IDENTIFICATION VALIDATION AND USE OF EST-DERIVED MOLECULAR MARKERS FROM THE GENOMES OF *Mycosphaerella fijiensis* AND *Musa* spp.

SUZANA ANTUNES LOURENÇONI GARCIA

2009

SUZANA ANTUNES LOURENÇONI GARCIA

IDENTIFICATION VALIDATION AND USE OF EST-DERIVED MOLECULAR MARKERS FROM THE GENOMES OF *Mycosphaerella fijiensis* AND *Musa* spp.

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, para obtenção do título do "Mestre".

Orientador

Dr. Manoel Teixeira Souza Júnior

LAVRAS
MINAS GERAIS - BRASIL
2009

Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da Biblioteca Central da UFLA

Garcia, Suzana Antunes Lourençoni.
Identification, validation and use of EST-derived molecular markers from the genomes of *Mycosphaerella fijiensis* and *Musa* spp. / Suzana Antunes Lourençoni Garcia. – Lavras : UFLA, 2009.
76 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2009.
Orientador: Manoel Teixeira Souza Júnior.
Bibliografia.

1. Banana. 2. *Mycosphaerella fijiensis*. 3. Marcadores moleculares VNTR e TRAP. I. Universidade Federal de Lavras. II. Título.

CDD – 584.21042115
581.2115

SUZANA ANTUNES LOURENÇONI GARCIA

IDENTIFICATION VALIDATION AND USE OF EST-DERIVED MOLECULAR MARKERS FROM THE GENOMES OF *Mycosphaerella fijiensis* AND *Musa* spp.

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, para obtenção do título do "Mestre".

APROVADA em 17 de Março de 2009

Prof. Dr. Luciano Vilala Paiva	UFLA
Prof. Dra. Dulcinéia de Carvalho	UFLA

Dr. Manoel Teixeira Souza Júnior
EMBRAPA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

“Porque dele e por ele, e para ele, são todas as coisas; glória, pois, a ele eternamente. Amém”

Romanos 11. 36

À Deus ofereço

AGRADECIMENTOS

Á Jesus Cristo, mestre da minha vida, meu maior exemplo de amor incondicional, humildade e determinação.

À Universidade Federal de Lavras, em especial ao Programa de Pós-Graduação em Biotecnologia Vegetal pela oportunidade de realização do curso.

Ao coordenador do curso, prof. Dr Luciano Vilela Paiva pelo incentivo, apoio, aprendizado e confiança.

Ao meu orientador Dr. Manoel Teixeira Souza Junior, pela dedicação, ensinamentos, conselhos, amizade e também por ter provido todos os meios necessários para o desenvolvimento desta dissertação no Plant Research International (PRI) da Universidade de Wageningen.

À minha co-orientadora, Cláudia Fortes pela disposição em sempre me ajudar, pelos conselhos e pela amizade.

Ao Dr. Gert kema por todo apoio durante todo o período que eu estive no PRI.

Ao Dr. Adilson Kobayashi, pela amizade, dedicando do seu precioso tempo dando me importantes conselhos referentes ao desenvolvimento da minha dissertação.

Aos professores do curso de Biotecnologia Vegetal.

Aos membros da banca examinadora.

Aos amigos do LCBM e em especial Dani, Felipe, Fabiane, Glacy, Rosane e Evânia pela convivência e aprendizado.

Ás amigas do PRI, Sarah e Caucasella.

À minha querida família, meus pais, irmãos, vovó Queta e Mônica pelo apoio constante e por não medirem esforços em me ajudar.

Ao meu amor Mark Spekken, pelo companheirismo, confiança e amor.

MUITO OBRIGADA

SUMÁRIO

Página

ABSTRACT

GARCIA, Suzana Antunes Lorençoni. **Identification, validation and use of EST-derived molecular markers from the genomes of *Mycosphaerella fijiensis* and *Musa* spp**. 2009. 76p. Dissertation (Master in Agronomy/ Plant Biotechnology) – Federal University of Lavras.*

The present master thesis was performed under the scope of the *Musa*Forever program. This program is a strategic alliance established by Brazilian Agricultural Research Corporation (EMBRAPA), Plant Research International - Wageningen UR (PRI-WUR), the Catholic University of Leuven (KUL), the Federal University of Lavras (UFLA), and the French Agricultural Research Centre for International Development (CIRAD), aiming scientific knowledge and biotechnology tools to optimize the possibilities for genetic control of Black leaf streak and Fusarium wilt in banana. This alliance is a capacity building and technology transfer platform, led by Embrapa and PRI-WUR. The data used in this thesis was obtained during nine months internship in the Biointeractions and Plant Health business unit (http://www.pri.wur.nl/UK/research/Biointeractions/) at the Plant Research International (http://www.pri.wur.nl/UK/), Wageningen University and Research Centre (http://www.wur.nl/UK/). These results are intended to be published as three different manuscripts in scientific journals in the near future; that is the reason for them to be presented already in this thesis as manuscript (Chapter 2, 3 and 4).

Guidance Committee: Dr. Manoel Teixeira Souza Júnior – EMBRAPA (Major Professor), Luciano Vilala Paiva – UFLA and Dulcinéia de Carvalho -UFLA

RESUMO

GARCIA, Suzana Antunes Lorençoni. **Identificação, validação, e uso de marcadores moleculares a partir de sequencias de DNA dos genomas de *Mycosphaerella fijiensis* and *Musa* spp.** 2009. 76p. Dissertação (Mestrado em Agronomia/ Biotecnologia Vegetal) – Universidade Ferderal de Lavras.*

A presente dissertação de mestrado foi realizada dentro do escopo do programa *Musa*Forever. Este programa é uma aliança estratégica estabelecida pela Empresa Brasileira de Pesquisa e Agropecuária (EMBRAPA), Instituto Internacional de Pesquisa em Planta - Wageningen UR (PRI-WUR), Universidade Católica de Leuven (KUL), Universidade Federal de Lavras (UFLA), e Centro Francês de Investigação para o Desenvolvimento (CIRAD), visando o conhecimento científico e o estabelecimento de ferramentas de biotecnologia para otimizar as possibilidades de controle genético da Sigatoka Negra e Fusarium em bananeiras. Esta aliança visa reforçar a capacidade de transferência de tecnologia, liderada pela Embrapa e PRI-WUR. Os dados utilizados nesta dissertação foram obtidos durante nove meses de estágio no laboratório Biointeraction (http://www.pri.wur.nl/UK/research/Biointeractions/) do Instituto de Pesquisa Internacional em Planta (PRI) (http:// www.pri.wur.nl/UK/), Universidade de Wageningen e Centro de Pesquisa (http://www.wur.nl/UK/). Os resultados obtidos geraram três manuscritos que serão publicados em três revistas científicas diferentes no futuro próximo, sendo essa a razão pela qual eles estão sendo apresentados nesta dissertação na forma de manuscritos (Capítulos 2, 3 e 4).

Comitê Orientador: Dr. Manoel Teixeira Souza Júnior – EMBRAPA (Orientador), Luciano Vilala Paiva – UFLA e Dulcinéia de Carvalho -UFLA

CHAPTER 1

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Transcriptome, EST database, Data*Musa* and EST-derived molecular markers

The transcriptome, defined as the complete set of RNA transcripts produced by the genome at any time, is dynamic and changes under different temporal and/or spatial circumstances due to different patterns of gene expression. Transcriptomics is the name given to the study of an organism's transcriptome (Roux et al., 2008).

The production of an EST (Expressed Sequence Tags) database is the initial step in the transcriptomics of a given organism. An EST is a unique DNA sequence derived from a cDNA library (therefore from a sequence which has been transcribed in some tissue or at some stage of development) and can be derived from a transcribed protein-coding or non-protein-coding nucleotide sequence (Roux et al., 2008).

Expressed Sequence Tags are created by one shot sequencing of the 5`and/or 3` ends of a cloned mRNA (i.e. sequencing several hundred base pairs from one or both ends of cDNA clones taken from a cDNA library isolated gene transcripts that have been converted into cDNA). The resulting sequence is a relatively low quality fragment whose length is limited by approximately 500 to 800 nucleotides. Because these clones consist of DNA that is complementary to mRNA, the ESTs represent expressed portions of genes (Adams et al., 1991).

Despite the fact that a typical EST represents only a portion of a coding sequence, this partial sequence data is of substantial utility. For example, EST collections are a relatively quick and inexpensive route for discovering new genes (Bourdon et al., 2002), confirm coding regions in genomic sequence, create opportunities to elucidate phylogenetic relationships (Nishiyama et al., 2003), facilitate the construction of genome maps (Paterson et al., 2000), can

sometimes be interpreted directly for transcriptome activity (Ewing et al., 1999; Ogihara et al., 2003), and provide the basis for development of expression arrays also known as DNA chips (Chen et al., 1998). In addition, high-throughput technology and EST sequencing projects can result in the identification of significant portions of an organism's gene content and thus can serve as a foundation for initiating genome sequencing projects (Hoeven et al., 2002).

The term "EST" has been used to describe genes for which no further information exists besides the tag. Often, they are expressed only in certain tissues at certain points in time however, some of them can be constitutively expressed (Venter, 1993).

The production of a banana EST database is one of the strategies used by the Global *Musa* Consortium (GMGC) for the characterization of the genome of this socially and economically important fruit crop (Consortium, 2001). The first known study on banana transcriptome was performed before the creation of the GMGC, by Syngenta, which produced and characterized two cDNA libraries of the Cavendish variety Grand Naine (AAA). These two cDNA libraries are available to members of the GMGC through signing a material transfer agreement.

Another banana transcriptomics study was a collaborative effort between EMBRAPA, the Catholic University of Brasilia (UCB) and the Center for Agriculture Research and Development (CIRAD), France, with the support of the National Council for Scientific and Technological Development (CNPq) in Brazil. This common effort enabled the production of the DATA*Musa.* The DATA*Musa* is a banana genomics database that contains information on structural genomic, transcriptome and resistance gene analogs. The transcriptome data in the DATA*Musa* database derived from the production and characterization of seven cDNA libraries, which were produced in order to discover genes expressed in leaves submitted to temperature stress, leaves in

early and late stages of infection by the pathogen *Mycosphaerella fijiensis*, male flowers, roots of *in vitro* plants and green peel of *M. acuminata* (A genome) (Souza et al., 2005).

EMBRAPA also produced and characterized, with the support of the Generation Challenge Programme, two cDNA libraries of the Pisang Klutuk (PKW) variety, an *M. balbisiana* diploid banana (B genome) (Generation Challenge Programme, 2006). So far, the transcriptome database from bananas (*Musa* spp.), DATA*Musa*, containing 42,724 ESTs from 11 different cDNA libraries, and encompassing approximately 24 Mb of DNA sequence. This database comprises a rich source for identification of EST-derived molecular markers.

Molecular markers detect DNA sequence variation among genotypes. DNA markers are found in abundance and are not influenced by the environment or plant developmental stages, making them ideal for genetic relationship studies (Reddy et al., 2002). Most molecular markers are related to genomic DNA, and therefore could belong to either the transcribed region or the non-transcribed region of the genome and have been described as random DNA markers (Anderson, 2003).

However the development of molecular markers from the transcribed region of the genome has become a very important tool for genetic relationship studies. The transcriptome-based molecular markers are able to detect the length of the polymorphism in the expressed sequence region of the genome. The molecular markers from the expressed region of the genome can be obtained using different sources; mRNA, cDNA, and ESTs (Gupta & Rustgi, 2004). It is possible to develop cDNA-AFLP, cDNA-SSR, and synthetic RFLP probes in order to detect cDNA-RFLP (Gupta & Rustgi, 2004) using cDNA or mRNA. cDNA-AFLP method is highly sensitive and is very specific in detecting poorly expressed genes and distinguishing between homologous sequences. Accurate

gene expression profiles can be determined by quantitative analysis of band intensities, and subtle differences in transcriptional activity can be revealed (Breyne et al., 2003). SSR-containing cDNAs are conserved across the species and consequently are very useful for comparative mapping of transcribed (Saha et al., 2003) regions. RFLP probes derived from cDNA clones has been shown to be efficient in identifying the location of specific genes in molecular mapping (Singh et al., 2008). Expressed sequence tag (EST) databases represent a potentially valuable resource for the development of simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs) both useful markers for creating genetic maps in plants (Morgante et al., 2002; Rafalski, 2002). SSRs are particularly common in the 5`-untranslated region (UTR) and, to a lesser extent, in the 3`-UTR of transcribed plant sequences. Moreover, it has been shown that EST-based SSR markers (EST-SSRs) can be rapidly and inexpensively developed from existing EST databases (Gupta & Rustgi, 2003; Bhat et al., 2005).

The primary limitation of using EST databases as a source of molecular markers is that this approach relies on existing genomic resources, and suitable databases are often only available to researchers who are studying economically important species (Pashley et al., 2006). Adding to this difficulty is the fact that only a fraction of all ESTs contain SSRs. For example, it has been suggested that the frequency of SSR-containing sequences in plant-derived EST databases is typically on the order of 2–5% (Kantety et al., 2002). Once developed, however, EST-SSRs are likely to be useful across a much broader taxonomic range than are anonymous SSRs owing to the fact that the former come exclusively from transcribed regions of the genome. It is therefore possible that these markers will prove to be particularly valuable for the investigation of population genetic phenomena in close relatives of species with existing genomic resources (Pashley et al., 2006).

Mycosphaerella fijiensis, Black sigatoka disease and Molecular markers

Mycosphaerella fijiensis (*Pseudocercospora fijiensis; Mycosphaerellaceae*) belongs to the genus *Mycosphaerella*, which is one of the largest genera of ascomycetes, comprising several thousands of species (Crous, 1998). It is also an heterothallic filamentous fungus, which can be cultured on both liquid and solid media (Mourichon & Zapater, 1990).

M. fijiensis is the causal agent of black Sigatoka or black leaf streak disease (BLSD), the most economically important banana disease (*Musa* spp.) in tropical and subtropical areas which is rapidly becoming the most devastating disease of banana production worldwide. It causes leaf necrosis and plant defoliation, resulting in yield losses and poor fruit quality due to a decrease in photosynthesis, reduction of fruit size and induction of premature maturation (Marin et al., 2003).

BLSD was first discovered on the South-eastern coast of Viti Levu in Fiji in 1963 (Rhodes, 1964). Subsequently, the disease was reported in the Pacific Islands, Asia, Africa and finally in Latin America in 1972 in La Lima, Honduras (Stover, 1976).

The genetic structure of *Mycosphaerella fijiensis* populations around the world was examined using DNA restriction fragment length polymorphism (RFLP) markers (Carlier et al., 1996). *M. fijiensis* isolates from five geographical populations, representative of different banana-producing areas (South-East Asia including the Philippines and Papua New Guinea, Africa, Latin America and Pacific Islands), were evaluated. Southeast Asia had the highest level of genetic diversity, supporting the hypothesis that the pathogen originated in the region. In the Latin America and Caribbean regions, the highest levels of genetic diversity were observed in populations from Honduras and Costa Rica, corroborating to hypothesis that the pathogen entered the continent in this area. In Africa, the levels of genetic diversity in most countries were similar, making

it difficult to locate the place(s) where the pathogen first entered the continent[31]. Hence, RFLP markers were very useful tools for detecting genetic variation in *M. fijiensis* isolates from different parts of the world.

Afterwards, PCR-RFLP and microsatellite molecular markers were used in order to infer gene flow and dispersal processes of *M. fijiensis* at the continental scale of population structure analysis from Latin America and the Caribbean and in Africa (Rivas et al., 2004).

Microsatellites markers (SSR) were also used to study the population genetics of *M. fijiensis* from Mexico and Nigeria (Neu et al., 1999) and from Colombia (Perea et al., 2005).

In 2007, a ~7.1X equivalent of the estimated 73.4 Mbp *M. fijiensis* (isolate CIRAD086) genome (http://genome.jgi-psf.org/Mycfi1/Mycfi1.home.html) was produced at the Joint Genome Institute of the United States Department of Energy (http://www.jgi.doe.gov/). In addition to that, a *M. fijiensis* EST sequencing project was also done at JGI-DOE, which has resulted in more than 30,000 ESTs (Kema, 2007). The development of *M.fijiensis* SSR markers has been possible due to Perl script based pipeline (Yang & Zhong, 2007).

The *M. fijiensis* sequencing databases provides a valuable source of data for the identification of molecular markers associated to black leaf streak disease (BLSD).

References

ADAMS, M.; KELLEY, J.; GOCAYNE, J. Complementary DNA sequencing: expressed sequence tags and human genome project. **Science**, New York, v.252, n.21, p.1651-1656, June 1991.

ANDERSON, J.R.L. Functional markers in plants. **Trends in Plant Science**, Oxford, v.8, n.12, p.554-560, Dec. 2003.

