silva Jr.³, G.J. Pappas Jr.^{2,3} and R.N.G. Miller¹

¹ Universidade de Brasília, Campus Universitário Darcy Ribeiro, Instituto de Ciências Biológicas, Departamento de Biologia Celular, Asa Norte, CEP 70.910-900, Brasília, DF, Brazil

² Universidade Católica de Brasília, SGAN 916, Módulo B, CEP 70.790-160, Brasília, DF, Brazil

³ EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, DF, Brazil

Keywords: *Musa acuminata*, *Mycosphaerella musicola*, 454 transcriptome sequencing

4: Passos MAN, de Oliveira Cruz V, Emediato FL, de Camargo Teixeira C, Souza MT Jr, Matsumoto T, *et al.* Development of expressed sequence tag and expressed sequence tag-simple sequence repeat marker resources for *Musa acuminata*. *AoB Plants*. 2012;2012:pls030.



AoB PLANTS

<http://aobplants.oxfordjournals.org/>

Open access – Research article

**THIS ARTICLE IS PART OF A SPECIAL ISSUE ENTITLED
'MOLECULAR TECHNOLOGIES TO IMPROVE
VEGETATIVELY PROPAGATED BANANA AND CASSAVA'**

Development of expressed sequence tag and expressed sequence tag-simple sequence repeat marker resources for *Musa acuminata*

Marco A. N. Passos¹, Viviane de Oliveira Cruz¹, Flavia L. Emediato¹, Cristiane de Camargo Teixeira², Manoel T. Souza Jr³, Takashi Matsumoto⁴, Vânia C. Rennó Azevedo³, Claudia F. Ferreira⁵, Edson P. Amorim⁵, Lucio Flavio de Alencar Figueiredo¹, Natalia F. Martins³, Maria de Jesus Barbosa Cavalcante⁶, Franc-Christophe Baurens⁷, Orzenil Bonfim da Silva Jr³, Georgios J. Pappas Jr^{1,3}, Luc Pignolet⁸, Catherine Abadie⁸, Ana Y. Ciampi³, Pietro Piffanelli^{7,9} and Robert N. G. Miller^{1*}

5: Miller, R.N.G. ; Passos, Marco Aurélio ; Emediato, Flavia Leonel ; Teixeira, C. C. ; Pappas Júnior, Georgios J.. Candidate Resistance Gene Discovery: Resistance Gene Analog Characterization and Differential Gene Expression Analysis in *Musa-Mycosphaerella* Host-Pathogen Interactions. *Acta Horticulturae*, v. 897, p. 179-186, 2011.

Candidate Resistance Gene Discovery: Resistance Gene Analog Characterisation and Differential Gene Expression Analysis in *Musa-Mycosphaerella* Host-Pathogen Interactions

R.N.G. Miller^{1,2}, M.A.N. Passos¹, F.L. Emediato¹, C. de Camargo Teixeira¹ and G.J. Pappas Júnior^{1,3}

¹Universidade Católica de Brasília, Brasília, DF, Brazil

²Universidade de Brasília, Departamento de Biologia Celular, Brasília, DF, Brazil

³EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, DF, Brazil

Keywords: black leaf streak, black Sigatoka, EST, plant disease resistance genes, RGA

ANEXO 2: Trabalhos científicos apresentados em congresso com resumo publicado em revista ou anais

1. Passos, MAN; Cruz, VO; Emediato, FL; Azevedo, VCR; Brasileiro, ACM; Amorim, EP; Ferreira, CF; Togawa, RC; Pappas, GJ; Silva, OB; Miller, RNG. Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. In: resumos, VII congresso brasileiro de micologia de 25 a 28 de novembro de 2013, Belem, PA, Brazil
2. Passos, MAN; Pappas Jr, GJ; Silva JR, O; Emediato, FL; Teixeira, CC; Franco, OC; Miller, RNG. 2013. Characterization of genetic components of resistance to *Mycosphaerella musicola* in *Musa acuminata*. In: IV Simpósio Brasileiro de Genética Molecular de Plantas, 8-12 de abril de 2013, Bento Gonçalves, RS, Brasil.
3. Miller, R. N. G.; Passos, M. A. N.; Emediato, F. L.; Teixeira, C. C.; Silva Junior, O. B. da; Pappas Junior, G. J. Gene expression analysis in *Musa acuminata*-*Mycosphaerella* interactions. In: Congresso Brasileiro de Fitopatologia, 45., 2012, Manaus. Fito 2012. Tropical Plant Pathology, Brasília, DF, v. 37, 2012. Suplemento.
4. Emediato, F.L., Passos, M.A.N., Teixeira, C.C., Pappas JR, G.J., Miller, R.N.G. 2012. Análise de expressão de genes análogos de resistência em *Musa acuminata* durante as interações compatíveis e incompatíveis com *Mycosphaerella musicola*. In: 45 Congresso Brasileiro de Fitopatologia, 2012, Manaus. Resumos, 45 Congresso Brasileiro de Fitopatologia.
5. Passos, MAN; Pappas Jr, GJ; Silva Jr, O; Emediato, FL; Teixeira, CC; Miller RNG. 2012. Identificação e caracterização de componentes genéticos de resistência a *Mycosphaerella musicola* em *Musa acuminata*. In: 45 Congresso Brasileiro de Fitopatologia, 2012, Manaus. Resumos, 45 Congresso Brasileiro de Fitopatologia.
6. Miller, RNG; James-Kay, A; Passos, MAN; Emediato, FL; Teixeira, CC; Pappas GJ. 2010. Candidate Resistance Gene Discovery in *Musa*-*Mycosphaerella* Host-Pathogen Interactions. VII Encuentro Latinoamericano y del Caribe sobre Biotecnología Agropecuaria, – REDBIO Mexico, 1 - 5 de Novembro de 2010, Expo Guadalajara, Guadalajara, México.
7. Passos, MAN; Pappas Jr, GJ; Emediato, FL; Teixeira, CC; Miller RNG. 2010. Identificação e caracterização de componentes genéticos de resistência e defesa aos patógenos *Mycosphaerella fijiensis* e *Mycosphaerella musicola* em *Musa* spp. In: 43 Congresso Brasileiro de Fitopatologia, 2010, Cuiabá. Resumos, 43 Congresso Brasileiro de Fitopatologia, 2010.
8. Emediato, FL, Passos, MAN; Teixeira, CC; Pappas Jr, GJ; Miller RNG. 2010. Caracterização de genes análogos de resistência (RGAs) em cultivares de *Musa acuminata* contrastantes em resistência à estresses bióticos. In: 43 Congresso Brasileiro de Fitopatologia, 2010, Cuiabá. Resumos, 43 Congresso Brasileiro de Fitopatologia, 2010.

