

**Genômica Estrutural de *Musa acuminata* ssp. *burmannicoides* var.
Calcutta 4**

República Federativa do Brasil

Luiz Inácio Lula da Silva
Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Roberto Rodrigues
Ministro

Empresa Brasileira de Pesquisa Agropecuária

Conselho de Administração

Luis Carlos Guedes Pinto
Presidente

Silvio Crestana
Vice-Presidente

Alexandre Kalil Pires
Ernesto Paterniani
Helio Tollini
Marcelo Barbosa Saintive
Membros

Diretoria-Executiva da Embrapa

Silvio Crestana
Diretor Presidente

José Geraldo Eugênio de França
Kepler Euclides Filho
Tatiana Deane de Abreu Sá
Diretores Executivos

Embrapa Recursos Genéticos e Biotecnologia

José Manuel Cabral de Sousa Dias
Chefe-Geral

Maurício Antônio Lopes
Chefe-Adjunto de Pesquisa e Desenvolvimento

Maria Isabel de Oliveira Penteado
Chefe-Adjunto de Comunicação e Negócios

Maria do Rosário de Moraes
Chefe-Adjunto de Administração

Boletim de Pesquisa e Desenvolvimento 108

**Genômica Estrutural de *Musa acuminata* ssp.
burmannicoides var. Calcutta 4**

Manoel Teixeira Souza Júnior

Candice Mello Romero Santos

Felipe Rodrigues da Silva

Roberto Coti Togawa

Robert Neil Gerald Miller

Ana Yamaguishi Ciampi

Pietro Piffanelli

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

Comitê de Publicações

Presidente: *Maria Isabel de Oliveira Penteado*

Secretário-Executivo: *Maria da Graça Simões Pires Negrão*

Membros: *Arthur da Silva Mariante*

Maria Alice Bianchi

Maria de Fátima Batista

Maurício Machain Franco

Regina Maria Dechechi Carneiro

Sueli Correa Marques de Mello

Vera Tavares de Campos Carneiro

Supervisor editorial: *Maria da Graça S. P. Negrão*

Normalização Bibliográfica: *Maria Iara Pereira Machado*

Editoração eletrônica: *Maria da Graça S. P. Negrão*

1ª edição

1ª impressão (2005):

G 335 Genômica estrutural de *Musa acuminata* ssp. *burmannicoides* var.

Calcutta 4 / Manoel Teixeira Souza Júnior ... [et al.]. – Brasília:

Embrapa Recursos Genéticos e Biotecnologia, 2005.

22 p. – (Boletim de Pesquisa e Desenvolvimento / Embrapa

Recursos Genéticos e Biotecnologia, 1676 – 1340; 108)

1. *Musa acuminata* - Calcutta 4 - genômica estrutural. 2. Análise da Estrutura Primária do Genoma A de *Musa acuminata* - projeto de pesquisa. 3. DATAMusa - banco de dados - genômica de banana - genômica estrutural - transcriptoma - análogos de genes de resistência. I. Souza Júnior, Manoel Teixeira. II. Série.

634.772 – CDD 21.

SUMÁRIO

Resumo	6
Abstract.....	7
Introdução.....	8
Material e Métodos	9
1. Seleção dos clones de BAC e produção das bibliotecas shotgun ..	9
2. Sequenciamento de DNA.....	10
3. Avaliação da qualidade.....	11
4. Agrupamento das seqüências e inspeção da montagem.....	11
5. Anotação das seqüências	11
Resultados	12
Caracterização do clone de BAC MA4_BAC008L021	12
Caracterização do clone de BAC MA4_BAC042M013	13
Caracterização do clone de BAC MA4_BAC078I012	14
Caracterização do clone de BAC MA4_BAC106O017	15
Caracterização do clone de BAC MA4_BAC111B014	16
Considerações Finais	17
Referências Bibliográficas	21

Manoel Teixeira Souza Júnior

Candice Mello Romero Santos

Felipe Rodrigues da Silva

Roberto Coti Togawa

Robert Neil Gerald Miller

Ana Yamaguishi Ciampi

Pietro Piffanelli

Resumo

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 Universidade Católica de Brasília (UCB) e o Centro Francês de Pesquisa Agrícola para o Desenvolvimento Internacional (CIRAD), no período de fevereiro de 2002 a junho de 2005, resultou na criação do DATAMusa. O DATAMusa é um banco de dados de genômica de banana composto de informações de genômica estrutural, de transcriptoma e de análogos de genes de resistência. A parte referente a genômica estrutural é resultado da produção e caracterização de cinco clones de BAC, selecionados utilizando-se sonda do projeto EGRAM. Os clones de BAC, totalizando 525 kb, foram seqüenciados mediante a 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 de expressão.