BHAT, P.R.; KRISHNAKUMAR, V.; HENDRE, P.S.; RAJENDRAKUMAR, P.; VARSHNEY, R.K.; AGGARWAL, R.K. Identification and characterization of expressed sequence tags-derived simple sequence repeats, markers from robusta coffee variety 'C x R': an interspecific hybrid of *Coffea canephora* x *Coffea congensis*. **Molecular Ecology Notes**, Oxford, v.5, n.1, p.80-83, Mar. 2005.

BOURDON, V.; NAEF, F.; RAO, P.; REUTER, V.; MOK, S.; BOSL, G.; KOUL, S.; MURTY, V.; KUCHERLAPATI, R.; CHAGANTI, R. Genomic and expression analysis of the 12p11-p12 amplicon using EST arrays identifies two novel amplified and over expressed genes. **Cancer Research**, Baltimore, v.62, n.21, p.6218-6223, Dec. 2002.

BREYNE, P.; DREESEN, R.; CANNOO, B.; ROMBAUT, D.; VANDEPOELE, K.; ROMBAUTS, S.; VANDERHAEGHEN, R.; INZE, D.; ZABEAU, M. Quantitative cDNA-AFLP analysis for genome-wide expression studies. **Molecular Genetics and Genomics**, Berlin, v.269, n.2, p.173-179, May 2003.

CARLIER, J.; ZAPATER, M.F.; DUBOIS, C.; MOURICHON, X. Genetic structure of the global population of banana black leaf streak fungus *Mycosphaerellla fijiensis.* **Molecular Ecology**, Oxford, v.5, n.4, p.499-510, Aug. 1996.

CHEN, J.; WU, R.; YANG, P.C. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. **Genomics**, San Diego, v.51, n.3, p.313-324, Aug. 1998.

CONSORTIUM, G.M.G. **Beyond arabidopsis and rice**: strategy for the global musa genomics consortium. Arlington: INIBAP, 2001. 30p.

CROUS, P.W. ***Mycosphaerella*** **spp. and their anamorphs associated with leaf spot diseases of Eucalyptus**. Saint Paul: APS, 1998. 170p. (Mycologia Memoir, 21).

EWING, R.; KAHLA, A.; POIROT, O.; LOPEZ, F.; AUDIC, S.; CLAVERIE, J. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. **Genome Research**, Cold Spring Harbor, v.9, n.10, p.950-959, Oct. 1999.

GENERATION CHALLENGE PROGRAMME. **Generation challenge programme competitive and commissioned research project mid-year reports**. Mexico, 2006. No page.

GUPTA, P.K.; RUSTGI, S. Molecular markers from the transcribed/expressed region of the genome in higher plants. **Functional and Integrative Genomics**, Berlin, v.4, n.3, p.139-162, July 2004.

HOEVEN, R. van der; RONNING, C.; GIOVANNONI, J.; MARTIN, G.; TANKSLEY, S. Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large EST collection and selective genomic sequencing. **Plant Cell**, Rockville, v.14, n.7, p.1441-1456, Sept. 2002.

KANTETY, R.V.; ROTAL, M.L.; MATTHEWS, D.E.; SORRELLS, M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. **Plant Molecular Biology**, Dordrecht, v.48, n.4, p.501-510, July 2002.

KEMA, G.H.J. Sequencing the major *Mycosphaerella* pathogens of wheat and banana. In: INTERNATIONAL SYMPOSIUM ON RECENT ADVANCES IN BANANA CROP PROTECTION FOR SUSTAINABLE PRODUCTION AND IMPROVED LIVELIHOODS, 15., 2007, San Diego. **Proceedings**... San Diego: ISHS, 2007. 1 CD-ROM.

MARIN, D.; ROMERO, R.; GUZMÁN, M.; SUTTON, T.B. Black sigatoka: an increasing threat to banana cultivation. **Plant Disease**, Quebec, v.87, n.3, p.208-222, Mar. 2003.

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with non repetitive DNA in plant genomes. **Nature Genetics**, New York, v.30, n.2, p.194-200, Feb. 2002.

MOURICHON, X.; ZAPATER, M.F. Obtention in vitro du stade *Mycosphaerella fijiensis*, forme parfaite de *Cercospora fijiensis*. **Fruits**, Paris, v.45, n.6, p.553-557, 1990.

NEU, C.; KAEMMER, D.; KAHL, G.; FISCHER, D.; WEISING, K. Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. **Molecular Ecology**, Oxford, v.8, n.3, p.523-525, Mar. 1999.

NISHIYAMA, T.; FUJITA, T.; SHIN-I, T. Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution. **Proceedings National Academy of Science**, Washington, v.100, n.13, p.8007-8012, June 2003.

OGIHARA, Y.; MOCHIDA, K.; NEMOTO, Y.; MURAI, K.; YAMAZAKI, Y.; SHIN-I, T.; KOHARA, Y. Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. **Plant Journal**, Oxford, v.33, n.6, p.1001-1011, Mar. 2003.

PASHLEY, C.H.; ELLIS, J.R.; MCCAULLEY, D.E.; BURKE, J.M. EST databases as a source for molecular markers: lessons from helianthus. **Journal of Heredity**, Washington, v.97, n.4, p.381-388, July 2006.

PATERSON, A.; BOWERS, J.; BUROW, M. Comparative genomics of plant chromosomes. **Plant Cell**, Rockville, v.12, n.9, p.1523-1540, Sept. 2000.

PEREA, I.; ARANGO, E.R.; MÁRQUEZ, E.; ARANGO, R. Genetic diversity of Colombian isolates of Mycosphaerella fijensis Morelet based on microsatellite markers. **InfoMusa**, Montpellier, v.14, n.2, p.1-4, Dec. 2005.

RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Cell Biology**, London, v.5, n.2, p.94-100, Apr. 2002.

REDDY, P.M.; SARAL, N.; SIDDIQ, E.A. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. **Euphytica**, Wageningen, v.128, n.1, p.9-17, Nov. 2002.

RHODES, P.L. A new banana disease in Fiji. **Phytopathological News**, Kew, v.10, p.38-41, 1964.

RIVAS, G.G.; ZAPATER, M.F.; ABADIE, C.; CARLIER, J. Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas Mycosphaerella fijiensis. **Molecular Ecology**, Oxford, v.13, n.2, p.471-482, Feb. 2004.

ROUX, N.; BAURENS, F.C.; DOLEZEL, J.; HRIBOVA, E.V.A.; HARRISON, P.H.; SASAKI, T.; MATSUMOTO, T.; AERT, R.; REMY, S.; SOUZA, M.; LGODA, P. Genomis of banana and plantain (*Musa* spp), major stable crops in the tropics. **Genomic of Tropical Crop Plants**, New York, v.1, n.1, p.83-111, Jan. 2008.

SAHA, S.; KARACA, M.; JENKINS, N.J.; ZIPF, A.E. Simple sequence repeats as useful resources to study transcribed genes of cotton. **Euphytica**, Wageningen, v.130, n.3, p.355-364, Apr. 2003.

SINGH, R.; SOON-GUAN, T.; PANANDAM, J.; RAHMAN, R.A. Identification of cDNA-RFLP markers and their use for molecular mapping in oil palm (*Elaeis guineensis*). **Asia Pacific Journal of Molecular Biology & Biotechnology**, Southampton, v.16, n.3, p.53-63, 2008.

SOUZA, M.T.; SANTOS, C.M.; MARTINS, N.F.; SILVA, F.R. Transcriptoma de Musa acuminata no DATA*Musa*. **Embrapa Recursos Genéticos e Biotecnologia**, Brasília, v.109, p.1-25, out. 2005.

STOVER, R.H. Banana leaf spot caused by *Mycosphaerella musicola* and *Mycosphaerella fijiensis* var *difformis*: a comparison of the first central American epidemics. **FAO Plant Protection Bulletin**, Rome, v.24, p.36-42, 1976.

STOVER, R.H. International spread of banana leaf spot (*Mycosphaerella fijinensis* Leach). **Tropical Agriculture**, Surrey, v.39, p.327-338, 1962.

VENTER, J.C. Identification of a new human receptor and transporter genes by high throughput cDNA EST sequecing. **Journal of Pharmacy and Pharmacology**, London, v.45, n.1, p.355-360, 1993. Supplement.

YANG, B.J.; ZHONG, S.B. Fourteen polymorphic microsatellite markers for the fungal banana pathogen *Mycosphaerella fijiensis*. **Molecular Ecology Resources**, Oxford, v.8, n.5, p.910-912, Mar. 2007.

CHAPTER 2

IDENTIFICATION AND VALIDATION OF EST-DERIVED MOLECULAR MARKERS, TRAP AND VNTRS, FOR BANANA RESEARCH.

ABSTRACT

The advent of high-throughput sequencing technology has generated abundant information on DNA sequences for the genomes of many plant species. Expressed Sequence Tag (EST), a unique DNA sequence derived from a cDNA library, and therefore representing a gene which has been transcribed in a specific tissue or at some stage of development, is one type of DNA sequence made highly available lately for many important crop species. Molecular markers are used for bridging DNA sequence information with particular phenotypes, and are useful tools for genotyping germplasm collections and also for tagging genes involved in desirable agronomic traits. In this sense, there is always a strong demand for suitable marker techniques to better utilize the existing sequence information. A transcriptome database from banana (*Musa* spp.), DATA*Musa*, containing 42,724 ESTs from 11 different cDNA libraries, encompassing approximately a 24 Mb-DNA sequence, was used in this study for primers' design to amplify two types of EST-derived molecular markers, Variable Nucleotide Tandem Repeat (VNTR) and Target Region Amplification Polymorphism (TRAP). These primers were then validated against a panel of 14 *Musa* diploid genotypes, and produced 32 (VNTR) and 119 (TRAP) alleles. Either separately or in combination, both types of markers were able to discriminate *Musa* genotypes from different genome backgrounds (A or B genomes). The identified TRAP alleles derived from only one unigene, while the VNTR derived from 12 unigenes. Based on the results from this study, it can be said that EST-derived markers are an important source of polymorphism to be used in genetic diversity and gene discovery studies in banana.

RESUMO

Com o surgimento de novas tecnologias de sequenciamento, têm sido geradas abundantes informações sobre sequências de DNA do genoma de muitas espécies de plantas. Sequências de EST (Expressed Sequence Tags), são sequências de DNA fita simples derivadas de uma biblioteca de cDNA, e, portanto podem representar um gene que foi transcrito em um tecido específico ou em algum estágio de desenvolvimento. É um tipo de sequência de DNA recentemente disponível para estudos de muitas espécies cultivadas importantes. Os marcadores moleculares são utilizados para gerar informações de sequências de DNA com fenótipos particulares, e são ferramentas úteis para genotipagem de coleções de germoplasma e também para marcação de genes envolvidos em características agronômicas desejáveis. Nesse sentido, há sempre uma grande demanda por técnicas de marcadores moleculares adequados para utilizar melhor as informações da sequência de DNA. Um banco de dados do transcriptoma de bananeira (*Musa* spp.) DATA*Musa*, contendo 42.724 ESTs de 11 diferentes bibliotecas de cDNA, abrangendo cerca de 24Mb sequências de DNA, foi utilizado neste estudo para o desenho de iniciadores para amplificar dois tipos de marcadores moleculares derivados de EST: Variable Nucleotide Tandem Repeat (VNTR) e Target Region Amplification Polymorphism (TRAP). Estes primers foram validados em um painel de 14 genótipos diplóides de *Musa*, e produziram 32 marcadores VNT e 119 TRAP.

Introduction

Banana (*Musa* spp.) is the world's largest perennial monocotyledon, belongs to the *Musaceae* family, and has a relative small genome (~ 600 Mb) distributed over eleven chromosomes. *Musa acuminata* (A genome) and *M. balbisiana* (B genome), the two major *Musa* species, are at the origin of edible bananas. These bananas comprise a diverse group, including cooking types such as plantains and a wide range of dessert types, derived basically from intra-specific crosses among several subspecies of *M. acuminata*, as well as from hybrids with *M. balbisiana* (Stover & Simmonds, 1987; Simmonds, 1995). Traditionally, the classification of the banana genome has been based on its phenotype and its similarity to *M. acuminata* and *M. balbisiana* (Simmonds & Shepard, 1995).

Most current cultivars are triploid ($2n = 3X = 33$), with a few parthenocarpic diploid and tetraploid varieties of regional importance, differing in genomic composition and ploidy level, with the main genomic groups being AA, AAA, AAB, ABB and AAAB (Simmonds, 1995).

Genetic analysis, propagation and conservation studies in *Musa* are of utmost importance, since the commercial banana varieties have sterile nature and consequently are less adapted to environmental changes, being continuously under the extinction threat (Venkatachalam & Bhagyalakshmi, 2008).

Banana breeding program worldwide is largely dependent on ploidy manipulation through interspecific and interploidy hybridization (Ortiz et al., 1995). Additionally, *Musa* breeding is time consuming and expensive due to the large amounts of space required for the cultivation of this crop (6 m^2 per plant) and their long growth cycles (10 to 18 months). The combination of these factors suggests that the pace and efficiency of banana genetic improvement could be greatly enhanced through the use of molecular markers (Crouch et al., 1999).

DNA markers are found in abundance and are not influenced by the environment or plant developmental stages, making them ideal for genetic relationship studies (Reddy et al., 2002). The development and application of technologies based upon molecular markers provides powerful tools that are adequate enough to reveal polymorphisms at the DNA sequence level, and are often robust to detect genetic variability within the populations and among individuals (Kresovich et al., 1995; Simmons et al., 2007).

Many studies of the *Musa* genetic diversity have been conducted by applying different molecular markers including isozyme polymorphism (Bhat et al., 1992), Restriction Fragment Length Polymorphisms (RFLP) (Jarret et al., 1992), Random Amplified Polymorphic DNA (RAPD) (Guimarães et al., 2009). Amplified fragments length polymorphism (AFLP) (Wong et al., 2001), and Variable Number Tandem Repeats or Single Sequence Repeat (SSR) (Kaemmer et al., 1997; Crouch et al., 1999; Guimarães et al., 2009).

Among these molecular markers, SSR or VNTR markers have been demonstrated be the most appropriate for *Musa* due to their reproducibility, co-dominant inheritance, multiallelic nature and wide genome coverage; therefore being the marker of choice for breeders (Varshney et al., 2002; Ning et al., 2007).

The variable number tandem repeats (VNTR) markers, also known as short tandem repeats (STR) or Single Sequence Repeats (SSR), consists of a specific sequence of nucleotides, typically 1-6, repeated in tandem, being found in coding as well as in non-coding DNA regions of the genome. Like all genetic loci, SSRs are subject to point mutation, but slipped-strand mispairing at meiosis or during DNA replication also may occur, changing the number of repeat units. This process can generate a large number of alleles at a single microsatellite locus, each differing by one or more copies of the same repeat unit (Levinson & Gutman, 1987; Lai & Sun, 2003).

Even though the usefulness of SSRs in population genetic studies is well established (Holton, 2001), the isolation and characterization of such markers via traditional methods (i.e., the screening of size-fractionated genomic DNA libraries) are costly and time consuming (Squirrell et al., 2003). Over the past decade, however, there has been a tremendous increase in the accessibility of DNA sequence data from a wide variety of taxa, including a wealth of expressed sequence tags (ESTs) that are typically unedited, automatically processed, single-pass sequences produced from cDNAs. Moreover, it has been shown that EST-based SSR markers (EST-SSRs) can be rapidly and inexpensively developed from existing EST databases (Gupta et al., 2003). Thus, the use of such databases for marker development appears to be a promising alternative to the development of traditional ‘‘anonymous’’ SSRs following standard methods.

There are a number of advantages by using expressed genes as compared to anonymous for use as genetic markers. First, if an EST marker may be associated to a trait of interest, it may be possible that this could be the gene affecting the trait directly (Thiel et al., 2003). Therefore, EST-derived markers can provide opportunities for gene discovery and enhance the role of genetic markers by assaying variation in transcribed and known function genes. Second, EST-derived markers are likely to be more conserved and therefore may be more transferable among species than anonymous sequence derived markers (Cordeiro et al., 2001; Decroocq et al., 2003). Third, ESTs that share homology candidate genes can be specifically targeted for genetic mapping and can be useful for aligning genome linkage across distantly related species for comparative analysis (Holton et al., 2002).

TRAP, known as target region amplification polymorphism is based on a simple and rapid PCR-based marker system. It was developed using EST information and a bioinformatics approach to generate polymorphic markers

around targeted candidate gene sequence. TRAPs are amplified by one fixed primer designed from a known Expressed Sequence Tag (EST) and by a second primer of arbitrary sequence with either an AT- or a GC-rich core to anneal with an intron or exon, respectively. The arbitrary primers are 5'end-labeled with IR dye 700 or IR dye 800 for self-detection of the amplified fragments with the Global DNA Sequencer (Li-Cor Biosciences) (Hu & Vick et al., 2003).