1: Marco A N Passos, Viviane Oliveira de Cruz, Flavia L Emediato, Vania C Renno Azevedo, Ana C M Brasileiro, Edson P Amorim, Claudia F Ferreira, Roberto C Togawa, Georgios J Pappas Junior, Orzenil Bonfim da Silva Jr, Robert NG Miller. Analysis of the leaf transcriptome of *Musa acuminata* during interaction with *Mycosphaerella musicola*: gene assembly, annotation and marker development. In: resumos, VII congresso brasileiro de micologia de 25 a 28 de novembro de 2013, Belem, PA, Brazil

LIVRO DE RESUMOS VII CONGRESSO BRASILEIRO DE MICOLOGIA DE 25 A 28 DE NOVEMBRO DE 2013

**ANALYSIS OF THE LEAF TRANSCRIPTOME OF MUSA ACUMINATA
DURING INTERACTION WITH MYCOSPHAERELLA MUSICOLA: GENE
ASSEMBLY, ANNOTATION AND MARKER DEVELOPMENT**

Marco A N Passos (1); Viviane Oliveira de Cruz (1); Flavia L Emediato (1); Vânia C Rennó Azevedo (2); Ana C M Brasileiro (2); Edson P Amorim (3); Claudia F Ferreira (3); Roberto C Togawa (2); Georgios J Pappas Júnior (1); Orzenil Bonfim da Silva Jr (2); Robert NG Miller (1)

(1) Universidade de Brasília, Departamento de Biologia Celular, Brasília, DF

(2) EMBRAPA Mandioca e Fruticultura Tropical, Cruz das Almas, BA

(3) EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, DF

2: Passos, MAN; Pappas Jr, GJ; Silva JR, O; Emediato, FL; Teixeira, CC; Franco, OC; Miller, RNG. 2013. Characterization of genetic components of resistance to *Mycosphaerella musicola* in *Musa acuminata*. In: IV Simpósio Brasileiro de Genética Molecular de Plantas, 8-12 de abril de 2013, Bento Gonçalves, RS, Brasil.



CHARACTERIZATION OF GENETIC COMPONENTS OF RESISTANCE TO *MYCOSPHAERELLA MUSICOLA* IN *MUSA ACUMINATA*

Passos, MAN¹; Pappas Jr, GJ¹; Silva JR, O²; Emediato, FL¹; Teixeira, CC³; Franco, OL³; Miller, RNG¹.

¹Universidade de Brasília/ IB/BIOCEL/ CEP 70910-900, Brasília DF, Brasil. ²EMBRAPA Recursos Genéticos e Biotecnologia, Parque Estação Biológica, CP 02372, CEP 70.770-900, Brasília, DF, Brazil. ³Universidade Católica de Brasília/ Pós-graduação em Ciências Genômicas e Biotecnologia/ SGAN 916, Módulo B/ CEP 70790-160, Brasília, DF, Brasil.

*E-mail: marconinomia@gmail.com.

Keywords: *Musa acuminata*, *Mycosphaerella musicola*, disease resistance, transcriptomics, proteomics.

3: Miller, R. N. G.; Passos, M. A. N.; Emediato, F. L.; Teixeira, C. C.; Silva Junior, O. B. da; Pappas Junior, G. J. Gene expression analysis in *Musa acuminata*-*Mycosphaerella* interactions. In: Congresso Brasileiro de Fitopatologia, 45., 2012, Manaus. Fito 2012. Tropical Plant Pathology, Brasília, DF, v. 37, 2012. Suplemento.



Tropical Plant Pathology 37 (Suplemento), agosto 2012
45º Congresso Brasileiro de Fitopatologia - Manaus, AM
Copyright the Brazilian Phytopathological Society <http://www.sbfito.com.br>

Gene expression analysis in *Musa acuminata*-*Mycosphaerella* interactions.
(Análise de expressão genética durante as interações *Musa acuminata*-*Mycosphaerella*)

Miller, R.N.G.¹, Passos, M.A.N.¹, Emediato, F.L.¹, Teixeira, C.C.², Silva Jr, O³; Pappas Jr, G.J.³.

¹Universidade de Brasília, ²Universidade Católica de Brasília, ³EMBRAPA Recursos Genéticos e Biotecnologia. *E-mail*: robertmiller@unb.br.

4: Emediato , F.L., Passos, M.A.N., Teixeira, C.C., Pappas JR, G.J., Miller, R.N.G. 2012. Análise de expressão de genes análogos de resistência em *Musa acuminata* durante as interações compatíveis e incompatíveis com *Mycosphaerella musicola*. In: 45 Congresso Brasileiro de Fitopatologia, 2012, Manaus. Resumos, 45 Congresso Brasileiro de Fitopatologia.



Tropical Plant Pathology 38 (Suplemento), agosto 2012
45º Congresso Brasileiro de Fitopatologia - Manaus, AM
Copyright the Brazilian Phytopathological Society <http://www.sbfito.com>

MELHORAMENTO GENÉTICO

708

Análise de expressão de genes análogos de resistência em *musa acuminata* durante as interações compatíveis e incompatíveis com *Mycosphaerella musicola*.

(Analysis of Resistance Gene Analog expression in *Musa acuminata* during compatible and incompatible interactions with *Mycosphaerella musicola*.)

Emediato, F.L.¹, Passos, M.A.N.¹, De Camargo Teixeira, C.², Pappas Jr, G.J.³, Miller, R.N.G.¹

¹Universidade de Brasília; ²Universidade Católica de Brasília; ³EMBRAPA Recursos Genéticos e Biotecnologia. *E-mail:* flavinhaleonel@gmail.com

5: Passos, MAN; Pappas Jr, GJ; Silva Jr, O; Emediato, FL; Teixeira, CC; Miller RNG. 2012. Identificação e caracterização de componentes genéticos de resistência a *Mycosphaerella musicola* em *Musa acuminata*. In: 45 Congresso Brasileiro de Fitopatologia, 2012, Manaus. Resumos, 45 Congresso Brasileiro de Fitopatologia.