Abstract

The research project entitled "Analysis of the Primary Structure of the A Genome of *Musa acuminata*", financed by the National Council for Scientific and Technological Development (CNPq), and executed by Embrapa Genetic Resources and Biotechnology, in partnership with the University Catholic of Brasilia (UCB) and the French Center of Agricultural Research for International Development (CIRAD), in the period from February 2002 to June 2005, resulted in the launching of the DATAMusa. The DATAMusa is a banana genomics database that contains information on structural genomic, transcriptome and resistance genes analogs. The structural data in the DATAMusa is resulted of the selection (using probes from the EGRAM project) and characterization of five clones from the *Musa acuminata* ssp. *burmannicoides* - variety Calcutta 4 BAC library. Together these BAC clones account for 525 kb, and were sequenced through the generation of 20,038 sequences. The preliminary analysis of the data, using the automatic annotation pipeline RiceGAAS, let to the identification of 113 genes and their respective promoter sequences.

Introdução

A banana (*Musa* spp.) é uma espécie cultivada em diversos países tropicais e possui um importante papel social e econômico. 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).

A bananeira é cultivada de Norte a Sul do País, sendo fundamental para a complementação da dieta alimentar das populações de baixa renda. Praticamente toda fruta produzida é comercializada no mercado interno. A maioria dos bananicultores é composta por pequenos produtores, e o setor da bananicultura no Brasil gera mais de 500 mil empregos diretos. A banana é considerada uma rica fonte de energia, minerais e vitaminas.

O Programa Internacional para o Melhoramento de *Musa* (PROMUSA), ligado à Rede Internacional para o Melhoramento de Banana e Plátano (INTERNATIONAL..., 2005), 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 *Musa* (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. Este consórcio é uma iniciativa que objetiva aplicar genômica para o melhoramento sustentável da banana, como também gerar recursos livremente acessíveis para a genômica de *Musa*, e usar os novos conhecimentos e ferramentas em apoio ao melhoramento convencional e

biotecnológico. A estratégia do GMGC também busca permitir um melhor uso e manutenção da biodiversidade deste gênero, e garantir a segurança alimentar e financeira de milhões de pessoas em países subdesenvolvidos e em desenvolvimento.

Este boletim de pesquisa relata as atividades de pesquisa em genoma estrutural de banana coordenadas e executadas pela Embrapa Recursos Genéticos e Biotecnologia dentro do projeto de pesquisa “Análise da Estrutura Primária do Genoma A de *Musa acuminata*”, em parceria com a Universidade Católica de Brasília (UCB) e o Centro Francês de Pesquisa Agrícola para o Desenvolvimento Internacional (CIRAD), no período de fevereiro de 2002 a junho de 2005, e que fazem parte do DATAMusa.

Material e Métodos

1. Seleção dos clones de BAC e produção das bibliotecas shotgun

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 (Tabela 1) foram preliminarmente caracterizados mediante geração de perfil de restrição com diferentes enzimas de restrição e sequenciamento das extremidades. As bibliotecas de subclones clone 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. Clones da biblioteca de BAC de *Musa acuminata* ssp. *burmannicoides* - variedade Calcutta 4 – selecionados e caracterizados.

Clone de BAC	Tamanho Estimado (Kb)	Marcador EGRAM
MA4_BAC106O017	120	SB851
MA4_BAC078I012	130	SB661
MA4_BAC008L021	110	SB373
MA4_BAC042M013	35	SB748
MA4_BAC111B014	130	SB854
TOTAL	525	--

2. Sequenciamento de DNA

As extremidades 5' e 3' dos subclones foram seqüenciadas na plataforma de sequenciamento de DNA da Embrapa Recursos Genéticos e Biotecnologia (<http://www/laboratorios/psd/psd.html>), utilizando-se os oligonucleotídeos iniciadores universais “M13 forward” (5' - TGT AAA ACG ACG GCC AGT - 3') e “M13 reverse” (3' - CAG GAA ACA GCT ATG ACC - 5'). Para o sequenciamento de DNA foi utilizado o BigDye Terminator Cycle Sequencing Kit (Applied Biosystems), e as reações de amplificação foram realizadas em termocicladores 9700 (Applied Biosystems). As reações de sequenciamento

foram lidas em seqüenciadores automáticos 3700 (Applied Biosystems). Os eletroferogramas foram submetidos ao Sistema GENOMA da Embrapa Recursos Genéticos e Biotecnologia (<http://genoma.cenargen.embrapa.br/genoma/>) e estocados no *MUSA_BACs* database do Laboratório de Bioinformática até processamento e análise de seqüências.