The TRAP technique was initially developed with sunflower (*Helianthus annus* L.) DNA samples and generated amplified fragments with sizes ranging from 50-900 pb (Hu et al., 2003). TRAP markers also have proven to be reliable and reproducible in plant species such as lettuce (*Lactua sativa* L.) cultivars (Hu et al., 2005), sugarcane (*Saccharum* L.ssp) (Alwala et al., 2006), and common beans (*Phaseouls vulgaris* L.) (Miklas et al., 2006).

A transcriptome database from banana (*Musa* spp.), DATA*Musa*, containing 42,724 ESTs from 11 different cDNA libraries, and encompassing approximately a 24 Mb-DNA sequence, was used in this study to design of primers for PCR amplification of VNTR and TRAP markers EST-derived.

In this report we assessed the feasibility of TRAP markers in a panel 14 banana diploid varieties, representing both the A and B *Musa* genomes and it was identified and validated EST-derived VNTR markers using the same panel. This is the first study showing the possibility in using EST-sequence for generating VNTR markers, and the application of the TRAP technique in *Musa* genotypes.

Material and methods

a) Plant Material

A panel of fourteen *Musa* genotypes, obtained at International Institute of Tropical Agriculture (IITA) breeding program, were used in this study; being six genotypes from the *Musa balbisiana* species (B genome) and eight from the *Musa acuminata* species (A genome) (TABLE 1).

Species	Cultivar name	Genome
1 - *M. balbisiana*	Singapuri	BB
2 - *M. balbisiana*	Los Banos	BB
3 - *M. balbisiana*	Tani	BB
4 - *M. balbisiana*	Butohan	BB
5 - *M.acuminata*	Pisang Lillin	AA
6 - *M. acuminata*	Calcutta 4	AA
7 - *M. balbisiana*	Etikehel	BB
8 - *M.acuminata*	Truncata BS-252-A	AA
9 - *M. balbisiana*	Montpellier	BB
10 - *M. acuminata*	Borneo	AA
11 - *M. acuminata*	Madang	AA
12 - *M. acuminata*	Pisang Mas	AA
13 - *M. acuminata*	Selangor	AA
14 - *M. acuminata*	Zebrina	AA

TABLE 1 Fourteen *Musa* spp. diploid varieties, representing both the A and B genome.

b) DNA isolation

Two grams of the cigar leaf were used to extract the DNA using the SDS method (Dellaporta et al., 1983). Due to the high content of poliphenols in banana leaves, the following modifications were introduced: the extraction buffer was supplemented with 2% PVP, and 4 uL of beta-mercaptoethanol (Sigma_Aldrich, St-Louis, USA) was added to each sample just after adding the extraction buffer to the powdered sample. The DNA was later precipitated by the isopropanol procedure.

c) VNTR identification and primer design

Data obtained from the DATA*Musa* database were submitted to analysis in a bioinformatics pipeline for VNTR detection. This pipeline consisted of a primer development module which choped genomic sequences into 1,500 bp size fragments with a 100 bp overlap, then detects VNTRs using a Perl script allowing imperfect repeat VNTR detection by using the scan-for-matches program (Dsouza et al., 1997). Finally, it generateed a maximum of five primer sets by using Primer3 software (Rozen & Skaletsky, 2000) that could amplify a VNTR by PCR followed by a screening for unique primer pairs. Additional optimization of the primer pairs was performed using the Fast PCR Professional 5.2 software (http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). All the primers were synthesized by Sigma-Aldrich (U.S.A).

d) VNTR PCR amplification

Each PCR reaction was performed in a 25 µL total volume, containing 50ng of template genomic DNA, 2mM $MgCl_2$, 600 µM dNTPs, 5 µM each primer, 0.4 U of Taq-DNA polymerase (Roche, Mannhein, German). Temperature cycling was conducted with the following program : 94^{o}C for 2 min., 13 cycles of 94^{o}C for 30 s, 66^{o}C for 30 s (-1^{o}C per cycle) and 72^{o}C for 30

s, followed by 28 cycles of 94^{o}C for 30 s, 53^{o}C for 30 s and 72^{o}C for 30 s and a final elongation period of 7 min. at 72^{o}C.

PCR products were separated by electrophoresis using 3.0% agarose gels containing 0.3 µg/mL ethidium bromide, in 0.5 x TBE buffer, at 120 V, for approximately 5 h. Amplification products were visualized using UV transilluminator and photographed using a Eagle Eye II still video system.

e) TRAP Primer design

TRAP marker analysis was conducted as described by Hu & Vick (2003). A single fixed primer in combination with 2 arbitrary primers, each labeled with an infrared dye (IR-700 or IR-800), were used in the PCR reactions.

Three fixed primers were designed against the EST MUC4FL1018_H03 at DATA*Musa* database; *Musa*.1 sequence forward 5' GGGAGGCCTTGATCTCGT 3', *Musa*.2 sequence forward 5' GCTGGCCAGATGGTGAAC 3', *Musa*.3 sequence forward 5' GAGGAGGACCGCAGATGA 3'.

The fixed primers were selected by using the web-based PCR primer designing program "Primer 3" (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi) (Rozen & Skaletsky, 2000) with the following parameters: primer optimum, maximum and minimum T_m at 53, 55 and 50^{o}C, respectively.

For the development of the arbitrary primers, the general principles of PCR primer design were upheld so as to avoid self-complementarities and improper GC content (40-60%). Three principals were considered in the construction of each arbitrary primer: (1) the selective nucleotides were 3 to 4 nucleotides at the 3` end, (2) the "core" consisted of 4 to 6 nucleotides with AT or GC rich regions, and (3) a filler sequence making up the 5` end. In addition,

the arbitrary primers were 3'end-labeled with IR dye 700 or 800 for self-detection of the amplified fragment with the Li-Cor Global DNA Sequencer (Li-Cor Biosciense, Lincoln, NE) (Hu & Vick, 2003). A total of four arbitrary primers were used in this study; T03 sequence reverse 5' CGTAGCGCGTCAATTATG 3', 5'end-labeled with IR dye 700, T04 sequence reverse 5' CGTAGTGATCGAATTCTG 3', 5'end-labeled with IR dye 700, T13 sequence reverse 5' GCGCGATGATAAATTATC 3', 5'end-labeled with IR dye 800, T14 sequence reverse 5' GTCGTACGTAGAATTCCT 3', 5'end-labeled with IR dye 800.

f) TRAP PCR amplification and visualization

For the Li-Cor Genotyper system, PCR was conducted with a final reaction volume of 10 µL in 96-well microtiter plates in a GanAmp 9700 Thermal Cycler MJ Research PTC-200 Peltier with the following components : 50ng of DNA sample, 1.0 µL of 10X reaction buffer, 1.0 µL of 25mM $MgCl_2$, 0.1 µL of 10mg/mL BSA, 0.4 µL of 5mM dNTP, 2 pmol each of 700- and 800 IR dye labeled arbitrary primers, 10 pmol of the fixed primer, and 0.5 units of *Taq* DNA polimerase (SuperTaq). Temperature cycling was conducted with the following program : 94^{o}C for 2 min., 1 cycle of 94^{o}C for 45 s, 35^{o}C for 45 s and 72 ^{o}C for 1 min. , followed by 35 cycles of 94^{o}C for 45 s, 35^{o}C for 45 s and 72^{o}C for 1 min.; plus an extra elongation period of 7 min. at 72^{o}C.

Afterwards, the reaction products were mixed with an equal volume (10 l) of formamide-loading buffer (98% formamide, 10 mM EDTA pH8.0 and 0.1 % Bromo Phenol Blue). The total mixture was carefully vortexed and heated for 5 minutes at 94 C in the denaturation hotblock and then quickly cooled in ice. A 0.5 l aliquot was loaded into a 6.5% denaturing polyacrylamide sequencing gel in a Li-Cor Global DNA Genotyper. The sequencing gel was prepared using protocols recommended by the manufacturer (Li-Cor Biosciences).

Electrophoresis was conducted at 1,500 V for 3.5 h and the images were collected by SAGA software (Li-Cor Biosciense).

g) Scoring and analyzing the amplified fragments

The polymorphic fragments generated by the VNTR and TRAP markers were visually scored from the printed images. The scoring codes were (1) for present, (0) for absent, and (9) for missing data. The Jaccard`s coefficient was used to calculate the pairwise genetic dissimilarity matrices using the GENES software (Cruz, 2001). The genetic dissimilarity matrices were then used to construct the dendrogram with the UPGMA (Unweighted Pair Group Method with Arithmetic mean) employing the STATISTICA software (Statistica for Windows, 2002). In order to evaluate the reliability of the clusters formed, the dataset was also submitted to bootstrapping with 300, 1000 and 100 simulations for the VNTR, TRAP and VNTR/TRAP data, respectively.

Results and discussion

1 - VNTR marker analysis

Microsatellite markers, developed from genomic libraries, can belong to either the transcribed region or the nontranscribed region of the genome, and rarely is there information available regarding their functions. By contrast, microsatellite markers derived from EST-sequence often have known or putative functions and are gene targeted markers with the potential of representing functional markers in those cases where polymorphisms in the repeat motifs affect the function of the gene in which they reside (Anderson, 2003). EST-SSR markers can contribute to ‘direct allele selection’, if they are shown to be completely associated or even responsible for a targeted trait (Sorrells & Wilson,

1997). Putative functions for a significant proportion of EST-SSR markers have been reported (Gao et al., 2004; Han, 2004).

A total of thirteen EST-VNTR loci were characterized using a panel of fourteen genotypes (Table 1) from *M. acuminata* Colla (A genome) and *M. balbisiana* Colla (B genome). Nine EST-VNTR loci (69%) generated amplification products, and detected polymorphisms between these two genotypes (Table 2).

Two primers (15%) Musa_Cachaco_O324TF and MUC4PE1021_A03_.b_017 detected more than three amplification products in the same genotype, and two primers (15%) Musa_Cachaco_N189TFB and MACVLIMFES026C14C8 did not generate polymorphic bands. Other studies in *Musa* showed that 21% (Ning et al., 2007) and 95% (Creste et al., 2006) of the SSR loci selected from genomic library were polymorphic. In addition, several studies in crop plants have shown that EST-SSR loci were less polymorphic compared to genomic SSRs because of greater DNA sequence conservation in transcribed regions (Scott, 2000; Gao, 2003).

EST-VNTR Locus	Repeated Sequence	Sequence (5` to 3`) F/R	N°. of alleles	PIC
1 - Musa_Cachaco_DH15TF	$(GTGTT)_4$	CCTTCGGTAGTTTGATCAGC AACAATTAGATAAGGAACTTG	3	0.69
2 - Musa_Cachaco_EB55TF	$(GAACC)_4$	CAACAGGGCAAGCTGGAGGTACAG ACTGTGAAACAATGTGAGATGGGTC	2	0.7
3 - Musa_Cachaco_ET53TF	$(GCAGT)_5$	GGAGGAGAGGATACATGGAA CTTGAATGATCAGCAAGCCAG	2	0.43
4 - Musa_Cachaco_FN22TF	$(ATGCC)_7$	CAACAGGGCAAGCTGGAGGTACAG ACTGTGAAACAATGTGAGATGGGTC	6	0.93
5 - Musa_Cachaco_0421TF	$(CGAGG)_4$	TGGAAGAGCGAGGGAGGAAGA GCCCTGGAATCGGAGTCGTGGTC	4	0.79
6 - MUC4LH1008_F03_.b_027	$(CACTG)_4$	TGTCTATGGGAATTGAGGATGAG GTACCGATCTCAAAATTGTCAGG	5	0.89
7 - MACVLIMFLS013C_B12_.b	$(TTTTTA)_4$	TTGGCAGGATTAAGCACAGTTTCC GAAGACGGATAATGCCTTTTGC	3	0.88
8 - MACVLIMFLS013D_E05_b	$(GAAAA)_4$	CCTTATTCTGTCGGGAAATCCA GTTCACAAGTGTCCCATGCAACAA	3	0.83
9 - Root1	$(CCGACG)_4$	ACACAAATATCCTTCCCGTCC TTCTTGGTAACGGTGAGATCG	4	0.75
10 - Musa_Cachaco_N189TFB	$(AGAGG)_4$	TAGCAGTGGAAGGGGCATGCA TTATGCTCACAACACACACAC	0	0
11 - Musa_Cachaco_O324TF	$(AGGGTT)_4$	TCACCGCCATGGACGTCGTCTA CATCAGTGACATTGCTCATCTC	>3*	0
12 - MACVLIMFES026C14C8	$(AGGGAG)_4$	AACCCCCCACCTTAAAAACATC ATCCCTCCAGCGTCAATCACCG	0	0
13 - MUC4PE1021_A03_.b_017	$(GTTTTT)_5$	CCTTATTCTGTCGGGAAATCCA GTTCACAAGTGTCCCATGCAACAA	>3*	0
Total of alleles			**32**	

TABLE 2 EST-VNTR Locus, primer sequence forward and reverse, number of alleles and Polymorphic Information Content (PIC). * More than two bands in diploid genotypes.

The average number of alleles amplified per VNTR primer was 3.55, ranging from 2 to 6, with a total of 32 alleles identified. Studies that using SSR from genomic library obtained higher values, 4.4 (Creste et al., 2006), 7.5 (Amorim et al., 2009) and 9.2 (Ning et al., 2007).

Polymorphism Information Content (PIC), a measure of allelic diversity at a locus, was estimated for each polymorphic EST-VNTR loci detected (Table 2). In comparison to genomic SSRs, SSR-ESTs revealed low polymorphic information content value in germplasm characterization and genetic diversity studies (Scott, 2000; Eujayl, 2001; Thiel et al., 2003). However in the present work the PIC values, ranged from 0.43 (Musa_Cachaco_ET53TF) to 0.93 (Musa_Cachaco_FN22TF), and the mean PIC value estimated across all the polymorphic VNTR loci was 0.76. This result is similar to other studies that used genomic SSRs (Creste et al., 2006; Amorim et al., 2009).

The number of alleles per genotype ranged from 7 to 13 for BB cultivars, and of 10 to 14 for AA cultivars. The *Musa acuminata* spp. *microcarpa* Borneo cultivar presented the highest number of alleles, 14; whereas the *Musa balbisiana* Singapure cultivar presented the smallest number of alleles, 7. The *Musa*_Cachaco_FN22TF locus presented the highest number of alleles, 6, whereas the loci *Musa*_Cachaco_EB55TF and *Musa*_Cachaco_ET53TF, presented the lowest number.

The dendrogram (FIGURE 1) indicates two main clusters, which correspond to the genome designation of the (AA) and (BB) genomes, showing that VNTR markers were able to separate the A from B genome. The bootstrap values were low due to the small number of analyzed bands, 32, but were sufficient to provide valuable information regarding the constitution of these genomes.

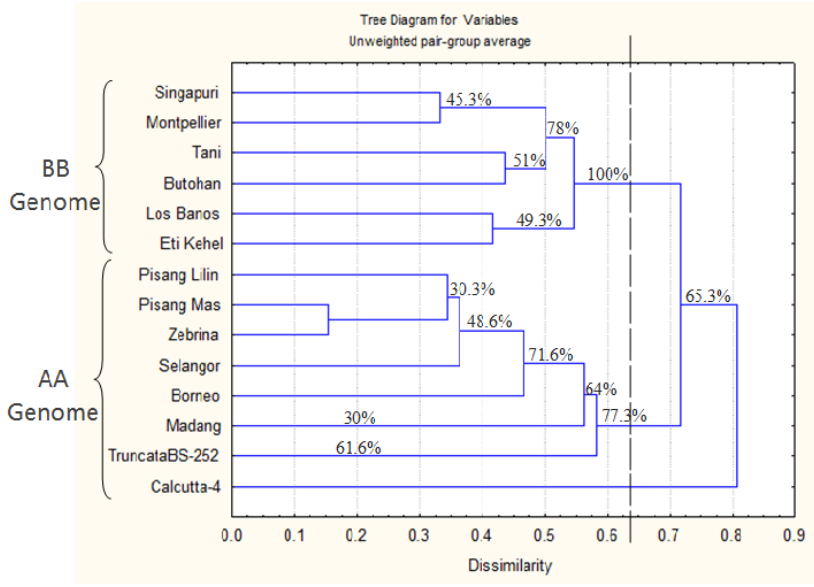


FIGURE 1 Dendrogram showing dissimilarity between *Musa* genotypes based on VNTR data.

In conclusion, the Data*Musa* database provides a valuable source of ESTs data for identification of polymorphism to be used in genetic diversity and gene discovery studies in banana.