Tropical Plant Pathology 38 (Suplemento), agosto 2012
45º Congresso Brasileiro de Fitopatologia - Manaus, AM
Copyright the Brazilian Phytopathological Society <http://www.sbfito.com>

MELHORAMENTO GENÉTICO

470

Identificação e caracterização de components genéticos de resistência a *Mycosphaerella musicola* em *Musa acuminata*.

(Identification and characterization of genetic components of resistance to *Mycosphaerella musicola* in *Musa acuminata*.)

Passos, M.A.N.¹, Pappas J.R., Silva J.R., O.², Emediato, F.L.¹, Teixeira, C.C.³, Miller, R.N.G.¹

¹Universidade de Brasília, ²EMBRAPA Recursos Genéticos e Biotecnologia. ³ Universidade Católica de Brasília *E-mail:* marconinomia@gmail.com

6: Miller, RNG; James-Kay, A; Passos, MAN; Emediato, FL; Teixeira, CC; Pappas GJ. 2010. Candidate Resistance Gene Discovery in Musa-Mycosphaerella Host-Pathogen Interactions. VII Encuentro Latinoamericano y del Caribe sobre Biotecnología Agropecuaria, – REDBIO Mexico, 1 - 5 de Noviembre de 2010, Expo Guadalajara, Guadalajara, México.

REDBIO México 2010

Simposio: Biotecnología de Frutos Tropicales

Responsable: Dr. Jorge Manuel Santamaría Fernández



***Candidate Resistance Gene Discovery in Musa-Mycosphaerella
Host-Pathogen Interactions***

Robert Miller

Universidad Católica de Brasília. Brasil.

Universidad Católica de Brasília

7: Passos, MAN; Pappas Jr, GJ; Emediato, FL; Teixeira, CC; Miller RNG. 2010. Identificação e caracterização de componentes genéticos de resistência e defesa aos patógenos *Mycosphaerella fijiensis* e *Mycosphaerella musicola* em *Musa* spp. In: 43 Congresso Brasileiro de Fitopatologia, 2010, Cuiabá. Resumos, 43 Congresso Brasileiro de Fitopatologia, 2010.

07.040

Identificação e caracterização de componentes genéticos de resistência e defesa aos patógenos *Mycosphaerella fijiensis* e *Mycosphaerella musicola* em *Musa* spp. Passos, M. A. N.¹; Pappas Jr., G. J.²; Emediato, F. L.¹; Teixeira, C. C.³; Miller, R. N. G.¹ - ¹Universidade de Brasília - Biologia Celular; ²Embrapa Recursos Genéticos e Biotecnologia - Laboratório de Bioinformática; ³Universidade Católica de Brasília - Ciências Genômicas e Biotecnologia. E-mail: marconinomia@gmail.com. Identification and characterization of genetic components of resistance and defense against pathogens *Mycosphaerella fijiensis* e *Mycosphaerella musicola* in *Musa* spp.

O desenvolvimento de programas de melhoramento genético possibilita a obtenção de genótipos resistentes a fitopatógenos. Nesse contexto, foi conduzida a análise da expressão gênica diferencial em folhas de cultivares contrastantes de *Musa acuminata* Calcutta 4 e *M. acuminata* Cavendish inoculadas in vitro com *Mycosphaerella fijiensis*, causador da Sigatoka Negra. O seqüenciamento foi realizado em 8000 clones. A análise diferencial in silico demonstrou que 220 genes foram diferencialmente expressos, sendo que mais de 20 deles estão relacionados à resistência e defesa, incluindo metalotioneína, germin-like proteínas, ferredoxinas, glutionas S-transferases, proteínas relacionadas a patogênese, tioredoxina, glicolato oxidase, proteínas responsivas a etileno, superoxidases dismutases, proteínas responsivas a stress e proteínas responsivas a auxina. Bioensaios foram conduzidos in planta inoculados com *M. fijiensis* e *M. musicola*. A utilização de seqüenciamento massal está sendo realizada visando o aumento da resolução da análise de genes diferencialmente expressos. **Apoio Financeiro:** IAEA, CNPq, UCB, FINEP.

8: FL, Passos, MAN; Teixeira, CC; Pappas Jr, GJ; Miller RNG. 2010. Caracterização de genes análogos de resistência (RGAs) em cultivares de *Musa acuminata* contrastantes em resistência à estresses bióticos. In: 43 Congresso Brasileiro de Fitopatologia, 2010, Cuiabá. Resumos, 43 Congresso Brasileiro de Fitopatologia, 2010.

XLIII Congresso Brasileiro de Fitopatologia

07.041

Caracterização de genes análogos de resistência (RGAs) em cultivares de *Musa acuminata* contrastantes em resistência à estresses bióticos Emediato, F. L.¹; Passos, M. A. N.¹; Teixeira, C. C.²; Pappas Jr., G. J.³; Miller, R. N. G.¹ - ¹Universidade de Brasília - Biologia Celular; ²Universidade Católica de Brasília - Ciências Genômicas e Biotecnologia; ³Embrapa Recursos Genéticos e Biotecnologia - Laboratório de Bioinformática. *E-mail: flavinhaleonel@gmail.com*. The characterization of Resistance Gene Analogs (RGAs) in *Musa acuminata* cultivars contrasting in resistance to biotic stresses.

A caracterização de genes análogos de resistência (RGAs) foi realizada em cultivares de *Musa acuminata* contrastantes em resistência à doença Sigatoka negra causada pelo fungo *Mycosphaerella fijiensis*. Primers direcionados à família de R-genes NBS-LRR geraram cerca de 200 seqüências organizadas em 38 contigs com similaridade à RGAs desta classe. A diferenciação de contigs com base nos genótipos não foi observada, mostrando em um mesmo contig seqüências provenientes tanto de cultivares resistentes à Sigatoka negra, como suscetíveis. Primers direcionados à família de R-genes RLK geraram 65 ORFs completas e mais uma vez sem a diferenciação dos contigs de proteínas quinase com base em cultivares contrastando em resistência a Sigatoka negra. A análise filogenética foi desenvolvida por máxima parcimônia com bootstrap de 1000 réplicas, demonstrando alta divergência entre as seqüências de RGAs tanto de NBS-LRR, quanto de RLK. A identificação e clonagem de R-genes em *Musa* proporcionará novas oportunidades em potencial para o melhoramento genético através de seleção assistida por marcadores e modificação genética. **Apoio Financeiro:** IAEA, CNPq, UCB, CAPES, FINEP.

