Agricultura
Ministério da Agricultura, Pecuária e Abastecimento

Boletim de Pesquisa 107
e Desenvolvimento ISSN 1676 - 1340
Outubro, 2005

DATA*MUSA*

BANCO DE DADOS DE GENÔMICA DE *MUSA* SPP

Embrapa

Embrapa
Recursos Genéticos e Biotecnologia

ISSN 1676 - 1340
Outubro, 2005

Boletim de Pesquisa e Desenvolvimento 107

DATA*MUSA*

BANCO DE DADOS DE GENÔMICA DE *MUSA* SPP

Manoel Teixeira Souza Júnior

Candice Mello Romero Santos

Natália Florêncio Martins

Felipe Rodrigues da Silva

Roberto Coti Togawa

Lisângela Pinheiro Cassiano

Elionor Rita Pereira de Almeida

Marly Catarina Felipe Coelho

Alexandre Rodrigues Caetano

Ana Yamaguishi Ciampi

Marcos Mota

Pietro Piffanelli

Robert Neil Gerald Miller

Brasília, DF
2005

Exemplares desta edição podem ser adquiridos na

Embrapa Recursos Genéticos e Biotecnologia
Serviço de Atendimento ao Cidadão
Parque Estação Biológica, Av. W/5 Norte (Final) –
Brasília, DF CEP 70770-900 – Caixa Postal 02372 PABX: (61) 3348-4739 Fax: (61) 3340-3666 http://www.cenargen.embrapa.br
e.mail:sac@cenargen.embrapa.br

1ª edição
1ª impressão (2005):

D 232 DATA*Mus*a: banco de dados de genômica de *Musa* spp / Manoel Teixeira Souza Júnior ... [et al.]. – Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2005.
24 p. – (Embrapa Recursos Genéticos e Biotecnologia / Boletim de pesquisa e desenvolvimento, 1676 – 1340; 107)

1. Banana - programa de genômica – biotecnologia. 2. DATA*Musa* - banco de dados - genômica de banana. 3. DATA*Musa* - banco de dados - genômica Estrutural. 4. DATA*Musa* - banco de dados – transcriptoma. 5. DATA*Musa* - banco de dados - análogos de genes de resistência. I. Souza Júnior, Manoel Teixeira. II. Série.

634.772 – CDD 21.

SUMÁRIO

Resumo **6**
Abstract **8**
Introdução **9**
Material e Métodos **10**
1. Seqüências de DNA obtidas a partir de subclones de BAC **10**
2. Seqüências de cDNA obtidas a partir de sete bibliotecas de folhas, raízes, casca verde e flor masculina de *Musa acuminata* **11**
3. Seqüências de DNA obtidas a partir de análogos de genes de resistência (RGAs) **14**
4. Sequenciamento **16**
5. Avaliação da qualidade e limpeza das seqüências **16**
6. Agrupamento das seqüências **17**
7. Identificação dos genes **17**
Resultados **17**
Considerações finais **19**
Referências Bibliográficas **20**

DATA*MUSA*

BANCO DE DADOS DE GENÔMICA DE *MUSA* SPP

Manoel Teixeira Souza Júnior[1]
Candice Mello Romero Santos[2]
Natália Florêncio Martins[3]
Felipe Rodrigues da Silva[4]
Roberto Coti Togawa[5]
Lisângela Pinheiro Cassiano[6]
Elionor Rita Pereira de Almeida[7]
Marly Catarina Felipe Coelho[8]
Alexandre Rodrigues Caetano[9]
Ana Yamaguishi Ciampi[10]
Marcos Mota[11]
Pietro Piffanelli[12]
Robert Neil Gerald Miller[13]

Resumo

A Embrapa Recursos Genéticos e Biotecnologia, a Universidade Católica de Brasília (UCB) e o Centro Francês de Pesquisa Agrícola para o Desenvolvimento Internacional (CIRAD) iniciaram, em 2002, o projeto de pesquisa intitulado “Análise da Estrutura Primária do Genoma A de *Musa acuminata*”, financiado pelo Conselho Nacional de Pesquisa e Desenvolvimento (CNPq), e com o objetivo de