2 - TRAP marker analysis

The same panel of *Musa* diploid genotypes (Table 1) was used in the validation of TRAP markers. From 12 TRAP PCR reactions that employed three fixed primers in combination with four random primers, all the primer combinations (100%) generated polymorphic bands and a total of 119 alleles (Table 3). This result means that the EST MUC4FL1018_H03 sequence alone provided 119 different markers identified in fourteen *Musa* genotypes.

The average number of alleles amplified per primer combination was 9.91, ranging from 5 to 16. The best primer combination was *Musa* SBP.1 x T03, providing the highest PIC value, 0.83. The number of alleles per genotype ranged from 42 to 74 for BB cultivars, and from 33 to 53 for AA cultivars. The *Musa balbisiana* "Butohan" cultivar presented the highest number of alleles, 74; whereas the cultivar *Musa acuminata* ssp. *burmannicoides* "Calcutta 4" presented the lowest number of alleles, 33.

The PIC value also was analyzed, ranging from 0.66 to 0.83 and the mean PIC value estimated across all the polymorphic TRAP loci was 0.77.

Primer Combination	N°. of bands	N° of polymorphic bands	% of Polymorphism	PIC
Musa SBP.1 x T03	128	9	7	0.83
Musa SBP.2 x T03	157	14	8.91	0.82
Musa SBP.3 x T03	144	13	9	0.78
Musa SBP.1 x T04	150	13	8.66	0.76
Musa SBP.2 x T04	128	16	12.5	0.79
Musa SBP.3 x T04	147	5	3.4	0.82
Musa SBP.1 x T13	142	5	3.52	0.66
Musa SBP.2 x T13	188	9	4.78	0.76
Musa SBP.3 x T13	173	6	3.46	0.73
Musa SBP.1 x T14	100	10	10	0.83
Musa SBP.2 x T14	172	12	7	0.71
Musa SBP.3 x T14	163	7	4.29	0.82
Total	1792	119	6.64	9.31

TABLE 3 Characteristics of the amplification products obtained with 12 TRAP primer pairs used to analyze the genetic diversity of *Musa* accessions.

Approximately 3.5%-12.5% of the amplified fragments, generated by the TRAP technique, were polymorphic markers. This percentage is in regard to the intensity of the amplified fragment that could be classified in strong, intermediate or weak. Only fragments with strong and intermediate band intensities were analyzed in order to provide more reliable results.

The IR-700 and IR-800 images generated by the Li-Cor system show a fraction of the amplified polymorphic bands generated by TRAP technique (FIGURE 2).

The excellent quality of the pictures provided by the Li-Cor DNA Sequencer makes the scoring of the amplified fragments faster, easier, and more reliable than VNTR. Moreover, the TRAP fragments can be detected using other amplified fragment detection system, such as the SRAP technique (Quiros, 2001).

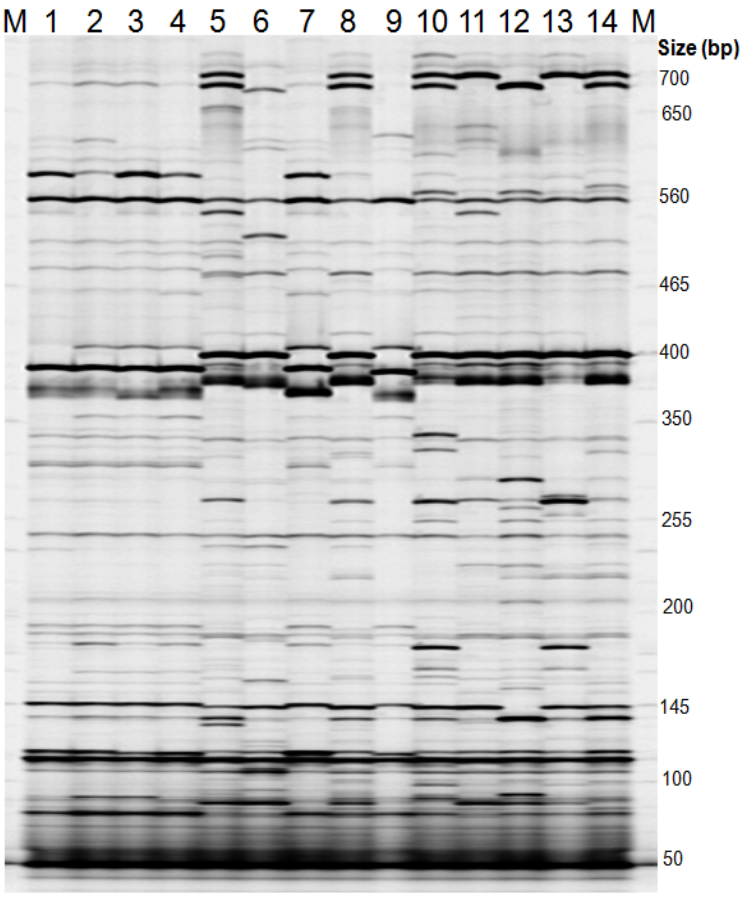


FIGURE 2 A portion of Target Region Amplification Polymorphism (TRAP) profile of DNA from 14 banana cultivars using the fixed primer *Musa* SBP.1 and the arbitrary primer T04. Lanes are: 1 Singapuri, 2 Los Banos, 3 Tani, 4 Butohan, 5 Pisang Lillin, 6 Calcutta 4, 7 Etikehel, 8Truncata BS-252-A,9 Montpellier,10 Borneo,11 Madang,12 Pisang Mas, 13 Selangor, 14 Zebrina. *M* = 1 kb molecular –weight marker.

Dendrogram in figure 3, indicated two main cluster groups at 99.8% and 100% of node consistency, which corresponded to the genome designation of plants (AA and BB).TRAP markers, as VNTR markers were also able to separate the A from B genome. The bootstrap values were high due to the large number of analyzed bands.

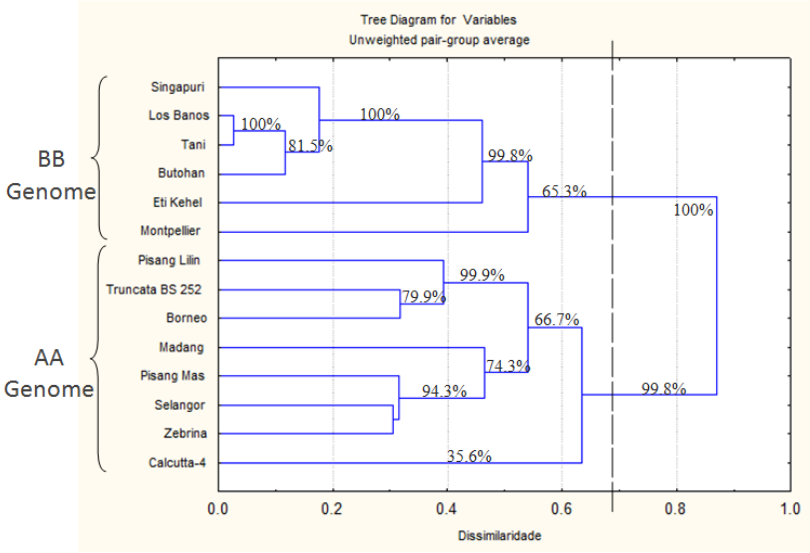


FIGURE 3 Dendrogram showing the dissimilarity between *Musa* genotypes based on TRAP data.

Several studies have been used the application of the TRAP technique with different purposes. One of them, using the TRAP technique for detecting quantitative trait loci in wheat, obtained a total of 357 markers and 24 markers per reaction using 15 fixed primers and 4 arbitrary primers. Probably this great difference in comparison to the results obtained by the study presented is due to the number of the fixed primers and to the number of genotypes used in the

characterization. In addition, this study revealed that the TRAP markers were highly efficient for genetic mapping in wheat (Liu et al., 2005). In another study using TRAP markers were used in lettuce genotyping via a panel of 53 cultivars, using 10 fixed primers and 4 arbitrary primers. Three-hundred and eighty-eight TRAP markers were encountered and the authors concluded that the TRAP makers are a powerful technique for fingerprinting lettuce cultivars (Hu et al., 2005).

An additional study used the TRAP technique for developing chromosome-specific TRAP-marker in wheat. Through 10 fixed primers and 2 arbitrary primers 307 TRAP markers were generated and the authors considered that the chromosome-specific markers developed provided an identity for each chromosome, facilitating the molecular and genetic characterization of the individual chromosome, including genetic mapping and gene identification (Li et al., 2007). Another study used the TRAP technique for mapping and tagging disease resistance traits in Common Bean. Seventy lines were used, 7 fixed primers and 11 arbitrary primers and 85 TRAP markers were generated. This study revealed that the TRAP-markers had great potential for mapping regions of common bean, linked to disease resistance and as observed in earlier studies, the TRAP technique detected numerous polymorphic markers that were reproducible and heritable as either dominant or codominant markers (Miklas et al., 2006).

The TRAP results presented in this work is in agreement with the results cited above and therefore demonstrates that is possible to use the TRAP technique with different aims in banana research. Additionally, TRAP technique proved to be a quick, reliable and efficient way to examine the variability genetic between *Musa* genotypes.

3- VNTR marker analysis X TRAP marker analysis

The dendrogram in figure 4 represents the data from VNTR/TRAP markers analyzed together. Two main cluster groups were formed at 99% and 100% of node consistency. The dendrogram also shows a clear separation between genotypes from the A and B genomes.

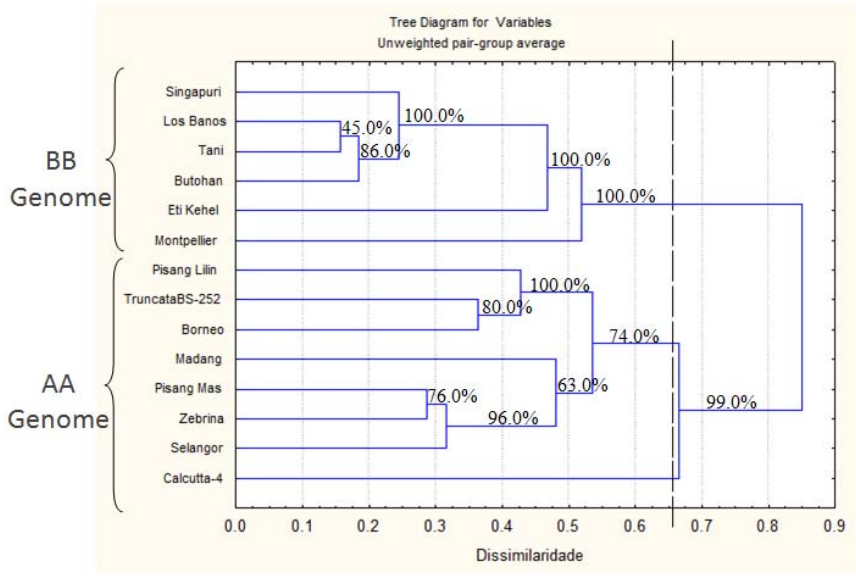


FIGURE 4 Dendrogram showing dissimilarity between *Musa* genotypes based on VNTR/TRAP data.

The cophenetic correlation value (Mantel, 1967), calculated for the VNTR and TRAP markers is r = 0.57, at 5% probability. This value indicates more reliability to our results, showing that both matrices were consistent in being able to differentiate both genomes (AA and BB).

The dendrograms generated for the VNTR primers and TRAP primers (Figure 1 and 3 respectively) show a clear distinction between the A and B genomes due to the presence of two main clusters (A and B). However the

subgroups varied for both. Probably these differences could be due to the template sequences used for primer designing. VNTR primers design used EST-sequences generated from the whole transcriptome from *M. acuminata* spp. *burmannicoides* “Calcutta 4”, while that the TRAP fixed primers were designed against only one gene sequence.

Nevertheless, the main focus of this report was not to discuss the similarity between the cultivars but, to show that the both techniques (VNTR and TRAP) proved to be very useful for molecular markers generation in *Musa* genotypes.

References

ALWALA, S.; SUMAN, A.; ARRO, J.A.; VEREMIS, J.C.; KIMBENG, C.A. Target region amplification polymorhysm (TRAP) for assessing genetic diversity in sugarcane germplasm collection. **Crop Science**, Madison, v.46, n.1, p.448-455, Jan. 2006.

AMORIM, E.P.L.; LEDO, C.A.D.; AMORIM, V.B.D.; REIS, R.V. dos; SANTOS-SEREJO, J.A.; SILVA, S.D.E. Agronomic and molecular characterization of diploid improved banana genotypes. **Revista Brasileira de Fruticultura**, Cruz das Almas, v.31, n.1, p.154-161, mar. 2009.

ANDERSON, J.R.L. Functional markers in plants. **Trends in Plant Science**, Oxford, v.8, n.12, p.554-560, Dec. 2003.

BHAT, K.V.; BHAT, S.R.; CHANDEL, K.P.S. Survey of isozyme polymorphisms for clonal identification in *Musa*. **HortScience**, Alexandria, v.67, n.4, p.501-507, 1992.

CORDEIRO, G.M.; CASU, R.; MCINTYRE, C.L.; MANNERS, J.M.; HENRY, R.J. Microsatellite markers from sugarcane *(Saccharum spp.)* ESTs cross transferable to erianthus and sorghum. **Plant Science**, New York, v.160, n.6, p.1115-1123, May 2001.

CRESTE, S.; BENATTI, R.T.; ORSI, M.R.; RISTERUCCI, A.M.; FIGUEIRA, A. Isolation and characterization of microsatellite loci from a commercial cultivar of Musa acuminata. **Molecular Ecology Notes**, Oxford, v.6, n.2, p.303-306, June 2006.

CROUCH, H.K.; JARRET, R.L.; TENKUANO, A.; ORTIZ, R. VNTR-based diversity analysis of 2x and 4x full-sib Musa hybrids. **Electronic Journal of Biotechnology**, Valparaíso, v.2, n.3, p.130-139, Dec. 1999.

CROUCH, J.H.; CROUCH, H.K.; CONSTANDT, H.; GYSEL, A. van; BREYNE, P.; MONTAGU, M. van; JARRET, R.L.; ORTIZ, R. Comparison of PCR-based molecular marker analyses of Musa breeding populations. **Molecular Breeding**, Dordrecht, v.5, n.3, p.233-244, Oct. 1999.

CRUZ, C.D. **Programa Genes**: versão Windows. Viçosa, MG: UFV, 2001. Software.

DECROOCQ, V.; FAVE, M.G.; HAGEN, L.; BORDENAVE, L.; DECROOCQ, S. Development and transferability of apricot and grape EST microsatellite markers across taxa. **Theoretical and Applied Genetics**, Berlin, v.106, n.5, p.411-422, Mar. 2003.

DELLAPORTA, S.L.; WOOD, J.; HICKS, J.B. A plant DNA minipreparation: version II. **Plant Molecular Biology Report**, Athens, v.1, n.1, p.19-21, Mar. 1983.

DSOUZA, M.; LARSEN, N.; OVERBEEK, R. Searching for patterns in genomic data. **Trends in Genetics**, London, v.13, n.12, p.497-498, Dec. 1997.

EUJAYL, I. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. **Euphytica**, Wageningen, v.119, n.1/2, p.39-43, May 2001.

GAO, L.F. Analysis of microsatellites in major crops assessed by computational and experimental approaches. **Molecular Breeding**, Dordrecht, v.12, n.3, p.245-261, Oct. 2003.

GAO, L.F.; JING, R.L.; HUO, N.X.; LI, Y.; LI, X.P.; ZHOU, R.H.; CHANG, X.P.; TANG, J.F.; MA, Z.Y.; JIA, J.Z. One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. **Theoretical and Applied Genetics**, Berlin, v.108, n.7, p.1392-1400, May 2004.

GUIMARÃES, N.C.C.; TORGA, P.P.; RESENDE, E.C.; PAIVA, E.; PAIVA, L.V. Identificação de variantes somaclonais em bananeiras 'Prata Anã', utilizando técnicas moleculares e citogenéticas. **Ciência e Agrotecnologia**, Lavras, v.33, n.2, p.448-454, mar./abr. 2009.

GUPTA, P.K.; SHARMA, S.; SINGH, R.; KUMAR, N.; BALYAN, H.S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. **Molecular Genetics and Genomics**, Berlin, v.270, n.4, p.315-323, Dec. 2003.

HAN, Z.G. Genetic mapping of EST-derived microsatellites from the diploid Gossypium arboreum in allotetraploid cotton. **Molecular Genetics Genomics**, Berlin, v.272, n.3, p.308-327, Oct. 2004.

HOLTON, T.A. Plant genotyping by analysis of microsatellites. In: HENRY, R.J. (Ed.). **Plant genotyping**: the DNA fingerprinting of plants. Wallingford: CABI, 2001. p.15-28.