Anexo 3: Lista de 125 pares de *primers* gerados para locos SSR presentes em seqüências de clones BAC de *M.acuminata* Calcutta 4 e Cavendish Grande Naine

Identificação BAC	SSR	Tamanho do produto	Primer FW	TM	Primer RV	TM2	Localização SSR	Localização FW	Localização RV	Loco
>MA4052E23	(TAA)9	199	ACAAATCTCCAAACCGATCAAC	60.2	TTATTGTGCAACGGAAGATGAC	60.0	9788	9646	9823	loco 1
>MA4052E23	(TTC)8	124	ACCAGAAGACGTTAGATGGTGG	60.4	GCTGATGCTAGTGCTGATGCTA	60.7	40885	12893	12995	loco 2
>MA4052E23	(AGA)13	373	AACCTCTGATGGACTTGAGAACA	60.1	CCCAGTTTGGAGTTTCTAGGTG	60.0	13515	13293	13644	loco 3
>MA4052E23	(AG)6	327	GCTATCTTACCATCGCTTGACAG	60.2	ATGTTGGCAGTGAAATGTCTTG	60.0	18840	18712	19017	loco 4
>MA4052E23	(TTC)9	399	TCAAGAGGTATGCTACATTCAGGA	60.1	GAAAGATGACCGAGGCAATTA	59.1	40885	40608	40986	loco 5
>MA4052E23	(AT)15	397	GCTCAAGAACTCACATGGACA	60.2	TGGAATCTATGGGTTATGGCTT	59.7	70490	45469	45844	loco 6
>MA4052E23	(GA)8	356	TTGAGACACAAAATCAGCAAG	60.2	ACGTAGTTCTGTTGCATTGGC	60.1	47436	47288	47623	loco 7
>MA4052E23	(AT)10	374	CAACCCATCATTCTTTGGTG	60.2	CTCACCTTCTGGTTTCAGCTTT	59.9	70490	53505	53857	loco 8
>MA4052E23	(CT)8	352	TTCAAGGAGCGTGAAGAAGCTG	59.5	TAGCCAGGTCACAAATTCCTAA	58.7	66640	66363	66693	loco 9
>MA4052E23	(AT)24	113	TTCTCTTCTCCCTCTCTCT	59.9	ATGACTCGGTTACGTTGATGTG	59.9	70490	70462	70553	loco 10
>MA4052E23	(TGA)7	236	TGGGTCTTCATCACCTACCTTT	59.8	TCAGTGGCTTGTTCTTGCTGT	59.9	79038	78867	79081	loco 11
>MA4105F04	(TTC)9	399	TCAAGAGGTATGCTACATTCAGGA	60.1	GAAAGATGACCGAGGCAATTA	59.1	6071	5794	6172	loco 12
>MA4113F17	(CT)17	360	TTCTTTGGCCTCTTTTCTCTCA	60.4	CTTCTTCCTCATTGCTTCCAT	59.7	52541	6459	6797	loco 13
>MA4113F17	(TA)13	357	TTTGGCCTCTTTTCTCTCACTC	59.9	CTTCTTCCTCATTGCTTCCAT	59.7	62481	6462	6797	loco 14
>MA4113F17	(TTC)13	148	ACCAGAAGACGTTAGATGGTGG	60.4	TAGTCGATGATGATGCTGCTTT	59.8	7953	7886	8012	loco 15
>MA4113F17	(CAG)5	154	ACCAGAAGACGTTAGATGGTGG	60.4	CAAACCTAGTCGATGATGATGC	59.5	7991	7886	8018	loco 16
>MA4113F17	(AGA)7	313	GGAAACCTCTGATGGACTTGAG	60.1	TTTGGGTGTGAAGCTACTGCTA	59.9	57361	8349	8640	loco 17
>MA4113F17	(TAA)10	212	ACCCAACCTTGATTGTACCGAC	60.1	CCCTTGTCTTCTCTCTCTTT	60.1	61341	12108	12298	loco 18
>MA4113F17	(TA)25	354	CGGTTCTACCTATAACGATGG	59.7	AACCAACGAGGACGAGATAGAC	59.6	62481	22240	22572	loco 19
>MA4113F17	(AT)8	285	GAACCCTATCTCCCAAAAGAC	60.1	AAGCTCATTCCATCGTAACACC	60.3	62540	47939	48202	loco 20
>MA4113F17	(AGA)6	398	AAAACCTCCGCTACCCAAT	59.8	TTCATCCACCATCATCTGTTGT	60.2	57361	49826	50202	loco 21
>MA4113F17	(CT)6	305	GTGAAGGAATCATCAAGCACAA	60.1	GCAAGAATAAAGATGGGTTTGG	59.8	52541	52515	52798	loco 22
>MA4113F17	(TA)38	293	AGACCAAACAATGAAACCACCT	59.7	GCTAACTGGCCTATCAACTTCG	60.2	62481	52960	53231	loco 23
>MA4113F17	(TA)10	306	AATCTGCCTCATCTTTCTGTCT	59.7	ACTATGCCTTTTGTGTTGTTGGC	60.4	62481	55786	56070	loco 24