[1] PhD - Embrapa Recursos Genéticos e Biotecnologia.
[2] Dr - Embrapa Recursos Genéticos e Biotecnologia.
[3] PhD - Embrapa Recursos Genéticos e Biotecnologia.
[4] Dr - Embrapa Recursos Genéticos e Biotecnologia.
[5] PhD - Embrapa Recursos Genéticos e Biotecnologia.
[6] MsC - Embrapa Recursos Genéticos e Biotecnologia.
[7] PhD - Embrapa Recursos Genéticos e Biotecnologia.
[8] MsC - Embrapa Recursos Genéticos e Biotecnologia.
[9] PhD - Embrapa Recursos Genéticos e Biotecnologia.
[10] Dr - Embrapa Recursos Genéticos e Biotecnologia.
[11] PhD - Embrapa Recursos Genéticos e Biotecnologia.
[12] PhD - Parco Tecnológico Padano / Lodi -Italy
[13] PhD - Universidade Católica de Brasília(UCB)

desenvolver as bases de um programa de genômica e biotecnologia de banana no Brasil. Este projeto resultou na criação do DATA*Musa*, um banco de dados de genômica de banana composto de informações de Genômica Estrutural (seqüências de clones de uma biblioteca de BAC), de Transcriptoma (Expressed Sequence Tags - ESTs) e de Análogos de Genes de Resistência (Resistance Genes Analogs - RGAs). As análises das seqüências hoje presente no DATA*Musa* permitiram a identificação de 5.317 *Musa acuminata* Assembled ESTs Sequence (MaAES), 113 genes e suas seqüências promotoras de expressão, e dezenas de RGAs. Uma análise preliminar dessas seqüências já permitiu identificar diversos genes candidatos de interesse para uso em apoio ao melhoramento genético da bananeira; cabendo destacar entre eles: análogos de genes de resistência a fungos, bactérias, a nematóides, insetos e estresses abióticos. O uso das informações do DATA*Musa* irá permitir ampliar as possibilidades de melhoramento genético e de transgenia direcionados para a cultura da banana, com vistas à geração de novas variedades superiores.

Abstract

Embrapa Genetic Resources and Biotechnology, the Catholic University of Brasilia (UCB) and the French Center of Agricultural Research for International Development (CIRAD) initiated, in 2002, the research project entitled "Analysis of the Primary structure of Genome of *Musa acuminata*", with financial support from The National Council for Scientific and Technological Development (CNPq), and with the objective of developing the basis for a banana genomic and biotechnology research & development program in Brazil. This project resulted in the creation of the DATA*Musa*, a banana genomic data base composed by information on Structural Genomics (sequences from BAC clones), Transcriptome (ESTs) and Resistance Genes Analogous (RGAs). The analysis of the sequences present in the DATA*Musa* database has allowed the identification of 5.317 *Musa acuminata* Assembled ESTs Sequence (MaAES), 113 genes and their promoter sequences, as well as several RGAs. A preliminary analysis of the data has already led to the identification of several candidate genes of potential use in banana genetic improvement, such as: genes for resistance to fungi, bacteria, nematodes, insects and abiotic stresses. The use of information from the DATA*Musa* database will increase the possibilities of development of new varieties by conventional breeding and transgenic strategies.

Introdução

A bananeira (*Musa* ssp.) é cultivada de Norte a Sul do País, sendo fundamental para a complementação da dieta alimentar das populações de baixa renda. O Brasil é o segundo maior produtor mundial de bananas, tendo produzido 6.469,470 Mt (9.5% da produção mundial) no ano 2003, em uma área de 507,000 hectares (FAO, 2004). Praticamente toda fruta produzida no Brasil é comercializada no mercado interno. A maioria dos bananicultores é composta por pequenos produtores, e o setor da bananicultura gera mais de 500 mil empregos diretos no Brasil.