HOLTON, T.A.; CHRISTOPHER, J.T.; MCCLURE, L.; HARKER, N.; HENRY, R.J. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. **Molecular Breeding**, Dordrecnt, v.9, n.2, p.63-71, June 2002.

HU, J.G.; OCHOA, O.E.; TRUCO, M.J.; VICK, B.A. Application of the TRAP technique to lettuce (Lactuca sativa L.) genotyping. **Euphytica**, Wageningen, v.144, n.3, p.225-235, Aug. 2005.

HU, J.G.; SEILER, G.S.; JAN, C.C.; VICK, B.A. **Assessing genetic variability among sixteen perennial helianthus species using PCR-based TRAP markers**. Fargo: Agricultural Research Service Northern Crop Science Laboratory, 2003. 5p.

HU, J.G.; VICK, B.A. Target region amplification polymorphism: a novel marker technique for plant genotyping. **Plant Molecular Biology Reporter**, Athens, v.21, n.3, p.289-294, Sept. 2003.

JARRET, R.L.; GAWEL, N.; WHITTEMORE, A.; SHARROCK, S. RFLP-based phylogeny of Musa species in Papua New Guinea. **Theoretical and Applied Genetics**, Berlin, v.84, n.5, p.579-584, June 1992.

KAEMMER, D.; FISCHER, D.; JARRET, R.L.; BAURENS, F.C.; GRAPIN, A.; DAMBIER, D.; NOYER, J.L.; LANAUD, C.; KAHL, G.; LAGODA, P.J.L. Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. **Euphytica**, Wageningen, v.96, n.1, p.49-63, July 1997.

KRESOVICH, S.; SZEWC-MCFADDEN, A.K.; BLICK, S.; MCFERSON, J.R. Abundance and characterization of simple sequence repeats (SSRs) isolated from a size fractionated genomic library of rassica napus L. (rapeseed). **Theoretical and Applied Genetics**, Berlin, v.106, n.2, p.411-422, July 1995.

LAI, Y.; SUN, F. The relationship between microssatelites slippage rate and the number of repeats units. **Molecular Biology and Evolution**, Chicago, v.20, n.12, p.2123-2131, Dec. 2003.

LEVINSON, G.; GUTMAN, G.A. Slipped misparing: a major machanism for DNA sequence evolution. **Molecular Biology and Evolution**, Chicago, v.4, n.3, p.203-221, May 1987.

LI, G.; KLINDWORTH, D.L.; SHIREEN, F.; CAI, X.; HU, J.; XU, S.S. Molecular characterization and chromosome-specific TRAP-marker development for Langdon durum D- genome disomic substitution lines. **Genome**, Ottawa, v.49, n.12, p.1545-1554, Dec. 2007.

LIU, Z.H.; ANDERSON, J.A.; HU, J.; FRIESEN, T.L.; RASMUSSEN, J.B.; FARIS, J.D. A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. **Theoretical and Applied Genetics**, Berlin, v.111, n.4, p.782-794, Aug. 2005.

MANTEL, N. The detection of disease clustering and generalized regression approach. **Cancer Research**, Birmingham, v.27, n.2, p.209-220, Feb. 1967.

MIKLAS, P.N.; HU, J.; GRUNWALD, N.J.; LARSEN, K.M. Potential application of TRAP (Targeted Region Amplified Polymorphism) markers for mapping and tagging disease resistance traits in common bean. **Crop Science**, Madison, v.46, n.1, p.910-916, Jan. 2006.

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**, Amsterdam, v.114, n.2, p.281-288, Oct. 2007.

ORTIZ, R.; FERRIS, R.S.B.; VUYLSTEKE, D.R. Banana and plantain improvement. In: GOWEN, S. (Ed.). **Bananas and plantains**. London: Chapman and Hall, 1995. chap.5, p.110-146.

QUIROS, G.L.C.F. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. **Theoretical and Applied Genetics**, Berlin, v.103, n.2/3, p.455-461, Aug. 2001.

REDDY, P.M.; SARAL, N.; SIDDIQ, E.A. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. **Euphytica**, Wageningen, v.128, n.1, p.9-17, Nov. 2002.

ROZEN, S.; SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers. In: KRAWETZ, S.A.; MISENER, S. **Bioinformatics methods and protocols**: methods in molecular biology. New Jersey: Humana, 2000. p.365-386.

SCOTT, K.D. Analysis of SSRs derived from grape ESTs. **Theoretical and Applied Genetics**, Berlin, v.100, n.5, p.723-726, Mar. 2000.

SIMMONS, M.P.; ZHANG, L.B.; WEBB, C.T.; MÜLLER, K. A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. **Molecular Phylogenetics and Evolution**, Orlando, v.42, n.2, p.528-542, Feb. 2007.

SIMMONDS, N.W. Bananas *Musa* (*Musaceae)*. In: _____. **Evolution of crop plants**. 2.ed. Essex: Longmans Scientific and Technical, 1995. p.370-375.

SIMMONDS, N.W.; SHEPARD, K. The taxonomy and origins of the cultivated bananas. **Journal of the Linnean Society of London (Botany)**, London, v.55, n.359, p.302-312, Dec. 1995.

SORRELLS, M.E.; WILSON, W.A. Direct classification and selection of superior alleles for crop improvement. **Crop Science**, Madison, v.37, n.33, p.691-697, 1997.

SQUIRRELL, J.; HOLLINGSWORTH, P.M.; WOODHEAD, M.; RUSSELL, J.; LOWE, A.J.; GIBBY, M.; POWELL, W. How much effort is required to isolate nuclear microsatellites from plants? **Molecular Ecology**, Oxford, v.12, n.6, p.1339-1348, June 2003.

STATISTICA FOR WINDOWS. **Computer program manual**. Version 6. Tulsa: Statsoft, 2002. 1 CD-ROM.

STOVER, R.H.; SIMMONDS, N.W. **Bananas**. 3.ed. London: Longman, 1987. 468p.

THIEL, T.; MICHALEK, W.; VARSHNEY, R.K.; GRANER, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, Berlin, v.106, n.3, p.411-422, Feb. 2003.

VARSHNEY, R.K.; THIEL, T.; STEIN, N.; LANGRIDGE, P.; GRANER, A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. **Cellular and Molecular Biology Letters**, Berlin, v.7, n.2A, p.537-546, Sept. 2002.

VENKATACHALAM, R.V.S.; BHAGYALAKSHMI, L. The use of genetic markers for detecting DNA polymorphism, genotype identification and phylogenetic relationships among banana cultivars. **Molecular Phylogenetics and Evolution**, Orlando, v.47, n.3, p.974-985, June 2008.

WONG, C.; KIEW, R.; LOH, J.P.; GAN, L.H.; SET, O.; LEE, S.K.; LUM, S.; GAN, Y.Y. Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. **Annals of Botany**, London, v.88, n.6, p.1017-1025, Dec. 2001.

CHAPTER 3

ALLELIC DIVERSITY OF *Mycosphaerella fijiensis* POPULATIONS IN COSTA RICA

ABSTRACT

Mycosphaerella fijiensis is the causal agent of black leaf streak of banana, the most important threat to banana production in many countries and particularly in Costa Rica due the favorable climate conditions for this disease development. Currently, the main control strategy is the frequent fungicide applications. However, apart from environmental concerns, this approach is not sustainable due to the abrupt or gradual development of fungicide resistance. To analyze the population dynamics of fungicide resistance, we developed a molecular diagnostics for strobilurin resistance, using the cytochrome b gene (*cytb*), in *M. fijiensis*. We also developed molecular markers for the mating type idiomorphs (*mat1-1* and *mat1-2*) and primers for five VNTR loci to estimate population genetic parameters. Monospore isolates obtained from three plantations that were 20-30 km apart (Cartagena, San Pablo and Zent) in the Limón province that represents the heart of the Costa Rican banana production area. Ninety-five isolates were obtained from a distant wild-type population that was never sprayed with fungicides, located in San Carlos region in the Alajuela province. A total of 665 isolates were assayed for *mat1-1*, *mat1-2*, VNTR and *cytb*. The mating type genes segregated in a 1:1 ratio indicating that the sampled populations most likely were randomly mating. The VNTR primers identified 32 alleles, among them, 21 were identified in wild type population, 20 in San Pablo population, 23 in Zent population and 24 in Cartagena population. The strobilurin diagnostic indicated that the wild-type population was entirely sensitive and that two of the three commercial populations, San Pablo and Cartagena were entirely resistant. The Zent population contained 8% of sensitive strains even though strobilurins were still used in that plantation. The next steps will be analysis of gene diversity within each population and the levels of gene flow. These analysis provide excellent basis for future research on fungicide resistance in *M. fijiensis* Costa Rica populations as well as comparative analysis with other banana-producing areas.

RESUMO

Mycosphaerella fijiensis é o agente causador da Sigatoka Negra em bananeira, sendo a principal ameaça para a produção de banana em muitos países, principalmente na Costa Rica, devido às condições climáticas favoráveis para o desenvolvimento da doença. Atualmente, a principal estratégia de controle é a freqüente aplicação de fungicidas. No entanto, além de preocupações ambientais, essa estratégia não é sustentável devido ao desenvolvimento abrupto ou gradual de resistência aos fungicidas. Para analisar a dinâmica populacional de resistência a fungicidas à base de estrobirulinas, foi desenvolvido um diagnóstico molecular para detectar isolados sensíveis e resistentes utilizando o gene citocromo b (*cytb*) de *M. fijiensis.* Foram também desenvolvidos marcadores moleculares para o gene relacionado à reprodução de *M. fijiensis* (Mating Types genes) do tipo mat1-1 e mat1-2 e também iniciadores para cinco locos de marcadores VNTRs para estimar a diversidade alélica de quatro populações de *M. fijiensis* da Costa Rica. Isolados monospóricos foram obtidos de três fazendas, a citar, Cartagena, San Pablo e Zent separadas por uma distância de 20 a 30 km, localizadas na provínica de Limón, que representa a principal área de produção de banana da Costa Rica. Noventa e cinco isolados foram obtidos de uma população selvagem que nunca foi pulverizada com fungicidas, localizada na região San Carlos, na província de Alajuela. Um total de 665 amostras foram analisadas para mat1-1, mat1-2, VNTR e *cytb*. Os alelos mat1.1 e mat1.2 segregaram em uma proporção de 1:1, indicando que as populações foram amostradas aleatoriamente para o este tipo de reprodução. Os primers VNTR identificaram 32 alelos, dentre eles, 21 foram identificados na população selvagem (WT), 20 na população de San Pablo, 23 na população Zent e 24 na população de Cartagena. O diagnóstico para estrobilurina indicou que a população selvagem (WT) foi totalmente sensível e que duas das três populações comerciais, San Pablo e Cartagena, foram totalmente resistentes.

Introduction

Mycosphaerella fijiensis (anamorphm *Paracercospora fijiensis*) is an ascomycete, haploid and heterothallic fungus (Mourichon & Zapater, 1990). It causes black leaf streak disease (BLSD), commonly known as black Sigatoka, the most economically important disease of banana (*Musa* spp.) in tropical and subtropical areas (Marin et al., 2003).

M. fijiensis spreads through either the transit of infected plant material (infected suckers and diseased leaves used to wrap food or other goods) or the ascospore (produced during sexual reproduction) or conidia dispersal (produced during asexual reproduction) (Gauhl et al., 2000). Whereas conidia are mainly dispersed over short distances within a same plant and to nearby plants, viable ascospores might be carried a few hundred kilometers by wind (Parnell & Wilson, 1998). Leaves are infected indirectly via stomata, and the abaxial surface is the primary affected site due to the great abundance of stomata on this leaf part (Washington et al., 1998). It causes large necrotic lesions on the banana tree leaves, decreases the photosynthetic capacity, resulting in fruit weight and quality loss. Field losses range from 30-50% depending on climatic conditions (Gauhl et al., 2000) and therefore impacting the food security and economy in the affected regions.

Black Sigatoka is controlled primarily by fungicides, due to the fact that nonchemical alternatives do not provide commercially acceptable control (Romero, 1997). The cost of controlling the disease in large plantations is about US$ 1,000 per hectare (Arias et al., 2003), but it is higher in smaller plantations where fungicide aerial spraying is not performed and in which crop losses can be as high as 50% (Mobambo et al., 1993).

The intensive use of these agents has imposed a selective pressure, favoring the development of resistant or tolerant strains, leading to an increase in application cycles. This was exemplified by the rapid development and spread of

strobilurin resistance in Central America[2]. From 1997 onward, the strobirulin fungicide azoxystrobin was widely used in the main banana production zone in Costa Rica against *M. fijiensis* var. *difformis* (Amil et al., 2007). Strobirulins are an important class of chemicals for the management of a broad range of fungal disease in the agricultural system. They have a single-site mode of action, inhibiting mitochondrial respiration by binding to the Qo site of the cytochrome bc1 enzyme complex, thus blocking electron transfer in the respiration pathway, thus interfering with ATP production (Barlett et al., 2002).

Sequencing of the *cytochrome b* gene of sensitive and resistant *M. fijiensis* isolates from Costa Rica revealed a difference in the nucleotide sequence leading to a single amino acid change from glycine to alanine at position 143 in the resistant isolate. This change is known to occur also in the naturally tolerant basidiomycete *Mycena galopoda*. It is suggested that the field isolates of *M. fijiensis* can acquire resistance to Qo inhibitors due to a target site alteration with a single base pair change (Sierotzki et al., 2000).

The sexual cycle of the *M. fijiensis* fungus is controlled by a single mating type locus (MAT) and plays an important role in BLSD epidemiology (Gauhl et al., 2000). This mating type locus is structurally unusual, because it contains one of two forms of dissimilar sequences occupying the same chromosomal position. The two non-allelic versions of the mating type locus in fungal species with heterothallic mating strategy were labeled ‘idiomorph’ (Metzenberg & Glass, 1990). By convention, the mating type idiomorphs of complementary isolates are termed MAT1-1 and MAT1-2 (Turgeon & Yoder, 2000). The number of genes in each idiomorph varies among different groups of fungi. However, they all contain homeodomain-encoding genes, either alpha box (for mat1-1) or high mobility group (HMG) domain transcription factors (for mat1-2) (Turgeon & Yoder, 2000).

Idiomorphs mat1-1 and mat1-2 from *M. fijiensis*, were isolated presuming substantial synteny between *M. graminicola* and *M. fijiensis* as a basis to isolate the mating genes of *M. fijiensis*[15]. The knowledge of the mat gene sequences in *M. fijiensis* is a starting point to a better understanding of the relevance of reproduction and recombination, in relation to the epidemiology of these important pathogens and the interaction with other species (Conde-Ferraez et al., 2007).

The population genetic structure of *M. fijiensis* has been studied on a global, country and field scale using molecular markers. These studies have shown that populations of *M. fijiensis* can maintain a high level of genetic diversity at a small scale, and that recombination plays an important role in this pathogen (Carlier et al., 1999).

Restriction Fragment Length polymorphism (RFLP) and PCR-RFLP markers were developed for the *M. fijiensis* genome, and used to characterize populations of *M. fijiensis* at a regional and global scale (Carlier et al., 1994, 1996; Rivas et al., 2004). Moreover, simple sequence repeat (SSR) markers also have been established for *M. fijiensis* (Carlier et al., 1999; Neu et al., 1999; Perea et al., 2005).

The genome of *M. fijiensis*, isolate CIRAD-086, was sequenced by DOE (Department of Energy) – JGI (Joint Genome Institute), California-USA. The release of the first draft of this genome, in August 2006, comprised a 7,1X of this 73.4 Mb genome. This genome sequencing enabled the identification of hundreds of potential Variable Number Tandem Repeats (VNTR) markers to be used in *M. fijiensis* research.

The objective of this study was to characterize four *M. fijiensis* populations from Costa Rica regarding the *cytochrome b* gene, the ratio between mat1-1 and mat1-2 genes, and 5 VNTR previously identified and validates for *M. fijiensis* studies. To do so, a study was carried out using 665 *M. fijiensis*

isolates from the main local banana and plantain producing regions in Costa Rica.

Material and methods

a) Sampling sites

Four different farms in Costa Rica were sampled in the beginning of 2008 (Figure 1). San Pablo, Zent, and Cartagena are located in the Limón province, where bananas are grown in high density on large plantations and black leaf streak is controlled by chemical fungicides. (Table 1).