>MA4113F17	(AGA)6	387	AAAACCTTCGCTACCCAAT	59.8	TCATCTGTTGTTTCGCTCAGTT	59.9	57361	57272	57637	loco 25
>MA4113F17	(TAA)8	109	ACCCAATTTTGATTGTACCGAC	59.9	TTATTGTGCAACGGAAGATGAC	60.0	61341	61286	61373	loco 26
>MA4140M09	(AT)28	398	TTATAGGATCGACGCTTTGTTG	59.2	AGTGCTATTTTACAGAGTGCGG	58.6	87298	25156	25532	loco 27
>MA4140M09	(TA)40	382	TATAGGATCGACGTTTTGTTGC	59.1	GGATACATTTCCGGATGCTGTC	59.4	51096	50870	51231	loco 28
>MA4140M09	(TGG)8	226	CGACTGCCTCCAAGGTATGTAT	60.3	GAACACTGAGAACGTCAGCATC	59.9	64223	64150	64354	loco 29
>MA4140M09	(AT)27	358	AACTGCACAGCCACAAACATAC	60.1	TTGTCCAACCTACTCTCTCTTTTC	59.4	87298	70060	70393	loco 30
>MA4140M09	(AC)12	172	CATAACCTATTGCACAGGCTCA	60.1	AGCCAACGGTGAACACATAAC	59.9	70840	70725	70876	loco 31
>MA4140M09	(AG)7	364	TCGATACAGAGGAGGGTCTGAT	60.0	TTACGTCAGCAAGCAAGGATTA	59.9	87067	87046	87388	loco 32
>MA4140M09	(AT)9	364	TCGATACAGAGGAGGGTCTGAT	60.0	TTACGTCAGCAAGCAAGGATTA	59.9	87298	87046	87388	loco 33
>MAC054B03	(AT)7	367	TATTTATGTCACGCACGTTTGC	60.9	TGCCCTCCTAATTCCTTATACCCA	60.9	99865	6796	7140	loco 34
>MAC054B03	(TA)7	397	TATAGGATCGACGTTTTGTTGC	59.1	GACAGAAGCAAATGGAGTGCTA	59.5	47016	22914	23289	loco 35
>MAC054B03	(AT)21	395	TATTTATGTCACGCACGTTTGC	60.9	TGCCCTCCTAATTCCTTATACCCA	60.9	99865	31279	31651	loco 36
>MAC054B03	(AT)19	319	CGAGAGAGACATTTGATTGACC	60.1	ATTTTACAGAGTGAGGCATGGG	60.3	99865	40081	40378	loco 37
>MAC054B03	(TA)11	330	TGTGTGTGCAAAGAGTAGTGTCAT	59.7	GCCCTCCTAATTCCTTATACCCAA	59.7	47016	46903	47210	loco 38
>MAC054B03	(TGG)6	227	CGACTGCCTCCAAGGTATGTAT	60.3	ACGGGAACACTGAGAACGTAAG	60.5	58200	58127	58332	loco 39
>MAC054B03	(AC)15	350	AACTGCACAGCCACAAACATAC	60.1	TTGTCCAACCTACTCTCTCTCTC	59.6	64442	63677	64002	loco 40
>MAC054B03	(AC)21	181	CATAACCTATTGCACAGGCTCA	60.1	GCCAACAGTGAACACACAACAT	60.9	64442	64335	64494	loco 41
>MAC054B03	(AG)8	239	AACCATCCAAGATCCAGAGCTA	60.0	TCGTCATCATCCGTAAGAGAGA	59.7	80687	80578	80795	loco 42
>MAC054B03	(AT)15	384	AAATCTTCGATACAGAGGACGG	59.6	TTACGTCAGCAAGCAAGGATTA	59.9	99865	80660	81022	loco 43
>MAC054B03	(AT)10	334	CACAACATCAGCAACACACTTG	60.2	CCTCCATTCATTCCAGTCTAGC	60.0	99865	98878	99190	loco 44
>MAC054B03	(AT)17	347	AAAAGCTGGAGGGTTAAGGTGT	60.3	GCGATCAGTAAAATCATCTCGG	60.9	99865	99650	99975	loco 45
>MAC054B03	(TCT)13	399	TGACTTTGTCCGAGTTCTTGTG	60.3	GTGGGTGGATTTGGTAGATTGT	59.9	102007	101834	102211	loco 46
>MAC063A04	(AGA)8	358	AACCTCTGATGGACTTGAGAACA	60.1	CCCAGTTTGAGTTTCTAGGTG	60.0	89550	2532	2868	loco 47
>MAC063A04	(AT)8	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	93010	3912	4260	loco 48
>MAC063A04	(AT)6	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	93010	3912	4260	loco 49
>MAC063A04	(TA)7	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	100919	3912	4260	loco 50
>MAC063A04	(TA)8	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	100919	3912	4260	loco 51

>MAC063A04	(AT)11	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	93010	3912	4260	loco 52
>MAC063A04	(TA)7	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	100919	3912	4260	loco 53
>MAC063A04	(TA)8	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAAACCGTCTTTATCGT	59.8	100919	3912	4260	loco 54
>MAC063A04	(AGA)9	364	ATGGACTTGAGAGCACACACAT	59.6	TGGAGTTCCTTGAAATAGGAGG	59.5	89550	12691	13033	loco 55
>MAC063A04	(CT)7	306	GTGAAGGAATCATCAAGCACAA	60.1	GCAAGAATAAGAATGGGTTTGG	59.8	79912	15514	15798	loco 56
>MAC063A04	(A)15	106	CTCCACTTTGATTGGTGCATA	59.9	GTGCTGAGTCTGTGTTTCATC	59.7	107871	17801	17885	loco 57
>MAC063A04	(TTC)7	153	GCAAAAGATGATTGTGAGTTGC	59.7	TCGAGACAAACCTAGTCGATGA	59.8	83045	18653	18784	loco 58
>MAC063A04	(AT)8	285	GAACCCTATCTCCCAAAAAGAC	60.1	AAGCTCATTCCATCGTAACACC	60.3	93010	25626	25889	loco 59
>MAC063A04	(AG)6	331	GGTGTACGATGGAATGAGCTT	60.3	ATCACCAATATGCGAAGAGACC	60.3	53504	25889	26198	loco 60
>MAC063A04	(TA)13	336	AGTCACGGTCTCTTCGCATATT	60.1	ATGTTGGCAGTGAAATGTCTTG	60.0	100919	26192	26506	loco 61
>MAC063A04	(AAG)9	400	AGACTCGGAAACAACAGAGGAG	59.9	CAACCTATGGTCACACGAAGAA	60.0	27397	27370	27748	loco 62
>MAC063A04	(TTC)6	174	AAGAACGAAACACAAAGCCGAT	60.1	TCGAAACAAACCTAGTCGATGA	59.7	83045	43762	43914	loco 63
>MAC063A04	(AGA)13	349	GGAAAGAACGAAGGAGGAGAAG	60.7	CACCATCATCTGTTGTTACGCT	60.0	89550	44451	44778	loco 64
>MAC063A04	(CT)6	305	GTGAAGGAATCATCAAGCACAA	60.1	GCAAGAATAAGAATGGGTTTGG	59.8	79912	47094	47377	loco 65
>MAC063A04	(AT)19	284	GGATGCTCCACATAGTCAATCA	59.9	CTCATCTCAAACCATGCCTTTT	60.4	93010	48526	48788	loco 66
>MAC063A04	(AGA)10	312	GGAAACCTCTGATGGACTTGAG	60.1	CTGCTGCTGCTACTGGTCAT	59.2	89550	50068	50360	loco 67
>MAC063A04	(AT)8	285	GAACCCTATCTCCCAAAAAGAC	60.1	AAGCTCATTCCATCGTAACACC	60.3	93010	53033	53296	loco 68
>MAC063A04	(AG)6	331	GGTGTACGATGGAATGAGCTT	60.3	ATCACCAATATGCGAAGAGACC	60.3	53504	53296	53605	loco 69
>MAC063A04	(TA)22	353	AGTCACGGTCTCTTCGCATATT	60.1	ATGTTGGCAGTGAAATGTCTTG	60.0	100919	53599	53930	loco 70
>MAC063A04	(TA)19	399	TTTAGCCATGTCCTTCAACGTA	59.6	TTTCTTCTCCCTCTGTGATGA	58.9	100919	63569	63946	loco 71
>MAC063A04	(AGA)10	367	ATGGACTTGAGAGCACACACAT	59.6	TGGAGTTCCTTGAAATAGGAGG	59.5	89550	77060	77405	loco 72
>MAC063A04	(CT)7	306	GTGAAGGAATCATCAAGCACAA	60.1	GCAAGAATAAGAATGGGTTTGG	59.8	79912	79886	80170	loco 73
>MAC063A04	(TTC)7	153	GCAAAAGATGATTGTGAGTTGC	59.7	TCGAGACAAACCTAGTCGATGA	59.8	83045	82946	83077	loco 74
>MAC063A04	(AGA)8	280	AACTTGGGACAATTATGGAGGA	59.7	CCAGTTTGGAGTTTCTTTTCGT	59.6	89550	89414	89672	loco 75
>MAC063A04	(TA)7	399	CGTCAGCCAAGACAAAGGTTTA	61.5	GCTTCGGATGGTTTGGTTATTT	61.3	100919	92203	92580	loco 76
>MAC063A04	(TA)8	264	ACCTCCTCTTTAGCCATGTCCT	60.4	CATGCAAGAAAATCATGTGGC	61.4	100919	100749	100992	loco 77
>MAC063A04	(TAA)7	199	ACAAATCTCAAACCGATCAAC	60.2	TTATTGTGCAACGGAAGATGAC	60.0	101737	101593	101770	loco 78