O Programa Internacional para o Melhoramento de *Musa* (PROMUSA), ligado à Rede Internacional para o Melhoramento de Banana e Plátano (INIBAP - www.inibap.org), é um mecanismo de colaboração e troca de informações entre pesquisadores envolvidos no melhoramento genético de *Musa* no mundo. Em 2001, o PROMUSA incentivou a formação e abrigou o consórcio internacional do genoma banana (Global *Musa* Genomics Consortium - GMGC), dos quais a Embrapa Recursos Genéticos e Biotecnologia, a Universidade Católica de Brasília (UCB), e o Centro Francês de Pesquisa Agrícola para o Desenvolvimento Internacional (CIRAD) são membros fundadores. O GMGC tem como objetivo decifrar o genoma de *Musa* para com isso garantir a sustentabilidade da banana como alimento básico para grande parte da população mundial. Isto será alcançado através do maior entendimento da genética e do genoma deste gênero, possibilitando a elaboração de novas estratégias de melhoramento genético e de transgenia direcionada.

O projeto de pesquisa intitulado “Análise da Estrutura Primária do Genoma A de *Musa acuminata*”, financiado pelo Conselho Nacional de Pesquisa e Desenvolvimento (CNPq), e executado pela Embrapa Recursos Genéticos e Biotecnologia, em parceria com a UCB e o CIRAD no período de fevereiro de 2002 a junho de 2005, resultou na criação do DATA*Musa*. O DATA*Musa é* hoje o segundo maior banco de dados de genômica de banana e é composto de informações de genômica estrutural (seqüências completas de clones de biblioteca

de BAC, de transcriptoma (Expressed Sequence Tags – ESTs) e de análogos de genes de resistência (Resistance Genes Analogs – RGAs).

Este boletim técnico descreve de forma sucinta o banco de dados DATA*Musa.*

Material e Métodos

1. Seqüências de DNA obtidas a partir de subclones de BAC

Cinco clones da biblioteca de BAC de *Musa acuminata* ssp. *burmannicoides* - variedade Calcutta 4 (VILARINHOS et al., 2003) foram selecionados utilizando-se cinco diferentes sondas obtidas junto ao projeto EGRAM - The European comparative gramineae mapping programme, e que estavam à disposição do CIRAD em Montpellier, França. Após a seleção, os clones de BAC foram preliminarmente caracterizados mediante sequenciamento das extremidades e geração de perfil de restrição com diferentes enzimas. As bibliotecas de subclones dos clones de BAC foram geradas, cada uma, com 3.072 clones, os quais foram arranjados em oito placas de 384 poços. O vetor utilizado para a produção da biblioteca de subclones foi o pcDNA 2.1 (Invitrogen Life Technologies, USA), que recebeu fragmentos de DNA na faixa de tamanho entre 5 e 10 Kb. A seleção dos clones de BAC e produção das bibliotecas shotgun foi realizada em laboratórios na França, dentro da cooperação estabelecida com o CIRAD.

Tabela 1. **Número de seqüências de DNA por subclone de BAC depositadas no DATA*Musa*.**

Nome do subclone de BAC	Nº de seqüências
MA4_BAC008L021	2.511
MA4_BAC042M013	2.491
MA4_BAC078I012	7.104
MA4_BAC106O017	2.385
MA4_BAC111B014	5.648
Total	**20.139**

2. Seqüências de cDNA obtidas a partir de sete bibliotecas de folhas, raízes, casca verde e flor masculina de *Musa acuminata*

A análise de transcriptoma de banana realizada no projeto de pesquisa "Análise da Estrutura Primária do Genoma A de *Musa acuminata*" utilizou seqüências de DNA obtidas a partir de sete bibliotecas de cDNA de banana (Tabela 2).

Cinco destas sete bibliotecas foram produzidas e caracterizadas na Embrapa Recursos Genéticos e Biotecnologia, tendo sido utilizado material vegetal da variedade Calcutta 4. Esta é uma variedade diplóide (*Musa* Germplasm Information System - MGIS accession number ITC0249) que pertence à seção EUMUSA. As plantas utilizadas foram obtidas inicialmente no Banco Ativo de Germoplasma de banana da Embrapa Mandioca e Fruticultura, em Cruz das Almas-Bahia (www.cnpmf.embrapa.br).