3. Avaliação da qualidade

Para avaliar a qualidade do sequenciamento os eletroferogramas gerados no sequenciamento dos subclones de BAC 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).

4. Agrupamento das seqüências e inspeção da montagem:

As seqüências foram submetidas, independente de sua qualidade, à montagem utilizando o programa CAP3 (HUANG e MADAN, 1999) e Phrap (GREEN, 1999), e estas inspecionadas utilizando-se o Consed e o Phrapview (GORDON et al., 1998).

5. Anotação das seqüências

Para realizar a anotação utilizou-se o sistema de anotação automática do arroz-RiceGAAS (SAKATA et al., 2002), o qual é composto por programas de predição e análises de códigos de proteínas e estrutura de genes.

Resultados

Caracterização do clone de BAC MA4_BAC008L021

O clone da biblioteca de BAC de *Musa acuminata* ssp. *burmannicoides* var. Calcutta 4 (VILARINHOS et al., 2003), denominado MA4_BAC008L021, foi selecionado com a sonda SB373 (Tabela 1). A seqüência desta sonda apresentou homologia em Blastn com a seqüência “*Zea mays* PCO148957 mRNA sequence AY107008 hypothetical protein”, e apresentou uma única cópia tanto no genoma de banana quanto no de arroz. A cobertura desta biblioteca de subclones variou de 140 a 280 vezes o tamanho estimado do clone MA4_BAC008L021, que era de 110 kb.

Os cromatogramas (2.511) resultantes do sequenciamento dos clones da biblioteca de shotgun do clone de BAC MA4_BAC008L021 foram submetidos à montagem. A inspeção da montagem mostrou que cinco contigs foram obtidos, sendo que estes contigs eram resultado da junção de 2.163 seqüências (ou 86,35% das seqüências utilizadas). O maior destes contigs, denominado contig 5, com 113.508 bases, resultou da junção de 2.155 seqüências, o que deixou oito seqüências para montar os outros três contigs.

Estes resultados, aliados à comparação do tamanho observado dos fragmentos obtidos na digestão do DNA deste clone de BAC com diversas enzimas de restrição com a “digestão eletrônica” da seqüência montada (teste de co-linearidade), assim como a estimativa da taxa de erro inferior a um por milhão de bases, nos permite declarar esta seqüência finalizada. O contig 5 apresentou uma cobertura de clones uniforme do seu início ao fim, com somente dois pequenos gaps na região 30.000-35.000 e na região 53.000-56.000, os quais não comprometeram a montagem.

A análise preliminar do conteúdo biológico das seqüências geradas através do sistema de anotação automática identificou de 28 regiões codificadoras de proteínas, sendo 15 na fita sense e 13 na antisense. Destas 28 regiões identificadas, oito são de proteínas hipotéticas, não possuindo homólogos ainda descritos no GenBank (Tabela 2).

Caracterização do clone de BAC MA4_BAC042M013

O clone da biblioteca de BAC de *Musa acuminata* da var. Calcutta 4 (VILARINHOS et al., 2003) denominado MA4_BAC042M013, foi selecionado com a sonda SB 748 (Tabela 1). Esta sonda, cuja seqüência apresentou homologia em Blast (n) com a seqüência *Zea mays* mRNA for Putative Porphobilinogen deaminase Y12809, apresentou uma única cópia no genoma de banana e duas no de arroz. A cobertura desta biblioteca de subclones variou de 439 a 878 o tamanho estimado do clone MA4_BAC042M013, que era de 35 kb.

Os cromatogramas (2.491) resultantes do sequenciamento dos clones da biblioteca de shotgun do clone de BAC MA4_BAC042M013 foram submetidos à montagem. A inspeção da montagem mostrou que 88 contigs foram obtidos, sendo que estes contigs eram resultado da junção de 1.440 seqüências (ou 57,81% das seqüências utilizadas). O maior destes contigs, denominado de Contig 88, com 29.568 bases, resultou da junção de 1.229 seqüências, restando 211 seqüências para montar os demais contigs.

Estes resultados, aliados à comparação do tamanho observado dos fragmentos obtidos na digestão do DNA deste clone de BAC com diversas enzimas de restrição com a “digestão eletrônica” da seqüência montada (teste

de co-linearidade), assim como a estimativa da taxa de erro inferior a um por milhão de bases, nos permite declarar esta seqüência finalizada. Porém, neste caso, o clone de BAC apresentou um tamanho inferior ao predito inicialmente, isto é, 29.568 bp e não 35.000 bp.