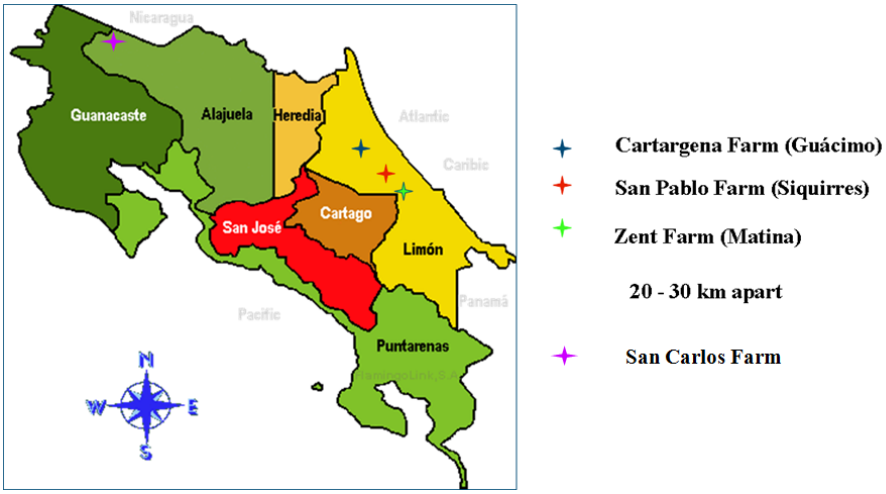


FIGURE 1 Costa Rican Map showing the locations of Cartagena, San Pablo, Zent and San Carlos farms, where *M. fijiensis* isolates used in this study were collected. Cartagena, San Pablo and Zent are 20 – 30 Km apart and are approximately 64 ha, 285 ha, 342 ha large, respectively. The wild type isolates were collected from a 0,5 ha-experimental plot at San Carlos regions.

These farms represent the main Costa Rica banana production area and are 20-30 km apart. They are are approximately 285 ha, 342 ha, 64 ha large, respectively. The fungicide spraying program adpted on the farms before samples were taken consisted of rotation of protectant and systemic fungicides. Generally 2-4 protectant (mainly mancozeb) followed by 1-2 systemic applications s (tridemorph, triazols, pyrimethanil or pyraclostrobin) (Table 1).

Fungicide	Dose	Number of sprayed/ year (2007)		
	g i.a /ha	San Pablo	Cartagena	Zent
Mancozeb	1.2	24	28	31
Chlrothalonil	720	2	2	0
Tridemorph	450	16	19	10
Triazols	100	10	9	5
Pyrimethanil	300	4	5	4
Pyraclostrobin	100	2	1	2

TABLE1 Fungicides, rate and the number of sprays used in the San Pablo, Cartagena and Zent banana plantation farms, in 2007, before samples were collected for this study in 2008.

The fourth farm is San Carlos, located at Alajuela province, isolated from the principal banana production area and was never sprayed with fungicides. The wild type isolates were collected from a 0.5 ha experimental plot within that area.

b) Isolate selection

One hundred ninety monoascosporic isolates of *M. fijiensis* were collected from each farm (San Pablo, Cartagena and Zent), from banana leaf tissue bearing abundant pseudothecia and ninety five collected from the San Carlos farm. Leaf tissue was collected from ten banana plants from each farm.

Ascospores were discharged from the pseudothecia onto water agar (Stover, 1962), and single ascospores were transferred immediately to 15 x 100-mm Petri dishes filled with potato dextrose agar (PDA). Eight to ten ascospores from each sampled point were placed on each plate containing PDA. After 4 days the colonies were visible to the naked eye, transferred to Mycophil agar (Becton Dickinson Microbiology Systems, Cockeysville, MD), and incubated for 15 at 25^{o}C under continuous fluorescent light for colony growth and conidial production.

c) DNA isolation

Before DNA isolation, mycelium were harvested and submitted to freeze-drying for 24 hours. A total of 190 samples from each farm, San Pablo, Zent, Cartagena, and 95 samples from San Carlos, were submitted to DNA isolation. Genomic DNA was extracted with the Kit Wizard Magnetic DNA Purification System for food Kit (Promega) according to manufacturer`s instructions. After DNA isolation, 2 µL of the DNA samples were quantified using NanoDrop.

d) Mating type idiomorphs (mat1-1 and mat1-2) primers

For amplification of mating type idiomorph (mat1-1 and mat1-2) markers, the following primers were used: Mat1.1 Forward: 5'-CATGAGCACGCTGCAGCAAG-3', Mat 1.1 Reverse: 5'-GTAGCAGTGGTTGACCAGG-3', Mat 1.2 Forward: 5'-GGCGCTCCGGCAAATCTTC-3', Mat 1.2 Reverse: 5'-CTTCTCGGATGGCTTGCGTG-3'.

e) Mating type idiomorphs (mat1-1 and mat1-2) markers amplification

Each PCR reaction was performed in a 50 µL total volume, containing 50ng of template genomic DNA, 2mM MgCl2, 600 µM dNTPs, 5 µM each primer, 0.4 U of Taq-DNA polymerase (Roche, Mannhein, German). Temperature cycling was carried out following the following program: 94^{o}C for 2 min, 40 cycles of 94^{o}C for 1 min, 70^{o}C for 30 s and 72^{o}C for 1 min, and a final elongation period of 10 min. at 72^{o}C. PCR products were separated electrophoretically using 1.0% agarose gels containing 0.3 µg/mL ethidium bromide, in 0.5 x TBE buffer at 120 V for approximately 1 hour. Amplification products were visualized and photographed using UV transilluminator**.**

f) VNTR identification and primer design

The entire genome sequence of *M. fijiensis* was submitted to a bioinformatics pipeline consisting of a VNTR detection / primer development module with the following steps: a) chopping contigs into 1500 bp size fragments with a 100 bp overlap ; b) VNTR detection using a Perl script allowing imperfect repeat VNTR detection by using the scan-for-matches program (Dsouza et al., 1997); c) generation of a maximum of 5 primer sets by using Primer3 (Rozen & Skaletsky, 2000) that could amplify the VNTR by PCR followed by a screening for unique primer pairs, d) additional screening for using the Fast PCR Professional 5.2 software. Out of eight VNTR primers designed from *M. fijiensis*, five were tested in the four populations from Costa Rica (Table 2).

Primer Name	Sequence (5`to 3`)	Repeated Sequence of VNTR marker
VNTR_ 1333_F	GAGTGAAGTACTGCGGAGGC	CGCCTTT (12)
VNTR_ 1333_R	AGTTGGAGAAAGGCGAAAGG	
VNTR_ 3959_F	GCGCGAGGCTTTCTATCTC	TATCTT (11)
VNTR_ 3959_R	ACCCCGATTAGGGAAGGTC	
VNTR_ 3786_F	GTGGTGGGATGAGGTGTGG	AGTGGAAGA (8)
VNTR_ 3786_R	CGCCGCGCGCTTTGACTCTTC	
VNTR_ 0252_F	TAGAGGCTACCCTGCCGTC	CTAGCTAATA (6)
VNTR_ 0252_R	GTATACTTCCGACCTCGGGC	
VNTR_0705_F	ATAGGATGCGGCAGACACTC	ACCACTC (6)
VNTR_ 0705_R	CGTCGCGATTTGAAGTGCC	

TABLE 2 Five VNTR primers were designed from the *Mycosphaerella fijiensis* genome, isolate CIRAD 086, in order to amplify the repeated sequences of VNTR markers and evaluate the polymorphism within *M. fijiensis* populations from Costa Rica.

g) VNTR PCR amplification

Each PCR reaction was performed in a 25 µL total volume, containing 50ng of template genomic DNA, 2mM MgCl2, 600 µM dNTPs, 5 µM each primer, 0.4 U of Taq-DNA polymerase (Roche, Mannhein, German). Temperature cycling was conducted with the following program: 94^{o}C for 2 min., 13 cycles of 94^{o}C for 30 s, 66^{o}C for 30 s (-1^{o}C per cycle) and 72^{o}C for 30 s, followed by 28 cycles of 94^{o}C for 30 s, 53^{o}C for 30 s and 72^{o}C for 30 s and a final elongation period of 7 min. at 72^{o}C.

PCR products were separated electrophoretically using 3.0% agarose gels containing 0.3 µg/mL ethidium bromide, in 0.5 x TBE buffer at 120 V for approximately 5 h. Amplification products were visualized and photographed using UV transilluminator.

h) Development of PCR primers to amplify a fragment of the cyt b gene from M. fijiensis isolates sensitive and resistant

The sequence from the strobilurine (QoI)-resistant isolate containing G143A mutation (Accession AF343069) (Sierotzki et al., 2000) was compared to the strobilurine (QoI)-sensitive one (Accession AF343070) (Sierotzki et al., 2000) using the DNASTAR software. Both sequences are available at National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). The only difference between the two sequences was on one nucleotide, guanine (G). The sequence of the resistant isolate, instead of guanine, there is one cytosine (C). This mutation leads to a single amino acid change from glycine to alanine at position 143 (G143A). Figure 2 shows part of the *cyt b* gene sequence and the point mutation. Three primers, one forward and two reverse were designed in order to diagnose sensitive and resistant isolates: MFcytFor_1(F1) 5'-CTCAATACTGCCTCAGC-3', MFcytRev_1 (R1) 5'-CCGTAATGTGGTTCATC-3', and MFcytRev_S (Rs) 5'-GTTATAACTGTAGCTCC-3'. The primer design was made using the software Primer 3.

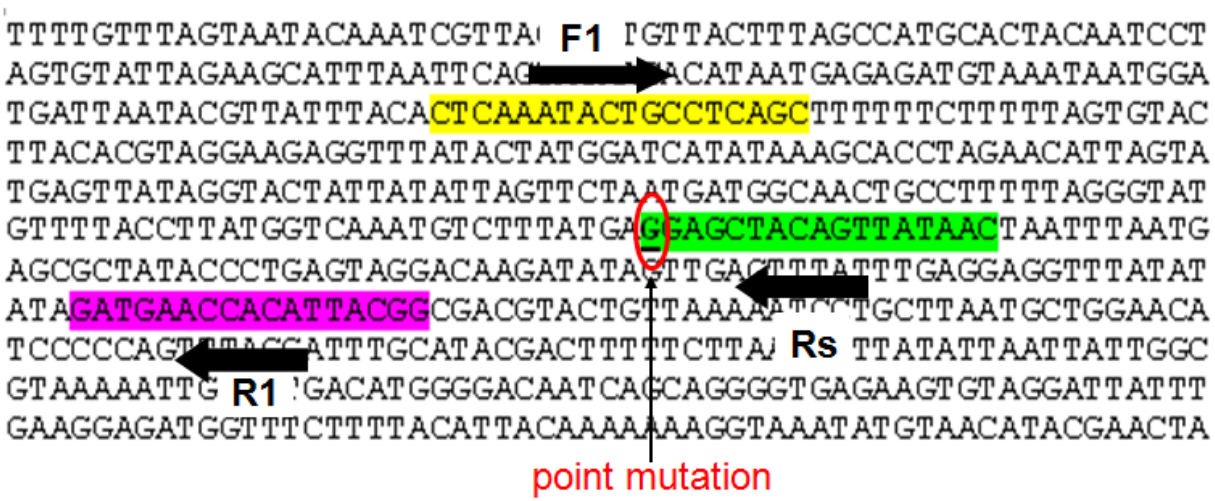


FIGURE 2 Part of *cyt b* gene sequence from a strobirulin sensitive *Mycosphaerella. fijiensis* isolate , with the located point mutation. The mutation leads to a change of a guanine (G) to a cytosine nucleotide (C). F1 means forward primer and the yellow highlighted is the position where the F1 primer annealed in the sequence. Rs means reverse primer and the green highlighted is the local where the Rs primer annealed only in the absence of mutation. R1 means reverse primer and the pink highlighted is the local where it annealed. F1 and R1 work as an internal control in a PCR reaction.

i) Multiplex PCR to detect Cytochrome b gene amplification

Each PCR reaction was performed in a 20 µL total volume, containing 50ng of template genomic DNA, 2mM MgCl2, 600 µM dNTPs, 5 µM each primer (MFcytFor_1, MFcytRev_1 and MFcytRev_S), 0.4 U of Taq-DNA polymerase (Roche, Mannhein, German). Temperature cycling was conducted with the following program: 94^{o}C for 2 min, 40 cycles of 94^{o}C for 1 min, 70^{o}C for 30 s and 72^{o}C for 1 min, and a final elongation period of 10 min. at 72^{o}C.

PCR products were separated electrophoretically using 1.0% agarose gels containing 0.3 µg/mL ethidium bromide, in 0.5 x TBE buffer at 120 V for approximately 1 hour. Amplification products were visualized and photographed using UV transilluminator and photographed using an Eagle Eye II still video system.

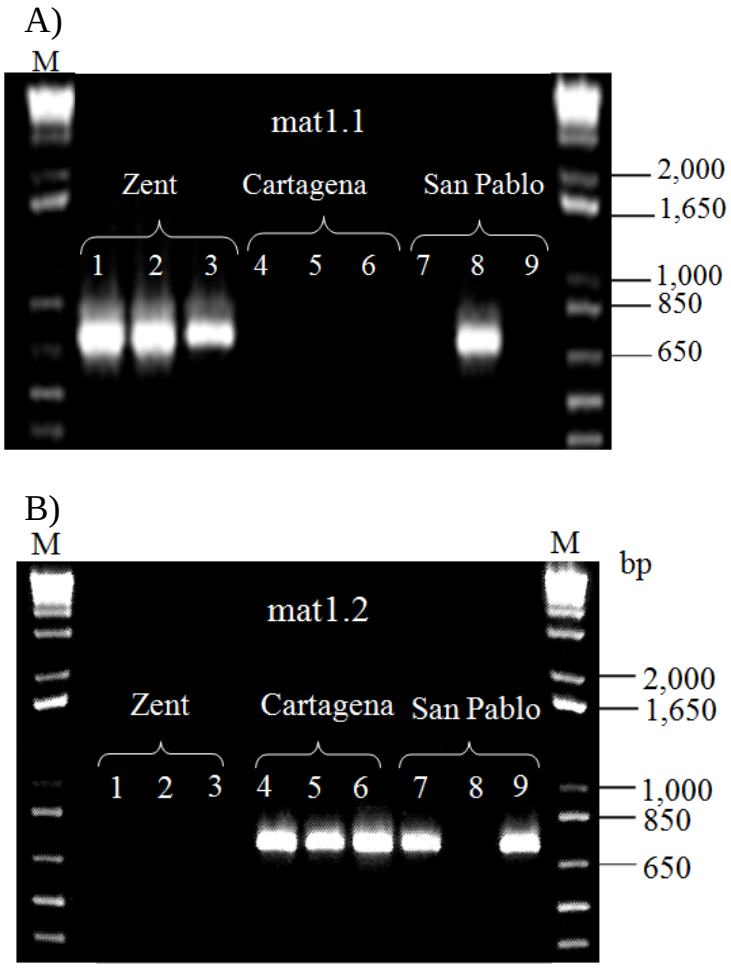
Results and discussion

1- Segregation of Mating Types (mat1.1 and mat1.2) in the Mycosphaerella fijiensis populations

A total of 665 *M. fijiensis*, 190 of which from San Pablo, 190 from Zent, 190 from Cartagena and 95 from San Carlos, were assayed for mat1-1 and mat1-2. The diagnostic band for ma1-1 and mat1-2 has the same size, 700 bp. The figure 3 shows the PCR result from the amplification of mating types (mat1.1 and mat1.2). The results show clearly that the isolates presented bands for mat1.1, did not presented for mat1.2. Therefore, primers used are highly specific to discriminate the idiomorphs mat1.1 and mat1.2.

The population from San Pablo presented 50.6% of mat1.1 and 49.4% from mat1.2, the population from Zent presented 44.5% of mat1.1 and 55.5% of mat1.2, the population from Cartagena presented 51.5% from mat.1.1 and 48.5% from mat1.2. The chi-squared test (Schuster & Cruz, 2008) on segregation was

calculated for each population. The value for wild type population was 0.025, for San Pablo population was 0.024, for Zent population was 2.09 and Cartagena population was 0.142. According to the percentile of chi-squared distribution, the *MAT* locus segregated in a 1:1 ratio indicating that the sampled populations were more likely randomly mating.