Santos et al. (2005) descrevem a produção e caracterização de duas bibliotecas de cDNA de folha de *M. acuminata* ssp. *burmannicoides* var. Calcutta 4 (AA) submetidas a estresse de temperatura (calor e frio). Para a construção da biblioteca de estresse de alta temperatura foram utilizadas folhas de plantas micropropagadas *in vitro* e que tinham sido aclimatadas em câmaras de ambiente controlado (80% de umidade relativa, 25 °C, e regime de 14 horas de luz). A

primeira coleta de folhas (H0) foi realizada após as plantas terem sido submetidas à temperatura de 25 °C por 72 horas. Após a primeira coleta a temperatura foi elevada para 35 °C, e novas amostras foram coletadas após uma (H1), duas (H2) e três (H3) horas da mudança na temperatura. A temperatura foi então elevada a 45 °C, e novas amostras foram coletadas após uma (H4), duas (H5), três (H6), nove (H7), e 21 horas (H8) da mudança na temperatura.

Para a construção da biblioteca de estresse de baixa temperatura, amostras de folha (C0) foram coletadas de plantas sob temperatura de 25 °C por 72 horas. Após a primeira coleta a temperatura foi reduzida para 15 °C, e novas amostras foram coletadas após uma (C1), duas (H2) e três (H3) horas da mudança na temperatura. A temperatura foi então reduzida para 5 °C, e novas amostras foram coletadas após uma (C4), duas (C5), três (C6), nove (H7), e 21 horas (H8) da mudança na temperatura. As folhas de bananeira coletadas foram imediatamente congeladas em nitrogênio líquido e transferidas para o freezer a -80 °C até o início do processo de extração do RNA total.

Para a produção das bibliotecas de cDNA de flores masculinas e casca verde de banana, os respectivos tecidos foram coletados de plantas de *M. acuminata* ssp. *burmannicoides* var. Calcutta 4 (AA) mantidas no campo experimental da Embrapa Recursos Genéticos e Biotecnologia. Para a produção da biblioteca de cDNA de raízes, raízes foram coletadas de plantas desta variedade mantidas em cultivo *in vitro*. O material vegetal depois de coletado, foi imediatamente congelado em nitrogênio líquido e transferido para freezer - 80 °C até início do processo de extração de RNA total.

Duas bibliotecas de cDNA foram construídas a partir de folhas de banana infectadas *in vitro* com o agente causal da Sigatoka Negra, o ascomiceto *Mycosphaerella fijiensis* Morelet (Forma perfeita) e *Paracercospora fijiensis* (Morelet) Deighton (Forma imperfeita). Uma das bibliotecas foi construída com a variedade *M. acuminata* ssp. *burmannicoides* var. Calcutta 4 (AA), altamente resistente a este fungo, enquanto que a outra foi construída com a variedade do subgrupo Cavendish denominada Grand Naine (AAA), altamente susceptível ao fungo. Estas duas bibliotecas de cDNA foram produzidas nas dependências do

CIRAD em Montpellier, França, haja vista que a Embrapa Recursos Genéticos e Biotecnologia não tinha permissão para manusear este fungo em Brasília.

Tabela 2. Número total de seqüências de ESTs obtidos a partir das bibliotecas de cDNA geradas no projeto.

Bibliotecas	***Nº de seqüências***
Folhas de bananeira submetidas ao estresse de alta temperatura	1.440
Folhas de bananeira submetidas ao estresse de baixa temperatura	1.440
Flor masculina	1.944
Raízes de plantas cultivadas *in vitro*	2.155
Casca	2.420
Folhas de bananeira infectadas com *Mycosphaerella fijiensis* em estágio inicial de infecção	3.902
Folhas de bananeira infectadas com *Mycosphaerella fijiensis* em estágio avançado de infecção	3.812
Total	**17.113**

O RNA total foi isolado das folhas (sadias e infectadas) e raízes, utilizando-se o kit “Plant RNA Reagent” (Invitrogen Life Technologies, USA), de acordo com o protocolo apresentado pelos fabricantes, enquanto que RNA total da casca verde e das flores masculinas foram isolados de acordo com Chang et al. (1993). Para todas as bibliotecas, o RNA total foi submetido à purificação de poli (A) + RNA utilizando-se o kit “Micro FastTrack 2.0 mRNA Isolation” (Invitrogen Life Technologies, USA), seguindo protocolo apresentado pelos fabricantes. As

bibliotecas foram construídas utilizando-se o kit “Creator Smart cDNA library” e o vetor pDNR-LIB (Clontech Laboratories, Inc., USA), também segundo protocolo apresentado pelos fabricantes.

