Author's post-print of the paper published in the Proceedings of the International ISHS-ProMusa Symposium on Global Perspectives on Asian Challenges, Acta Horticulturae 897, ISHS 2011. Available from www.musalit.org

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. Passos[1], F.L. Emediato[1], C. de Camargo Teixeira[1] and G.J. Pappas Júnior[1,3]

[1] Universidade Católica de Brasília, Brasília, DF, Brazil
[2] Universidade de Brasília, Departamento de Biologia Celular, Brasília, DF, Brazil
[3] EMBRAPA Recursos Genéticos e Biotecnologia, Brasilia, 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 & 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 (Mindrinos 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 guardee 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., 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 hybrization 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 *R*-genes or RGAs	Total number and percentage of sequences with homology to *R*-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 *R*-genes or RGAs	Total number and percentage of sequences with homology to *R*-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%)