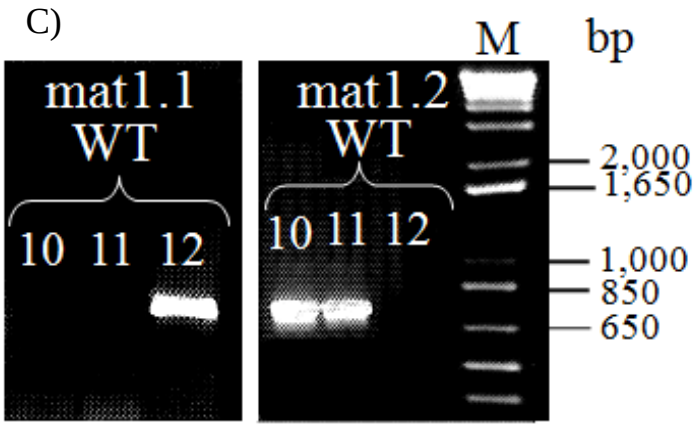


FIGURE 3 Partial results of the mating type marker amplification from PCR. 1, 2 and 3 were DNA samples from the Zent population, 4, 5 and 6 were DNA samples from the Cartagena population, 7,8 and 9 were DNA samples from the San Pablo population, 10, 11 and 12 were DNA sample from the wild type (WT) population. A) mat1.1 amplification. B) mat1.2 amplification C) mat1.1 and mat1.2 amplification. Agarose gel at 1% stained with 0.3 µg/mL of ethidium bromide. Amplification products were visualized and photographed using UV transilluminator. *M* = 1 kb Plus DNA Ladder molecular weight standard.

2- VNTR markers identified in the Mycosphaerella fijiensis populations

The alleles identified for each VNTR locus in each population, and their size (pb) is shown in the table 1. The VNTR locus 1333 and 0252 provided 2 different alleles, the VNTR locus the VNTR locus 0705 provided 3 alleles the VNTR locus 3786 provided 10 alleles, the VNTR locus3959 provided 15 alleles. Out of 32 different alleles identified, 15 were shared between isolates from all populations.

The number and size (pb) of identified alleles from the VNTR loci 3959, 3786 and 0705 were different among the populations. The VNTR locus 3959 provided 8 alleles in the Wild Type population, 7 alleles in the San Pablo

population, 9 alleles in the Zent population and 10 alleles in the Cartagena population. The VNTR locus 3786 provided 7 alleles in the Wild Type, San Pablo and Zent populations, and 8 alleles in the Cartagena population. The VNTR locus 0705 provided 2 alleles in the Wild Type, San Pablo and Cartagena populations, and 3 alleles in the Zen population. The VNTR loci 1333 and 0252 provided the same alleles in all the populations. The VNTR locus 3959 was the most polymorphic.

Figure 4 shows the amplification of some alleles from VNTR loci. The markers allowed robust scoring on agarose gel 3% and proved to be useful for allelic diversity studies.

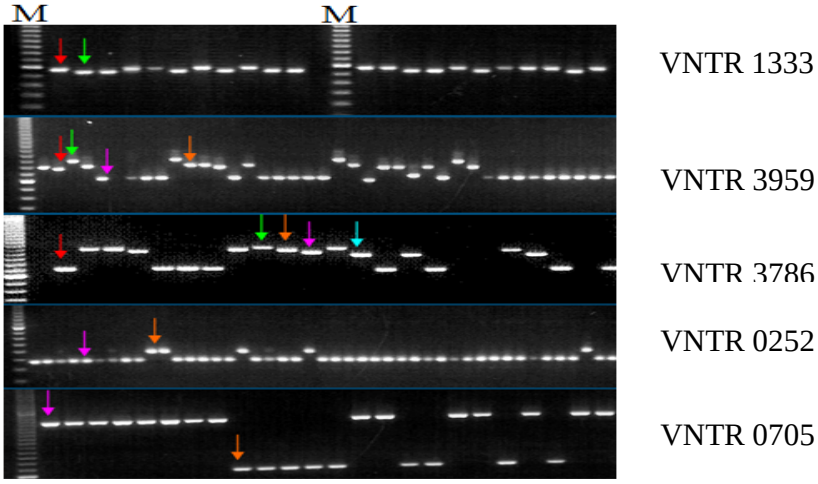


FIGURE 4 Partial results of allele amplification from the VNTR loci from the San Pablo population. Arrows indicate the different alleles amplified. The samples were applied in 3% agarose gel, stained with 0.3 µg/mL of ethidium bromide. Amplification products were visualized and photographed using UV transilluminator. *M* = EZ Load 20 pb Molecular Ladder (Invitrogen).

In the Wild Type population a total of 21 alleles were identified, in the San Pablo population 20 alleles, in the Zent population 23 alleles, in the Cartagena population 24 alleles. The Cartagena population showed the highest allelic diversity in relation to the other populations (Table 3).

VNTR locus	Wild type	San Pablo	Zent	Cartagena
1333	2	2	2	2
3959	8	7	9	10
3786	7	7	7	8
0252	2	2	2	2
0752	2	2	3	2
Total of alleles	21	20	23	24

TABLE 3 Total number of alleles identified from each VNTR locus in each *Mycosphaerella fijiensis* population. VNTR 1333 and VNTR 0252 presented the same number of alleles in all populations. VNTR 3959, VNTR 3786 and VNTR 0752 presented different number of alleles among the populations.

On average, the total sample presented 6.6 alleles per locus. The wild type population had 4.2 alleles, San Pablo population had 4, Zent population had 4.6, and Cartagena population had 4.8. This result indicated that the allelic variability between populations in regard to VNTR loci was very similar.

Several studies have used molecular markers to describe the genetic variability of *M. fijiensis*. Fourteen polymorphic microsatellites markers were identified in the sequenced genome of *M.fijiensis* using a Perl script (Yang & Zhong, 2007). The SSR primers were tested in 36 isolates from Hawaii. The average of alleles per locus was 6.43, very similar with our results.

The genetic structure of *M. fijiensis* populations around the world was examined using DNA restriction fragment length polymorphism (RFLP)

markers. Nineteen RFLP loci were tested in 136 *M. fijiensis* isolates from five geographical populations representative of banana-producing areas (South-East Asia including the Philippines and Papua New Guinea, Africa, Latin America and Pacific Islands). The mean of alleles per locus was 10.1 (Carlier et al., 1996). Microsatellite markers were also used to study the genetic diversity of Colombia isolates of *M. fijiensis* and were identified from microsatellite-enriched libraries. Seven SSR primers were tested in 36 isolates generating 2.03 alleles per locus (Perea et al., 2005). The same SSR primers were used to evaluated the genetic diversity of Mexico and Nigerian isolates (Neu et al., 1999), and the number of alleles per locus was 2.6 for the isolates from Mexico and 2.7 for the isolates from Nigeria. These values were lower than the one reported in this study (6.6). This fact could be related to the number of isolates analyzed or that the VNTR loci tested in our study were more informative.

To perform reliable comparisons, further analysis, as gene diversity (h), genotype diversity (G) and the level of genetic differentiation (Fst) between the *M. fijiensis* populations need to be made. This analysis will be presented soon.

Polymorphism Information Content (PIC), a measure of allelic diversity at a locus, was estimated for each of the polymorphic VNTR detected locus (Table 4). The PIC value from the VNTR locus 1333 ranged from 0.61 to 0.69, from the VNTR locus 3959 ranged from 0.95 to 0.97, from the VNTR locus 3786 ranged from 0.95 to 0.96, from the VNTR locus 0252 ranged from 0.51 to 0.71, and from the VNTR locus 0705 ranged from 0.6 to 0.74. The average PIC value of 0.76, demonstrates the high informativeness of these markers. The VNTR markers developed will be useful for M. fijiensis population genetics studies.

VNTR Locus	PIC Wild type	PIC San Pablo	PIC Zent	PIC Cartagena
VNTR1333	0.67	0.61	0.67	0.69
VNTR 3959	0.96	0.96	0.95	0.97
VNTR 3786	0.96	0.95	0.95	0.95
VNTR 0252	0.71	0.51	0.52	0.56
VNTR 0705	0.6	0.62	0.74	0.61

TABLE 4 Average PIC values for each *Mycosphaerella fijiensis* population, Wilde type, San Pablo, Zent, Cartagena provided by each VNTR locus. VNTR 3959 and 3786 loci presented the highest PIC values in all populations. VNTR 3959 locus presented the highest PIC value, 0.97.

3- Cytochrome b analysis

The point mutation in the *cytochrome b* gene was assayed in 665 *M. fijiensis* isolates in order to carry out strobirulin diagnosis. The strobilurin diagnosis indicated that the wild-type population was entirely sensitive. Only 8% of the Zent isolates showed to be sensitive to strobirulin, and two of the three commercial San Pablo and Cartagena populations were entirely resistant to strobirulin.

Figure 5 shows the molecular analysis of the strobirulin diagnosis. The 285bp-band indicated an internal control of the PCR and it was provided by the primers MFcytFor_1 and MFcytRev_1. Such band was present in all PCR products. The 200bp ~198 bp band was the diagnostic band for the strobirlulin sensitivity, it was provided by the primers MFcytFor_1 and MFcytRev_S. The other bands were unspecific and therefore the PCR condition will need to be optimized. The isolates that amplified the 198 bp-band were strobirulin sensitive, and the isolates that did not amplify this fragment were strobirulin resistant.

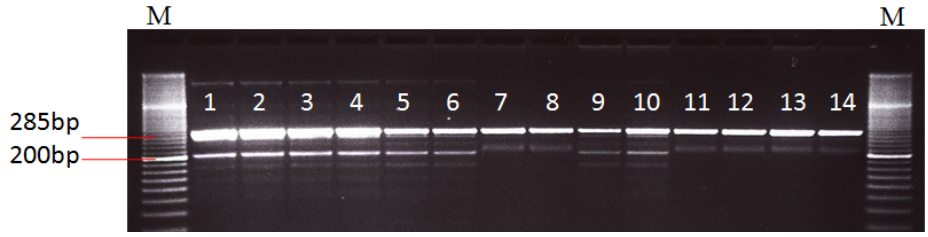


FIGURE 5 Strobirulin diagnosis. PCR products from reference isolates (1, 2, 3 and 4),Wild Type (5 and 6) , Zent (7, 8, 9, and 10), San Pablo (11 and 12) and Cartagena (13 and 14) isolates were applied in the agarose gel 3% stained with 0.3 µg/mL of ethidium bromide and visualized and photographed using UV transilluminator. *M* = EZ Load 20 pb Molecular Ladder (Invitrogen). The 285 bp band indicates the internal control of the PCR. The 200 bp band is the diagnostic band, which is present only in sensitive isolates.

The numbers 1, 2, 3 and 4 were respectively reference isolates, CIRAD 86 from Cameron, CIRAD 139 from Colombia, X845 from Indonesia, X846 from Phillipnes. They were used as a positive controls, since it is known that these isolates are sensitive to strobirulin. The numbers 5 and 6 were Wild Type isolates, and they presented a profile similar to the control isolates. Numbers 7, 8, 9, and 10 were isolates from the Zent population, two of them, 7 and 8 were resistant and 9 and 10 were sensitive. The numbers 11 and 12 were isolates from San Pablo and were resistant. The numbers 13 and 14 were isolates from Cartagena and were also resistant.

This result probably was related to the recommendations followed from the beginning of the strobirulin use in banana plantations in Costa Rica (Amil et al., 2007). However, our results show that the efforts made in order to prevent the increase of the strobirulin resistance were not satisfactory.

A study made from 2000 to 2003 evaluated the dynamics of strobirulin sensitivity in *M. fijiensis* isolates from Costa Rica in banana production areas submitted and not submitted to treatment with the fungicide. The results revealed there was no direct relationship between the number of applications and the frequency of the mutation on individual farms. On the other hand, the frequency converged toward regional averages, presumably due to the large-scale mixing of ascospores by wind. The immigration of resistant ascospores was detected as far as 6 km away both with and against the prevailing wind (Amil et al., 2007). Viable ascospores might be carried a few hundred kilometers by wind (Parnell et al., 1998). Therefore, if only one or two of the farms, San Pablo, Zent and Cartagena, quit using the fungicide, there will not be a decrease of resistant isolates, because those farms are 20-30 km apart.

References

AMIL, A.F.H.; STANGER, C.; SHAW, M.W. Dynamics of QoI sensitivity in *Mycosphaerella fijiensis* in Costa Rica during 2000 to 2003. **Proceedings of the American Phytopathological Society**, Saint Paul, v.97, n.11, p.1451-1457, Nov. 2007.

ARIAS, P.; DANKERS, C.; LIU, P.; PILKAUSKAS, P. **The world banana economy 1985-2002**. Rome: Food and Agriculture Organization of the United Nations, 2003. 94p. (FAO Commodity Studies).

BARLETT, D.W.C.; GODWIN, J.R.; HALL, A.A.; HAMER, M.; PARR-DOBRZANSK, B. The strobirulin fungicides. **Pest Management Science**, Sussex, v.58, n.7, p.649-662, July 2002.

CARLIER, J.; HADRAMI, A.; HAYDEN, H.; ZAPATER, M.F.; LAPEYRE, F. Populations study of *Mycosphaerella fijiensis* and genetic improvement of bananas for resistance to black leaf streak disease. In: INTERNATIONAL SYMPOSIUM ON THE MOLECULAR AND CELLULAR BIOLOGY OF BANANA, 1., 1999, New York. **Proceedings**... New York: Boyce Thompson Institute for Plant Research, 1999. p.40.

CARLIER, J.; LÉON, D.G.; ZAPATER, M.F.; LEBRUN, M.H. DNA restriction fragment length polymorphisms in Mycosphaerrella species that cause banana leaf spot diseases. **Phytopatology**, Saint Paul, v.84, n.7, p.751-756, July 1994.

CARLIER, J.; ZAPATER, M.F.; DUBOIS, C.; MOURICHON, X. Genetic structure of the global population of banana black leaf streak fungus *Mycosphaerellla fijiensis.* **Molecular Ecology**, Oxford, v.5, n.4, p.499-510, Aug. 1996.

CONDE-FERRAEZ, L.W.; CANTO-CANCHÉ, B.B.; KEMA, G.H.J.; CROUS, P.W.; JAMES, A.C.; ABELN, E.C. Isolation and characterization of the mating type locus of Mycosphaerella fijiensis, the causal agent of black leaf streak disease of banana. **Molecular Plant Pathology**, London, v.8, n.1, p.111-120, Jan. 2007.

DSOUZA, M.; LARSEN, N.; OVERBEEK, R. Searching for patterns in genomic data. **Trends in Genetics**, London, v.13, n.12, p.497-498, Dec. 1997.

GAUHL, F.; GAUHL, C.P.; JONES, D.R. Fungal diseases of the foliage: sigatoka leaf spots: black leaf streak, disease cycle and epidemiology. In:

JONES, D.R. (Ed.). **Diseases of bananas, abacá and ensete**. Wallingford: CAB International, 2000. p.56-62.

MARIN, D.H.R.; GUZMÁN, M.; SUTTON, T.B. Black sigatoka: an increasing threat to banana cultivation. **Plant Disease**, Quebec, v.87, n.3, p.208-222, Mar. 2003.

METZENBERG, R.L.; GLASS, N.L. Mating type and mating strategies in Neurospora. **BioEssays**, Cambridge, v.12, n.2, p.53-59, Feb. 1990.

MOBAMBO, K.N.; GAUHL, F.; VUYLSTEKE, D.; ORTIZ, C.; PASBERG-GAUHL, C.; SWENNEN, R. Yield loss in plantain from black Sigatoka leaf spot and field performance of resistant hybrids. **Field Crops Research**, Amsterdam, v.35, n.1, p.35-42, Oct. 1993.

MOURICHON, X.; ZAPATER, M.F. Obtention in vitro du stade *Mycosphaerella fijiensis*, forme parfaite de *Cercospora fijiensis*. **Fruits**, Paris, v.45, n.6, p.553-557, 1990.

NEU, C.; KAEMMER, D.; KAHL, G.; FISCHER, D.; WEISING, K. Polymorphic microsatellite markers for the banana pathogen *Mycosphaerella fijiensis*. **Molecular Ecology**, Oxford, v.8, n.3, p.523-525, Mar. 1999.

PARNELL, M.; BURT, P.J.A.; WILSON, K. The influence of exposure to ultraviolet radiation in simulated sunlight on ascospores causing black Sigatoka disease of banana and plantain. **International Journal of Biometeorology**, Lisse, v.42, n.1, p.22-27, Sept. 1998.

PARNELL, M.; WILSON, K. The influence of exposure to ultraviolet radiation in simulated sunlight on ascospores causing black Sigatoka disease of banana and plantain. **International Journal of Biometeorology**, Lisse, v.42, n.1, p.22-27, Sept. 1998.

PEREA, I.; ARANGO, E.R.; MÁRQUEZ, E.; ARANGO, R. Genetic diversity of Colombian isolates of Mycosphaerella fijensis Morelet based on microsatellite markers. **InfoMusa**, Montpellier, v.14, n.2, p.1-4, Dec. 2005.

RIVAS, G.G.; ZAPATER, M.F.; ABADIE, C.; CARLIER, J. Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas Mycosphaerella fijiensis. **Molecular Ecology**, Oxford, v.13, n.2, p.471-482, Feb. 2004.

ROMERO, R.A.S. Sensitivity of mycosphaerella fijiensis causal agent of black sigatoka of banana to propiconazole. **Phytopathology**, Saint Paul, v.87, n.1, p.96-100, Jan. 1997.