3. Seqüências de DNA obtidas a partir de análogos de genes de resistência (RGAs)

Plantas de Calcutta 4 foram micropropagadas *in vitro*. Folhas jovens foram coletadas e utilizadas para extração de DNA genômico. O ácido nucléico foi extraído de acordo com o método de CTAB (ROGERS e BENDICH, 1988), quantificado em géis de agarose de 1% via comparação com o “Low DNA Mass Ladder” (Invitrogen Life Technologies, USA), e sua qualidade examinada via digestão com a endonuclease *Eco*R1, e amplificação com 10mer primers em reação de RAPD.

A amplificação por PCR tendo como alvo os “motifs” conservados no domínio NBS em *Arabidopsis* foi conduzida usando combinações de oito primers degenerados, tendo como primers senso: P1a, P1b (BERTIOLI et al., 2003) e LM638-for (KANAZIN et al., 1996) e primers anti-senso: P3a, P3d (BERTIOLI et al., 2003) e RNBS-D-rev (PEÑUELA et al., 2002). Oito combinações de primers foram testadas: LM638-P3A, LM638-RNBS-D, P1A-P3A, P1A-P3D, P1A-RNBS-D, P1B-P3A, P1B-P3D, and P1B-RNBS-D) (Tabela 3). Sete combinações de primers degenerados foram também testados, com primers desenhados a partir de motivos conservados em domínios NBS e LRR em monocotiledôneas, e motivos não-TIR em *Arabidopsis* (Tabela 3), usando as programas HMMSearch (Hammer package), MEME Motif discovery, e Codehop. As combinações de primers testados foram: 39F1-1R1, 1F-P3b, P1c-P3b, 3F2-13R1, 3F2-11R1, 2F-13R1, 2F-11R1.

As reações de amplificação foram realizadas em 25 µl de solução contendo 50 ng de DNA genômico, 0,2 mM de cada dNTP, 0.5µM de cada primer, 1.25 unidades de Taq DNA Polymerase, 1X de Taq polymerase buffer, 2,5 mM de $MgCl_2$, completando para o volume final com água milliQ autoclavada. O programa empregado foi de um ciclo de 96°C/5’, 35 ciclos de 96°C/1’, 40°C/1’ e

72°C/1', e um ciclo de 72°C/10'. Após a amplificação, procedeu-se à separação dos fragmentos de DNA em gel de 1% de agarose, em TBE 1X, e visualização sob luz ultravioleta.

Os produtos de PCR foram purificados usando o kit de purificação Qiagen QIAquick PCR purification kit (Qiagen, USA) e clonados usando o vetor comercial pCR2.1TOPO (Invitrogen Life Technologies, USA.) ou pGEM-T-Easy (Promega, USA). A ligação dos produtos de PCR com os vetores descritos acima, foi dialisada usando membranas de 0.02μM, e células competentes de DH5α *Escherichia coli* transformados por eletroporação ou choque térmico. As células transformadas foram plaqueadas em meio LB de seleção com ampicilina (100μg/ml), X-Gal (20mg/ml) e IPTG (200mg/ml), e crescidas na estufa a 37ºC. Os clones recombinantes foram selecionados e o DNA extraído e purificado através de lise alcalina (SAMBROOK e RUSSELL, 2001).

Tabela 3. Seqüências de primers degenerados e motivos alvos para isolamento de RGAs.