>MAC091O16	(AGA)6	274	ATGGACTTGAGAGCACACACAT	59.6	CTGCTACTGGAAATGATCGTGA	60.2	49153	2749	3001	loco 79
>MAC091O16	(TA)36	395	ACCTCCTCTTTAGCCATGTCCT	60.4	CATGCAAGAAAATCATGTGGC	61.4	83239	5461	5835	loco 80
>MAC091O16	(TAA)10	346	ACCCAACCTTTGATTGTACCGAC	60.1	CAGAAGACCACGCTCTAATCG	60.0	6578	6523	6848	loco 81
>MAC091O16	(CT)8	161	TCAACCTCCTCTTTAGCCATGT	60.1	GCAGCTATGACTTCCACCTACC	60.1	25659	25546	25685	loco 82
>MAC091O16	(TTC)11	165	GCAAAAGATGATTGTGAGTTGC	59.7	TCGAGACAAACCTAGTCGATGA	59.8	85366	28734	28877	loco 83
>MAC091O16	(AGA)12	311	GGGCAATTTACGAGGAACAAT	60.1	CTAGGAAAGCTCCCCAGTTTG	60.2	49153	29303	29593	loco 84
>MAC091O16	(TA)18	399	AGCTTGATGCCAAAGGAACTTA	60.2	TAGCTTCTGTGCTTATGAACGC	59.6	83239	33572	33949	loco 85
>MAC091O16	(AT)19	344	AATCTGCCTCATCCTTTCTGTC	59.7	ACTATGCCTTTTGTGTTGTGGC	60.4	89620	36453	36775	loco 86
>MAC091O16	(AGA)7	365	ATGGACTTGAGAGCACACACAT	59.6	CCCAGTTTGGAGTTTCTTTTCA	60.4	49153	37838	38181	loco 87
>MAC091O16	(AAG)6	104	AAAATAAGAGGATAGGGAGGCG	59.9	TGCCAAATCACAGGTCAAAGT	60.5	41583	41534	41617	loco 88
>MAC091O16	(AGA)10	400	GGAAACCTCTGATGGACTTGAG	60.1	CTAGGAAAGCTCCCCAGTTTG	60.2	49153	48922	49301	loco 89
>MAC091O16	(TA)26	278	CCAACTTAGAGAGAGAGATTCCA	58.2	TATGCCTTTTGTGTTGTGGC	58.6	83239	54471	54729	loco 90
>MAC091O16	(CTT)18	332	TCGATCTGCACTTCTTCTCAAA	60.1	TCGAGACAAACCTAGTCGATGA	59.8	55619	55399	55709	loco 91
>MAC091O16	(CAT)7	317	TGACGCCTTCTTCTTCTTCTTC	60.1	TCATTTACCTGAGTGGTTGACG	60.0	55691	55614	55909	loco 92
>MAC091O16	(AT)13	330	CAAGCATTAGCTCGAAGGAACT	60.0	TAGCCCATATACCCCTCTATGC	59.3	89620	71927	72235	loco 93
>MAC091O16	(TTC)17	376	GGTATGTACATTCAGGAAACGA	60.3	CCTGCACCTTCGAGACAAAT	60.2	85366	85095	85451	loco 94
>MAC091O16	(AT)19	337	CACAATCCTAACGGAACCTCAAA	59.1	CCCACATTTTGCATGTATCAAC	60.1	89620	89484	89799	loco 95
>MAC139M12	(AT)25	346	CTCGCTTTCTCCGCAAGATT	60.8	CAACATCTCTGTCTAAGTGCCG	59.9	81181	4960	5284	loco 96
>MAC139M12	(GA)6	284	CAAAATGGAGGATGCTAAAAGG	59.9	TGTATGATTTCTGCCTGTGTCC	60.0	14118	12103	12365	loco 97
>MAC139M12	(AG)17	284	CAAAATGGAGGATGCTAAAAGG	59.9	TGTATGATTTCTGCCTGTGTCC	60.0	36586	12103	12365	loco 98
>MAC139M12	(GA)14	367	ATGAATGCAAGCAATGACTGAC	60.1	TATGAAAGCAATACTACTGCGG	60.1	14118	13074	13419	loco 99
>MAC139M12	(AG)9	347	GGCTGTAGCTTCTCTTGACAG	60.5	TCTTTATTCACCCAACACCTCC	60.2	36586	13631	13956	loco 100
>MAC139M12	(GA)9	161	ATGGGAAAAGAGGAGAACACAA	59.9	AATCCACATAAAGCAGCCAAAG	60.4	14118	14073	14212	loco 101
>MAC139M12	(AG)7	363	AAGTGGAGGAATCAAGACAGGA	60.1	AGGGGAAATAGAGAGGAGATGG	59.9	36586	15239	15580	loco 102
>MAC139M12	(AG)15	349	AAATGCTCTACTCATGGCTTCC	59.7	TCCTTCTTAACCTCGATCTTG	59.7	36586	16975	17302	loco 103
>MAC139M12	(ATA)6	380	GTGTGTGGAGAATGTTTGAGGA	60.0	CTTGACAGAGTCCAACCAAGTTA	59.4	20190	19912	20270	loco 104
>MAC139M12	(AG)9	388	ACAGTTTTGACTACATCGGGT	59.7	TAATACCTCTCTCCTCCTCCCC	59.9	36586	26333	26699	loco 105