A análise preliminar do conteúdo biológico da seqüência gerada através do sistema de anotação automática identificou cinco regiões codificadoras de proteínas, sendo três na fita sense e duas na fita antisense (Tabela 3).

Caracterização do clone de BAC MA4_BAC078I012

O clone da biblioteca de BAC de *Musa acuminata* da var. Calcutta 4 (VILARINHOS et al., 2003), denominado MA4_BAC078I012, foi selecionado com a sonda SB 661 (Tabela 1). Esta sonda, cuja seqüência apresentou homologia em Blast (n) com a seqüência *Zea mays* PCO104916 mRNA sequence AY105359 putative thioredoxin-like U5 small ribonucleoprotein particle *Oryza sativa*, apresentou duas cópias no genoma de arroz e de 2 a 3 cópias no genoma de banana. A cobertura desta biblioteca de subclones variou de 118 a 236 vezes o tamanho estimado do clone MA4_BAC078I012, que era de 130 kb.

Os cromatogramas (7.104) resultantes do sequenciamento dos clones da biblioteca de shotgun do clone de BAC MA4_BAC078I012 submetidos à montagem. A inspeção da montagem mostrou que 66 contigs foram obtidos, sendo que estes contigs eram resultado da junção de 4.608 seqüências (ou 65.74% das seqüências utilizadas). O maior destes contigs, denominado Contig 66, com 88.441 bases, resultou da junção de 2.620 seqüências, restando 4.389 seqüências para montar os demais contigs.

A análise preliminar do conteúdo biológico da seqüência gerada através do sistema de anotação automática identificou 22 regiões codificadoras de proteínas, sendo 14 na fita sense e oito na fita antisense. Destas, cinco regiões codificadoras identificadas deste BAC ainda não possui homólogos descritos no GenBank (Tabela 4).

Caracterização do clone de BAC MA4_BAC106O017

O clone da biblioteca de BAC de *Musa acuminata* da var. Calcutta 4 (VILARINHOS et al., 2003), denominado MA4_BAC106O017, foi selecionado com a sonda SB851 (Tabela 1). Esta sonda, cuja seqüência apresentou homologia em Blast (n) com a seqüência “*Zea mays* PCO105631 mRNA sequence/*T.aestivum* gene for phosphoglycerate kinase X73528”, apresentou duas cópias no genoma de arroz e de duas a três no genoma de banana. A cobertura desta biblioteca de subclones variou de 128 a 256 vezes o tamanho estimado do clone MA4_BAC106O017, que era de 120kb.

Os cromatogramas (2.385) resultantes do sequenciamento dos clones da biblioteca de shotgun do clone de BAC MA4_BAC106O017 foram submetidos à montagem. A inspeção da montagem mostrou que 13 contigs foram obtidos, sendo que estes contigs foram resultado da junção de 2.032 seqüências (ou 85,2% das seqüências utilizadas). Os maiores contigs foram denominados 13 e 12, com 116.333 e 25.357 bases, respectivamente. O Contig 13 resultou da junção de 1.628 seqüências, enquanto que o Contig 12 resultou da junção de 380 seqüências. Desde modo, restaram 377 seqüências para montar os demais contigs.

A análise preliminar do conteúdo biológico da seqüência gerada através do sistema de anotação automática identificou seis regiões codificadoras de proteínas, sendo três na fita sense e três na fita antisense, no Contig 12. No entanto, para o Contig 13 foram identificadas 32 regiões codificadoras, sendo 18 na fita sense e 14 na fita antisense, destas 32 regiões codificadoras identificadas neste BAC, 10 ainda não possui homólogos descritos no GenBank (Tabela 5).

Caracterização do clone de BAC MA4_BAC111B014

O clone da biblioteca de BAC de *Musa acuminata* da var. Calcutta 4 (VILARINHOS et al., 2003), denominado MA4_BAC111B014, foi selecionado com a sonda SB854 (Tabela 1). Esta sonda, cuja seqüência apresentou homologia em Blast (n) com a seqüência *Zea mays* mitochondrial Rieske Fe-S protein mRNA - M77224, apresentou uma duas cópias no genoma de arroz e cinco no genoma de banana (Tabela 1). A cobertura desta biblioteca de subclones variou de 118 a 136 vezes o tamanho estimado do clone MA4_BAC111B014, que era de 130 kb.

Os cromatogramas (5.648) resultantes do sequenciamento dos clones da biblioteca de shotgun do clone de BAC MA4_BAC111B014 foram submetidos à montagem. A inspeção da montagem utilizando mostr que 31 contigs foram obtidos, sendo que estes contigs foram resultado da união de 3.842 seqüências (ou 68% das seqüências utilizadas). O maior destes contigs, denominado Contig 31, com 75.660 bases, resultou da junção de 1.922 seqüências, restando 3.726 seqüências para montar os demais contigs.