Primer	Motivos	Seqüência do Primer (5' a 3')
P1a (forward)	P-loop	GGIATGCCIGGIIIIGGIAARACIAC*
P1b (forward)	P-loop	GGIATGGGIGGIIIIGGIAARACIAC
LM638-for (forward)	P-loop	GGIGGIGTIGGIAAIACIAC
P3a (reverse)	GLPL	AIITYIRIIRYIAGIGGYAAICC
P3d (reverse)	GLPL	AIITYIRIIRYYAAIGGIAGICC
RNBS-D-rev (reverse)	RNBS-D non-TIR	GGRAAIARISHRCARTAIVIRAARC
39F1 (forward)	Non NBS (n-terminal) monocot	TCATCAAGGACGAGCTGgarwbnatgma
1F (forward)	P-loop – GKTT monocot	GGCGGGGTGGGCaaracnacnht
P1c (forward)	P-loop - GKTT non TIR *Arabidopsis*	GGICGICCIGGIIIIGGIAARACIAC
3F2 (forward)	Kinase 2	GAGGTACTTCCTGGTGCTGgaygayrtbtgg
2F (forward)	RNBS-2	AACGGCTGCAGGATCATGrtbachachmg
1R1 (reverse)	P-loop	CGTGCTGGGCCAGGgtngtyttncc
P3b (reverse)	GLPL – non Tir *Arabidopsis*	AIITYIRIIRYIAGIGGIAGICC
13R1 (reverse)	LRR - C-terminal monocot	CGGCCAAGTCGTGCAyvakrtcrtgca
11R1 (reverse)	LRR - C-terminal monocot	TCAGCTTGCCGATCCACtydggsagbyt

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

4. Sequenciamento

As extremidades 5' dos clones de cDNA, as 5' e 3' dos subclones de BAC, e as 5' e 3' dos clones RGAs foram seqüenciadas na plataforma de sequenciamento de DNA da Embrapa Recursos Genéticos e Biotecnologia (http://www/laboratorios/psd/psd.html) e da UCB, utilizando-se os primers "M13 senso" (5`- TGT AAA ACG ACG GCC AGT - 3`), "M13 anti-senso"(3´ - CAG GAA ACA GCT ATG ACC - 5´), e o seqüenciador automático ABI3700. Os eletroferogramas gerados foram então submetidos ao Sistema GENOMA da Embrapa Recursos Genéticos e Biotecnologia (http://genoma.cenargen.embrapa.br/genoma/) e estocados no *MUSA*_ESTs database do Laboratório de Bioinformática até processamento e análise de seqüências.

5. Avaliação da qualidade e limpeza das seqüências

Os eletroferogramas gerados no sequenciamento dos clones de cDNA foram inicialmente analisados pelo programa Phred (EWING et al., 1998), que

avaliou a qualidade dos picos correspondentes a cada base seqüenciada, conferindo um valor de qualidade a cada uma. Para esta análise foram estabelecidos os parâmetros de aceitação das seqüências conforme Telles e Silva (2001) com Phred superior a 20, correspondendo a um erro a cada 1000 bases. A remoção de seqüências ribossomais, de poli-(A), de seqüências de baixa qualidade, de regiões do vetor e de adaptadores, foi conduzida conforme Telles e Silva (2001).

6. Agrupamento das seqüências

As seqüências de alta qualidade foram submetidas à montagem utilizando o programa CAP3 (HUANG e MADAN, 1999).

7. Identificação dos genes

Utilizando o programa BLASTx (ALTSCHUL et al., 1997), com filtro para de resultados $<10^{-5}$, verificou-se a similaridade entre os "MaAES" e as seqüências no banco de dados GenBank nr (BENSON et al., 2002), MIPS *Arabidopsis thaliana* (SCHOOF et al., 2002) e SwissProt (GASTEIGER et al., 2001), e para predizer a função das proteínas codificadas pelos genes expressos realizou-se o Blastx contra o banco de dados KOG (Eukaryotic Orthologous Groups) (TATUSOV et al., 2003).

Resultados

Os cinco clones de BAC selecionados foram seqüenciados, montados e submetidos à anotação automática mediante geração de 20.038 seqüências. A análise preliminar do conteúdo biológico das seqüências geradas através do sistema de anotação automática (RiceGAAS) identificou 113 genes e suas respectivas seqüências promotoras.

Foram produzidas 17.113 seqüências de cDNA de diferentes bibliotecas de ESTs, sendo que após a análise de qualidade e limpeza das seqüências restaram 13.297 seqüências de alta qualidade, ou 78% das seqüências. A análise destas seqüências permitiu a identificação de 5.317 clusters, sendo 1.647 contigs e 3.570

singlets, os quais foram denominados *Musa acuminata* Assembled EST Sequences (MaAES). Dos 5.317 MaAES, 1.975 (37.2%) não apresentaram similaridade com as seqüências depositadas nos bancos de dados públicos. Todas as seqüências consenso dos 5.317 MaAES foram depositadas no DATA*Musa.*

Para a identificação dos Análogos de Genes de Resistência (RGAs) um total de 1.593 seqüências foram geradas e analisadas. De um total de 1.593 seqüências originárias de amplificações com nove combinações diferentes de primers, 335 mostraram similaridade com genes de resistência (R-genes) ou RGAs, através de análise com os programas Blastx e estwisedb.