ROZEN, S.; SKALETSKY, H.J. Primer3 on the WWW for general users and for biologist programmers. In: KRAWETZ, S.; MISENER, S. (Ed.). **Bioinformatics methods and protocols**: methods in molecular biology. Totowa: Humana, 2000. p.365-386.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações de cruzamentos controlados**. 2.ed. Viçosa, MG: UFV, 2008. 562p.

SIEROTZKI, H.P.; 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 Management Science**, Sussex, v.56, n.10, p.833-841, Oct. 2000.

STOVER, R.H. International spread of banana leaf spot (*Mycosphaerella fijinensis* Leach). **Tropical Agriculture**, Surrey, v.39, p.327-338, 1962.

TURGEON, B.G.; YODER, O.C. Proposed nomenclature for mating type genes of filamentous ascomycetes. **Fungal Genetics and Biology**, Orlando, v.31, n.3, p.1-5, Oct. 2000.

WASHINGTON, J.R.C.; LOPEZ, F.; FAJARDO, M. Infection studies of Mycosphaerella fijiensis on banana and the control of black Sigatoka with chlorothalonil. **Plant Disease**, Quebec, v.82, n.11, p.1185-1190, Nov. 1998.

YANG, B.J.; ZHONG, S.B. Fourteen polymorphic microsatellite markers for the fungal banana pathogen *Mycosphaerella fijiensis*. **Molecular Ecology Resources**, Oxford, v.8, n.5, p.910-912, Mar. 2007.

CHAPTER 4

VARIABLE NUMBER OF TANDEM REPEAT MARKERS IN THE GENOME SEQUENCE OF *Mycosphaerella fijiensis*, THE CAUSAL AGENT OF BLACK LEAF STREAK DISEASE OF BANANA (*Musa* spp.), ENABLE RAPID GENETIC DIVERSITY ANALYSIS

ABSTRACT

Mycosphaerella fijiensis, the causal agent of banana leaf streak disease (commonly known as black Sigatoka), is the most devastating pathogen attacking bananas (*Musa* spp). Recently the whole genome sequence of *M. fijiensis* became available. This sequence was screened for the presence VNTR markers, and 42 primer pairs were selected for validation based on repeat type, length and number of repeat units. Eight VNTR markers showing multiple alleles in a reference set of isolates from different parts of the world were then tested on a set of 48 isolates collected in a banana farms in Costa Rica. A total of 30 alleles were identified in the reference set and 22 were identified in the population from Costa Rica. For the eight VNTR markers tested, in the reference set and natural population, the number of alleles per locus varied between 2 - 5 and 0 - 6, respectively. The markers allowed robust scoring on agarose gels and proved useful for variation and population genetic studies using basic molecular instrumentation.

RESUMO

Mycosphaerella fijiensis, agente causal da Sigatoka Negra, sendo considerado o patógeno mais devastador da bananeira (*Musa* spp). Recentemente, a sequência completa do genoma de *M. fijiensis* tornou-se disponível. Esta sequência foi rastreada para a presença de marcadores VNTR e 42 pares de primers foram selecionados para a validação dos maracadores com base no tipo de repetição, tamanho e número de unidades de repetição. Oito marcadores VNTR que apresentaram alelos múltiplos em um conjunto de isolados de referência provenientes de diferentes partes do mundo, foram também testados em um conjunto de 48 isolados coletados em uma plantação de banana da Costa Rica. Um total de 30 alelos foram identificados nos isolados de referência e 22 foram identificados na população da Costa Rica. Para os oito marcadores VNTR testados nos isolados de referência e da população natural, o número de alelos por loco variou de 2 a 5 e de 0 a 6, respectivamente. Os marcadores VNTR foram facilmente detectados em géis de agarose e também provaram ser úteis para estudos de diversidade genética de populações usando equipamentos básicos de laboratório.

Banana leaf streak disease (also known as black Sigatoka), caused by the heterothallic fungus *Mycosphaerella fijiensis*, is the most devastating disease affecting bananas (*Musa* spp.) (Marin et al., 2003; Cordeiro et al., 2004; Conde et al., 2007). In humid, high rainfall areas, even well-managed plantations are weekly sprayed with fungicides throughout the year, in order to control this disease. These inputs often represent a considerable percentage of total production costs and add-up to hundreds of millions of US dollars per year for the global banana industry.

Black Sigatoka was originally identified in Fiji in 1963 (Rhodes, 1964) and nowadays it is present in most important banana growing areas worldwide (Arzanlou et al., 2007). *M. fijiensis* is a heterothallic Ascomycete, which is haploid for the vegetative part of its lifecycle. The sexual cycle results in air-borne ascospores that play an important role in the epidemiology of this disease. After deposition, germ tubes of ascospore or asexual conidia penetrate the stomata and during the biotrophic phase, the filamentous hyphae grows slowly and intercellularly in the mesophyll. This is followed by a necrotrophic phase, resulting in abundant cell death that turns the leaves fully necrotic. As a result, *M. fijiensis* decreases photosynthesis, reduces fruit size and induces early maturation in susceptible banana cultivars. *M. fijiensis* is related to the Sigatoka pathogen *M. musicola*, however it is much more aggressive and hence, it is globally replacing that pathogen even at higher altitudes (Arzanlou et al., 2007).

The objective of the current work is to present a set of variable number of tandem repeat markers (VNTRs) that were identified in the *M. fijiensis* genome sequence. The versatility and efficiency of these markers in genotyping and studying diversity were tested for both laboratory and field *M. fijiensis* isolates.

In 2007, a ~7.1X equivalent of the estimated 73.4 Mbp of the *M. fijiensis* (isolate CIRAD086) genome (http://genome.jgi-

psf.org/Mycfi1/Mycfi1.home.html) was produced at the Joint Genome Institute of the United States Department of Energy (http://www.jgi.doe.gov/). The consensus sequences of this genome were submitted to analysis in a bioinformatics' pipeline for VNTR detection, consisting of a primer development module. This pipeline chops genomic sequences into 1,500 bp size fragments with a 100 bp overlap, then detects VNTRs using a Perl script allowing imperfect repeat VNTR detection by using the scan-for-matches program (Dsouza et al., 1997) and finally generates a maximum of five primer sets that could amplify a VNTR by PCR using the Primer3 (Rozen & Skaletsky, 2000) software followed by a screening for unique primer pairs. Additional optimization of the primer pairs was performed using the Fast PCR Professional 5.2 software (http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm).

The initial screening for polymorphism was carried out using a reference set of eight *M. fijiensis* isolates: CIRAD086 (Cameroon, mat1-1), CIRAD139a (Colombia, mat1-2), and isolates CBS845 (Indonesia), CBS846 (Philippines), CBS848 (New Caledonia), CBS849 (Burundi), CBS851 (Gabon), and CBS852 (Tanzania) from the Fungal Diversity Centre (http://www.cbs.knaw.nl/) Utrecht, The Netherlands. The validation and versatility with the emphasis on differentiation and allele size stability were assessed on a population of *M. fijiensis* isolates (48 individuals), obtained from single ascospore cultures, that had been collected in the San Pablo banana plantation in Costa Rica.

Total DNA was isolated from 30 mg of mycelium using the Wizard®Magnetic DNA Purification System for food (PROMEGA food kit, Wisconsin, USA), following the manufacture's instructions; re-suspended in 100 µL of nuclease-free water, and stored at -20° C until use. VNTR markers were amplified in a 20 µL PCR reaction aliquot consisting of: 1X reaction buffer [10 mM Tris-HCl (pH 9.0), 1.5 mM MgCl2, 50 mM KCl], 10 ng of total DNA, 20 pmol of each forward and reverse primer, 200 µM of each dNTPs, and 0.2 units

of Taq (Roche Diagnostics, Leiden, The Netherlands). PCR was carried out in the PTC-200 Peltier Thermal Cycler (MJ Research, Waltham, MA) with the following amplification program: 94° C for 2 min., 13 cycles of 94° C for 30 sec, 66° C for 30 sec (-1°C / cycle), and 72° C for 30 sec; followed by 28 cycles of 94° C for 30 sec, 53° C for 30 sec, and 72° C for 30 sec and an extension cycle of 72° C for 7 min. The amplicons were separated in 3% agarose gels (Roche Diagnostics, Leiden, NL) in TBE buffer with 0.5 µg/mL ethidium bromide, followed by visualization under an UV transilluminator and estimated using a 1 kb molecular marker (Invitrogen, California-USA). A total of 1,528 candidate VNTR loci were identified in the *M. fijiensis* genome. Several represented imperfect mono-, di-, or tri- nucleotide repeats. A set of 42 VNTR markers displaying a larger number of repeat units, as well as repeats with a high level of sequence conservation between the repeat units, were selected for further analysis. Eight VNTRs (VNTR_0252, VNTR_0705, VNTR_1333, VNTR_1890, VNTR_3786, VNTR_3831, VNTR_3907, and VNTR_3959) (Table 1), with variable (4-12) repeat units, showed polymorphism in the reference set and were tested on the population that was collected from the aforementioned San Pablo farm (Figure 1).

The VNTR primers generated a total of 30 alleles in the *M. fijiensis* reference set and 22 in the San Pablo population. Nine alleles are shared between the reference set and the San Pablo population. Fourteen alleles present in the population were not identified in the reference set. The VNTR 3959_6594, presented a higher number of alleles in the San Pablo population than in the reference set. The VNTR 3907-122975 did not show any allele in this population (Table 1). The number of alleles per locus, in the reference set and natural population, varied between 2 - 5 and 0 - 6, respectively. For the fifty-two total polymorphic loci evaluated, the PIC value varied from 0.30 to 0.75 and

from 0.0 to 0.82, with averages of 0.57 and 0.32, for the reference set and natural San Pablo population, respectively. For the joint analysis, reference set and natural population, the PIC value varied from 0.16 to 0.85, with mean PIC value of 0.46. Fifty-six percent of the VNTR markers showed a PIC value higher than 0.50.

The versatility of VNTR markers as shown in this report is a great advantage. They are easy to score and can be run in agarose gels making them an interesting tool specially attractive because highly equipped laboratories are not required. Rapid population genetic analysis are very important in the air-borne black leaf streak disease that is a great threat to global banana production and is still occupying new territories. Due to the intensity of fungicide applications, resistance development is a frequently occurring phenomenon; therefore, rapid diagnostics for population analysis are indispensable. The markers presented here provide a useful and unique set of VNTRs that can be used in genetic variability and population studies, also in the support of fungicide resistance management and breeding strategies to control banana black leaf streak disease.

References

ARZANLOU, M.; ABELN, E.C.A.; KEMA, G.H.J.; WAALWIJK, C.; CARLIER, J.; CROUS, P.W. Molecular diagnostics in the Sigatoka disease complex of banana. **Phytopatology**, Saint Paul, v.97, n.10, p.1112-1118, Oct. 2007.

CONDE, L.; WAALWIJK, C.; CANTO-CANCHÉ, B.B.; KEMA, G.H.J.; CROUS, P.W.; JAMES, A.C.; ABELN, E.C.A. Isolation and characterization of the mating type locus of Mycosphaerella fijiensis, the causal agent of black leaf streak disease of banana. **Molecular Plant Pathology**, London, v.8, n.1, p.111-120, Jan. 2007.

CORDEIRO, Z.J.M.; MATOS, A.P.M.; SILVA, S.O.; GASPAROTTO, L.; CAVALCANTE, M.J.B. Impact and management of black-Sigatoka in Brasil. In: REUNIÃO INTERNACIONAL ACORBAT, 16., 2004, Joinvile. **Anais**... Joinvile: ACORBAT, 2004. p.63-69.

DSOUZA, M.; LARSEN, N.; OVERBEEK, R. Searching for patterns in genomic data. **Trends in Genetics**, London, v.13, n.12, p.497-498, Dec. 1997.

MARIN, D.; ROMERO, R.; GUZMÁN, M.; SUTTON, T.B. Black sigatoka: an increasing threat to banana cultivation. **Plant Disease**, Quebec, v.87, n.3, p.208-222, Mar. 2003.

RHODES, P.L. A new banana disease in Fiji. **Phytopathological News**, Kew, v.10, p.38-41, 1964.

ROZEN, S.; SKALETSKY, H.J. Primer3 on the WWW for general users and for biologist programmers. In: KRAWETZ, S.; MISENER, S. (Ed.). **Bioinformatics methods and protocols**: methods in molecular biology. Totowa: Humana, 2000. p.365-386.

ANEXOS

Name	Primer Sequence	Genome Position	Repeat Unit	Size (bp)	Number of alleles	
					Reference set	Natural population
VNTR_3907-122975	F: AATAGAGGCGCCGAGATAGG R: CTCTTACTAGGTGCCCTTGCC	Scafold 36-1c 545605/546088	GGCAAGGGTAAG (4)	100	2	0
VNTR_1890-11246	F: GTTATTGCCGATTAGGCCG R: ATACGTTAGATTCGCGGCAG	Scafold 10-1c 1483445/1484487	GAATCTACCGC (4)	104	3	2
VNTR_1333-7690	F: GAGTGAAGTACTGCGGAGGC R: AGTTGGAGAAAGGCGAAAGG	Scafold 7-1c 1163754/1164780	CGCCTTT (12)	337	4	1
VNTR_0705-143294	F: ATAGGATGCGGCAGACACTC R: CGTGCTGATTTGAAGTGCC	Scafold 3-1c 242566/243620	ACCACTCACCACTC (6)	214	4	2
VNTR_3959-6594	F: GCGCGAGGCTTTCTATCTC R: ACCCCGATTAGGGAAGGTC	Scafold 38-1c 96151/96578	TATCTT (11)	242	4	6
VNTR_0252-102618	F: TAGAGGCTACCCTGCCGTC R: GTATACTTCCGACCTCGGGC	Scafold 1-1c 5638704/5639742	CTAGCTAATA (6)	136	3	2
VNTR_3831-152531	F: AGCCCCGAAGGAGGGTAG R: CGAGCCTACGACTAGCCCG	Scafold 35-1c 305952/305981	TAGAAAGAAAATAGC (8)	350	5	5
VNTR_3786-91361	F: GCAGCGGAGTGCTAGTAACC R: CGCGCTTTTGACTCTTCTTC	Scafold 34-1c 157858/158319	AGTGGAAGA (8)	344	5	4

TABLE 1 Validation of variable number of tandem repeat markers (VNTRs) that were identified in the genome of *Mycosphaerella fijiensis* (isolate CIRAD086) on a reference set of eight *M. fijiensis* isolates originating from Cameroon, Colombia, Indonesia, The Philippines, New Caledonia, Burundi, Gabon and Tanzania as well as a wild natural *M. fijiensis* population in Costa Rica.

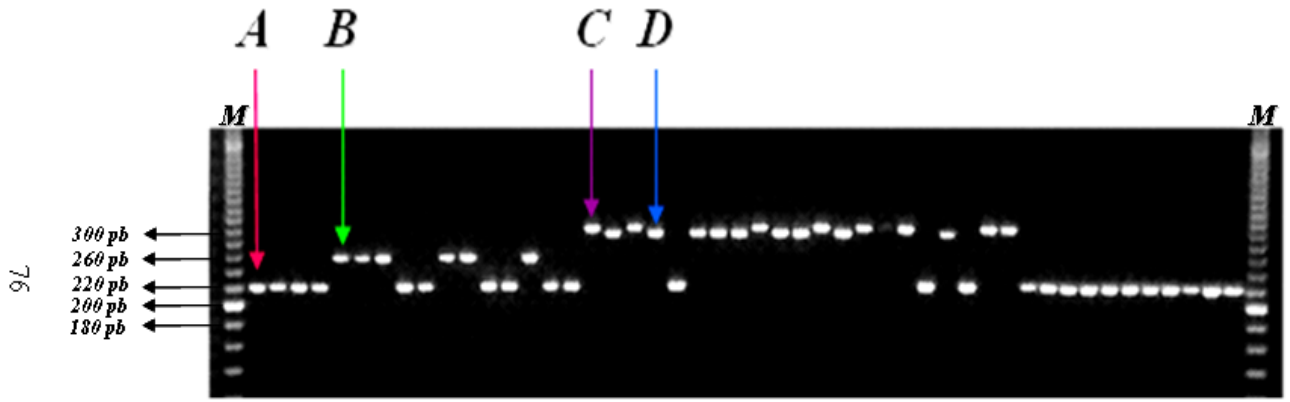


FIGURE 1 Validation of variable number of tandem repeat marker VNTR_3786 primers that generated four alleles (A, B, C, and D) on a set of 48 single ascospore of *Mycosphaerella fijiensis* isolates that were obtained from a natural population in the San Pablo banana plantation in Costa Rica. *M* = EZ Load 20 pb Molecular Ladder.