>MAC139M12	(AG)11	208	AACCTGCGAGGCTAGTAACAGA	60.4	TTGTTACCTTTTCTCTCTCCG	60.7	36586	36486	36672	loco 106
>MAC139M12	(AGA)6	372	GAAACAACGAAGGAGGAAAGT	58.8	ACACGGAGGAAGAGATTGAAAG	59.7	89903	38567	38917	loco 107
>MAC139M12	(AGA)15	393	CGGAAACAACAAAGGAGGAG	59.7	ATTGAAAGAATGGAGCAGGTGT	60.0	89903	45857	46228	loco 108
>MAC139M12	(TA)31	279	CCAACCTAGAGAGAGAGATTCCA	58.2	TATGCCTTTTGTGGG	58.6	100324	51198	51457	loco 109
>MAC139M12	(AGA)10	285	CAAAACCTTCTGCTACCCAATC	60.0	TGTGACTATTCTTTAAGCTCCTCG	60.2	89903	52837	53097	loco 110
>MAC139M12	(AT)24	384	AGAAGCTCAAAAGCAAAGATGG	60.0	TGACTTGTTTCGTGACAGACCT	59.8	81181	62652	63014	loco 111
>MAC139M12	(AGA)8	280	AACCTGGGACAATTATGGAGGA	59.7	CCAGTTTGGAGTTTCTTTTCGT	59.6	89903	64140	64398	loco 112
>MAC139M12	(TAA)6	179	TAGCTGTACTCACAAATCCCA	59.6	TGCAACGGAAGATGACACTAAC	60.1	75977	67647	67804	loco 113
>MAC139M12	(TAA)8	209	TACTTTGATTGTACCGACGCAC	60.0	CACTCATCCCTTTGTCCTTCTC	60.1	75977	75926	76113	loco 114
>MAC139M12	(AGA)8	358	AACCTCTGATGGACTTGAGAACA	60.1	CCCAGTTTGGAGTTCTAGGTG	60.0	89903	79521	79857	loco 115
>MAC139M12	(AT)8	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	81181	80901	81249	loco 116
>MAC139M12	(AT)6	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	81181	80901	81249	loco 117
>MAC139M12	(TA)7	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	100324	80901	81249	loco 118
>MAC139M12	(TA)8	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	100324	80901	81249	loco 119
>MAC139M12	(AT)11	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	81181	80901	81249	loco 120
>MAC139M12	(TA)8	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	100324	80901	81249	loco 121
>MAC139M12	(TA)7	370	GATACACGGGTGGTGAGGAG	60.3	TTACCGAAACCGTCTTTATCGT	59.8	100324	80901	81249	loco 122
>MAC139M12	(AGA)9	364	ATGGAAGTGGAGAGCACACACAT	59.6	TGGAGTTCCTTGAAATAGGAGG	59.5	89903	89680	90022	loco 123
>MAC139M12	(CT)7	306	GTGAAGGAATCATCAAGCACAA	60.1	GCAAGAATAAAGATGGGTTTGG	59.8	92529	92503	92787	loco 124
>MAC139M12	(TTC)7	153	GCAAAAGATGATTGTGAGTTGC	59.7	TCGAGACAAACCTAGTCGATGA	59.8	95662	95563	95694	loco 125

Anexo 4: Informações adicionais dos locos SSR derivados de transcritos de *M. acuminata* Calcutta 4 durante a interação com *M. musicola* mapeados no genoma de *M. acuminata* DH Pahang.