A montagem destas seqüências resultou em 62 contigs de RGAs de *Musa* e sete singletons foram identificados. Os contigs foram considerados completos após a remoção dos primers anti-senso e senso e análise da qualidade. O tamanho médio de um contig de RGA foi de 636 bp, dos quais 335 bp foram analisados filogeneticamente.

Dos contigs considerados completos após a remoção dos primers anti-senso e senso e a análise da qualidade, os que mostraram mudanças de janela de leitura foram excluídos da análise filogenética. Devido o fato que os RGAs foram gerados com várias combinações de primers direcionados para diversos motivos, uma seqüência comum de 335 pb para todos os contigs entre os motivos de kinase 2 e GLPL foi selecionada para a análise filogenética.

Uma análise filogenética preliminar de 28 *Musa* RGA contigs com contíguos ORFs foi conduzido com 50 *Arabidopsis thaliana* NB-ARC RGAs e R-genes, usando o programa ClustalW e a geração de uma árvore de média distância com o programa Jalview. Os RGAs formaram dois grupos, configurando as famílias NBS TIR e não-TIR. Todos os genes de *Musa acuminata* agruparam-se na classe não-TIR, como esperado para os RGAs de monocotiledôneas. Dentro desta classe, três subclasses foram observadas, sendo duas classes de *Arabidopsis thaliana* e um grupo *Musa acuminata*. As distâncias dentro da classe de *Musa acuminata* correlacionam-se com as combinações dos diferentes primers.

Considerações finais

O banco de dados de genômica de banana que resultou do projeto "Análise da Estrutura Primária do Genoma A de *Musa acuminata*", o DATA*Musa*, é composto por informações de Genômica Estrutural (BAC´s), de Transcriptoma (ESTs) e de Análogos de Genes de Resistência (RGAs). A análise das quase 40.000 seqüências de DNA presentes no DATA*Musa* permitiu a identificação de aproximadamente 5.500 unidades gênicas de banana, um aumento de quase 20 vezes o número de genes de banana disponíveis nos bancos de dados públicos como o GenBank – NCBI por ocasião do início do projeto. O DATA*Musa* é um banco de dados em contínuo enriquecimento, sendo que seqüências de DNA do genoma B de *Musa balbisiana* estão sendo produzidas e serão adicionadas ao banco de dados no futuro. Utilizando as informações já contidas no DATA*Musa*, o grupo de pesquisa em genômica e biotecnologia de banana da Embrapa Recursos Genéticos e Biotecnologia já desenvolve atividades de identificação e validação de seqüências promotoras de expressão gênica, uso de micro-arranjo de DNA para estudo da interação banana x *Mycosphaerella fijiensis*, desenvolvimento e mapeamento de marcadores moleculares do tipo SSR, e identificação, caracterização e validação de genes candidatos de interesse para a agroindústria da bananicultura. Todas as seqüências geradas estão disponíveis para acesso pelos interessados, mediante assinatura de acordo de confidencialidade e de transferência de material, no endereço http://genoma.embrapa.br/musa.

Referências Bibliográficas

ALTSCHUL, S. F.; MADDEN, T. L.; SCHAFFER, A. A.; ZHANG, J.; ZHANG, Z.; MILLER, W.; LIPMAN, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, Oxford, GB, v. 25, n. 17, p. 3389-3402, 1997.

BENSON, D. A.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; WHEELER, D. L. GenBank. **Nucleic Acids Research**, Oxford, GB, v. 30, n. 1, p. 17-20, 2002.

BERTIOLI, D. J.; LEAL-BERTIOLI, S. C. M.; LION, M. B.; SANTOS, V. L.; PAPPAS JUNIOR, G.; CANNON, S. B.; GUIMARÃES, P. M. A large scale analysis of resistance gene homologues in *Arachis.* **Mgg Molecular Genetics And Genomics**, Berlin, v. 270, p. 34-45, 2003.

CHANG, S.; PURYEAR, J.; CAIRNEY, J. A. Simple and efficient method for isolating RNA from Pine trees. **Plant Molecular Biology Reporter**, Athens, GR, v. 11, n. 2, p. 115-116, 1993.

EWING, B.; HILLIER, L.; WENDL, M. C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, Cold Spring Harbor, US, v. 8, n. 3, p. 175-85, 1998.

FAO. **WAICENTPORTAL**: statistics. Disponível em: <http://www.fao.org/waicent/portal/statistics_en.asp>. Acesso em: abr. 2004.

GASTEIGER, E.; JUNG, E.; BAIROCH, A. SWISS-PROT: connecting biological knowledge via a protein database. **Curr Issues Mol Biol**., v. 3, n. 3, p. 47-55, 2001.

HUANG, X.; MADAN, A. CAP3: a DNA sequence assembly program. **Genome Research**, Cold Spring Harbor, US, v. 9, n. 9, p. 868-877, 1999.

KANAZIN, V.; MAREK, L. F.; SHOEMAKER, R. C. Resistance gene analogs are conserved and clustered in soybean. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, US, v. 93, v. 11746-11750, 1996.

PEÑUELA, S.; DANESH, D.; YOUNG, N. D. Targeted isolation, sequence analysis, and physical mapping of nonTIR NBS-LRR genes in soybean. **Theoretical and Applied Genetics**, Berlin, v. 104, n. 2/3, p. 261-272, 2002.

ROGERS, S. O.; BENDICH, A. J. Extraction of DNA from plant tissues. In: GELVIN, S.; SCHILPEROOT, R. A. (Ed). **Plant molecular biology manual**. Boston: Kluwer, 1988. p. A6: 1-10.

SAMBROOK, J.; RUSSELL, D. W. **Molecular cloning**: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2001.

SANTOS, C. M. R.; MARTINS, N. F.; HORBERG, H. M.; ALMEIDA, E. R. de; COELHO, M. C.; TOGAWA, R. C.; SILVA F. R. da; CAETANO, A. R.; MILLER, R. N.; SOUZA JUNIOR, M. T. Analysis of expressed sequence tags from *Musa acuminata* ssp. *burmannicoides,* var. Calcutta 4(AA) leaves submitted to temperature stresses. **Theoretical and Applied Genetics**, Berlin, v. 110, p. 1517–1522, 2005.

SCHOOF, H.; ZACCARIA, P.; GUNDLACH, H.; LEMCKE, K.; RUDD, S.; KOLESOV, G.; ARNOLD, R.; MEWES, H. W.; MAYER, K. F. MIPS Arabidopsis thaliana Database (MAtDB): an integrated biological knowledge resource based on

the first complete plant genome. **Nucleic Acids Research**, Oxford, GB, v. 30, n. 1, p. 91-93, 2002.

TATUSOV, R. L.; FEDOROVA, N. D.; JACKSON, J. D.; JACOBS, A. R.; KIRYUTIN, B.; KOONIN, E. V.; KRYLOV, D. M.; MAZUMDER, R.; MEKHEDOV, S. L.; NIKOLSKAYA, A. N.; RAO, B. S.; SMIRNOV, S.; SVERDLOV, A. V.; VASUDEVAN, S.; WOLF, Y. I.; YIN, J. J.; NATALE, D. A. The COG database: an updated version includes eukaryotes. **Bioinformatics**, v. 4, n. 1-14, 2003.

TELLES, G. P.; SILVA, F. L. da. Trimming and clustering sugarcane ESTs. **Genetics Molecular Biology**, Ribeirao Preto, v. 24, p. 17-23, 2001.

VILARINHOS, A. D.; PIFFANELLI, P.; LAGODA, P.; THIBIVILLIERS, S.; SABAU, X.; CARREEL, F.; D'HONT, A. Construction and characterization of a bacterial artificial chromosome library of banana (Musa acuminata Colla). **Theoretical and Applied Genetics**, Berlin, v. 106, n. 6, p. 1102-1106, 2003.