Loco	Gene NB-ARC DH Pahang mais próximo	Chr	Unigene ID	Distância gene NB-ARC (pb)	Score
454_RES_PAHANG_locus3299	GSMUA_Achr9G02765_001	chr9	C4IGN96RJG01CJOKS 282	196	1,0
454_RES_PAHANG_locus1444	GSMUA_Achr1G12540_001	chr1	<i>Musa_c4_small_c9219</i>	2118	0,7
454_RES_PAHANG_locus3316	GSMUA_Achr5G07080_011	chr5	C4IGN96RJG02I66EQ 437	10061	1,0
454_RES_PAHANG_locus1458	GSMUA_Achr11G07440_001	chr11	<i>Musa_c4_small_c9422</i>	11255	1,0
454_RES_PAHANG_locus1835	GSMUA_Achr11G07440_001	chr11	<i>Musa_c4_small_rep_c1566</i>	11683	0,5
454_RES_PAHANG_locus3952	GSMUA_Achr11G07440_001	chr11	C4NIGN96RJG01CL6I3 318	13220	1,0
454_RES_PAHANG_locus2300	GSMUA_Achr5G07080_010	chr5	<i>Musa_c4_small_rep_c2707</i>	13762	0,8
454_RES_PAHANG_locus407	GSMUA_Achr5G07080_009	chr5	<i>Musa_c4_small_c15503</i>	14413	0,8
454_RES_PAHANG_locus2811	GSMUA_Achr6G32050_001	chr6	<i>Musa_c4_small_rep_c804</i>	14716	1,0
454_RES_PAHANG_locus2810	GSMUA_Achr6G32050_001	chr6	<i>Musa_c4_small_rep_c804</i>	14854	1,0
454_RES_PAHANG_locus513	GSMUA_Achr1G08690_001	chr1	<i>Musa_c4_small_c17577</i>	16957	1,0
454_RES_PAHANG_locus1102	GSMUA_Achr3G09290_001	chr3	<i>Musa_c4_small_c4617</i>	17694	0,8
454_RES_PAHANG_locus2274	GSMUA_Achr3G27820_001	chr3	<i>Musa_c4_small_rep_c2576</i>	18313	1,0
454_RES_PAHANG_locus2273	GSMUA_Achr3G28030_001	chr3	<i>Musa_c4_small_rep_c2576</i>	18498	0,8
454_RES_PAHANG_locus1096	GSMUA_Achr1G12540_001	chr1	<i>Musa_c4_small_c4536</i>	23820	0,8
454_RES_PAHANG_locus3926	GSMUA_Achr5G07080_008	chr5	C4IGN96RJG02HAUDB 213	30545	0,3
454_RES_PAHANG_locus3698	GSMUA_Achr9G15960_001	chr9	C4IGN96RJG02GECMD 340	37877	1,0
454_RES_PAHANG_locus311	GSMUA_Achr11G07440_001	chr11	<i>Musa_c4_small_c14010</i>	39111	1,0
454_RES_PAHANG_locus14	GSMUA_Achr3G27820_001	chr3	<i>Musa_c4_small_c10088</i>	43502	1,0
454_RES_PAHANG_locus3990	GSMUA_Achr10G07065_001	chr10	C4IGN96RJG01AIP0H 456	44645	0,7
454_RES_PAHANG_locus2243	GSMUA_Achr8G03665_001	chr8	<i>Musa_c4_small_rep_c24931</i>	45354	0,7
454_RES_PAHANG_locus549	GSMUA_Achr6G32050_001	chr6	<i>Musa_c4_small_c17975</i>	51757	0,8
454_RES_PAHANG_locus3360	GSMUA_Achr5G07080_007	chr5	C4NIGN96RJG01D7QPM 397	54482	1,0
454_RES_PAHANG_locus1635	GSMUA_Achr4G26175_001	chr4	<i>Musa_c4_small_rep_c12383</i>	55686	1,0

454_RES_PAHANG_locus3561	GSMUA_Achr3G27820_001	chr3	C4NIGN96RJG02H0PT2 328	57050	0,5
454_RES_PAHANG_locus3050	GSMUA_Achr6G32050_001	chr6	<i>Musa_c4_clusters_pos_c1777_1</i>	57839	1,0
454_RES_PAHANG_locus2334	GSMUA_Achr8G03665_001	chr8	<i>Musa_c4_small_rep_c2917</i>	58118	0,8
454_RES_PAHANG_locus1633	GSMUA_Achr4G26175_001	chr4	<i>Musa_c4_small_rep_c12383</i>	58331	0,8
454_RES_PAHANG_locus1634	GSMUA_Achr4G26175_001	chr4	<i>Musa_c4_small_rep_c12383</i>	60380	1,0
454_RES_PAHANG_locus1890	GSMUA_Achr4G26175_001	chr4	<i>Musa_c4_small_rep_c16665</i>	60600	0,8
454_RES_PAHANG_locus3310	GSMUA_Achr1G12540_001	chr1	C4NIGN96RJG02FIOGA 450	63604	0,8
454_RES_PAHANG_locus852	GSMUA_Achr8G03665_001	chr8	<i>Musa_c4_small_c23529</i>	65329	1,0
454_RES_PAHANG_locus4029	GSMUA_Achr4G26175_001	chr4	C4NIGN96RJG02F4R19 209	69056	1,0
454_RES_PAHANG_locus3231	GSMUA_Achr1G23280_001	chr1	C4NIGN96RJG02I06NW 384	69848	0,5
454_RES_PAHANG_locus1101	GSMUA_Achr11G07440_001	chr11	<i>Musa_c4_small_c4602</i>	72394	0,8
454_RES_PAHANG_locus38	GSMUA_Achr1G12540_001	chr1	<i>Musa_c4_small_c10369</i>	73359	0,8
454_RES_PAHANG_locus3080	GSMUA_Achr10G07065_001	chr10	<i>Musa_c4_clusters_pos_c486_1</i>	77816	0,8
454_RES_PAHANG_locus2246	GSMUA_Achr6G32050_001	chr6	<i>Musa_c4_small_rep_c25198</i>	85811	0,5
454_RES_PAHANG_locus82	GSMUA_Achr5G07080_013	chr5	<i>Musa_c4_small_c10992</i>	88123	1,0
454_RES_PAHANG_locus1081	GSMUA_Achr3G09290_001	chr3	<i>Musa_c4_small_c4315</i>	92183	0,7
454_RES_PAHANG_locus1593	GSMUA_Achr11G07440_001	chr11	<i>Musa_c4_small_rep_c11497</i>	94999	0,5
454_RES_PAHANG_locus2809	GSMUA_Achr1G08716_001	chr1	<i>Musa_c4_small_rep_c8036</i>	99102	0,8
454_RES_PAHANG_locus3083	GSMUA_Achr10G07065_001	chr10	<i>Musa_c4_clusters_pos_c534_1</i>	102644	0,8
454_RES_PAHANG_locus2616	GSMUA_Achr4G26175_001	chr4	<i>Musa_c4_small_rep_c530</i>	104075	1,0
454_RES_PAHANG_locus2722	GSMUA_Achr1G08690_001	chr1	<i>Musa_c4_small_rep_c675</i>	104154	1,0
454_RES_PAHANG_locus2641	GSMUA_Achr1G08716_001	chr1	<i>Musa_c4_small_rep_c563</i>	105932	1,0
454_RES_PAHANG_locus2642	GSMUA_Achr1G08690_001	chr1	<i>Musa_c4_small_rep_c563</i>	109404	1,0
454_RES_PAHANG_locus2488	GSMUA_Achr9G02765_001	chr9	<i>Musa_c4_small_rep_c3963</i>	114593	1,0
454_RES_PAHANG_locus1168	GSMUA_Achr1G08690_001	chr1	<i>Musa_c4_small_c5739</i>	114856	1,0
454_RES_PAHANG_locus3160	GSMUA_Achr4G26175_001	chr4	C4IGN96RJG01DQ3JN 433	115595	1,0