A análise preliminar do conteúdo biológico da seqüência gerada através do sistema de anotação automática identificou 20 regiões codificadoras de proteínas, sendo 12 na fita sense e oito na fita antisense (Tabela 6).

Considerações Finais

Foram selecionados cinco clones de BAC, que juntos representam 525 kb (ou aproximadamente 0,1%) do genoma A de banana, que é estimado em 500-600 Mb. No total foram geradas 20.038 seqüências de DNA a partir das cinco bibliotecas de subclones de clones de BAC. Destas, um total de 5.984.726 bases foram aceitas e submetidas à montagem dos clones de BAC, visando obter as seqüências consensos que foram submetidas à anotação automática. A anotação automática identificou 113 genes distribuídos nos cinco clones de BAC. Juntamente com esses, foram identificadas 113 regiões potencialmente promotoras de expressão gênica. Todos os cromatogramas gerados e os resultados da montagem e anotação dos cinco clones de BAC fazem parte do DATA_ *Musa* (banco de dados de seqüências de DNA e análises de seqüências).

Tabela 2. Lista de genes preditos no subclone de BAC MA4_008L021 através da anotação automática RiceGaas.

#	Função dos genes predita
1	hypothetical protein
2	putative Guanidinoacetate Methyltransferase With Gd Ion
3	putative subtilisin-like serine proteinase (EC 3.4.21.-)
4	putative COG4531: ABC-type Zn ²⁺ transport system, periplasmic component/surface adhesin
5	putative Myosin heavy chain, cardiac muscle isoform
6	putative expressed protein, having alternate splicing products
7	hypothetical protein
8	putative COG0463: Glycosyltransferases involved in cell wall biogenesis
9	putative Crystal Structure Of 5,8-Di-Amino-1,4-Di-Hydroxy-AnthraquinoneCK2 KINASE COMPLEX
10	hypothetical protein
11	putative Crystal Structure Of 5,8-Di-Amino-1,4-Di-Hydroxy-AnthraquinoneCK2 KINASE COMPLEX
12	putative INDEHISCENT protein
13	putative more than blood homolog
14	putative Two-component response regulator ARR18
15	hypothetical protein
16	hypothetical protein
17	putative ferrichrome receptor
18	putative prolamin box binding factor
19	hypothetical protein
20	putative low molecular weight neurofilament protein XNF-L
21	putative Structure Of A Beta-Trcp1-Skp1-Beta-Catenin Complex: Destruction Motif Binding And Lysine Specificity On The Scfbeta-Trcp1 Ubiquitin Ligase
22	putative Prolyl 4-hydroxylase alpha subunit (4-PH alpha) (Procollagen-proline,2-oxoglutarate-4-dioxygenase alpha subunit)
23	putative protein disulfide-isomerase (EC 5.3.4.1) precursor
24	hypothetical protein
25	putative cold acclimation protein WCOR413
26	putative Ring-H2 Finger Domain Of EI5
27	hypothetical protein
28	putative nirC protein

Tabela 3. Lista de genes preditos no subclone de BAC MA4_042M013 através da anotação automática RiceGaas.

#	Função dos genes predita
1	putative P54 protein precursor
2	unknown protein similar to ENSANGP00000015645
3	putative Porphobilinogen deaminase (PBG) (Hydroxymethylbilane synthase) (HMBS) (Pre-uroporphyrinogen synthase)
4	putative Porphobilinogen deaminase (PBG) (Hydroxymethylbilane synthase) (HMBS) (Pre-uroporphyrinogen synthase)
5	putative hsp 90-binding protein p59

Tabela 4. Lista de genes preditos no subclone de BAC MA4_078I012 através da anotação automática RiceGaas.

#	Função dos genes predita
1	putative Alpha-glucosidase precursor (Maltase)
2	putative HIV-1 retropepsin (EC 3.4.23.16)
3	putative gag-pol polyprotein
4	putative pol polyprotein homolog
5	putative pol polyprotein
6	putative polyprotein
7	putative gag protein homolog
8	putative pol protein
9	hypothetical protein
10	hypothetical protein
11	unknown protein
12	putative ATP-binding cassette transporter Atr5
13	putative gag-pol polyprotein
14	putative Uncharacterized protein family UPF0021
15	hypothetical protein
16	putative gag-protease polyprotein
17	putative ALR protein homolog
18	hypothetical protein
19	putative pol polyprotein
20	putative COG1960: Acyl-CoA dehydrogenases
21	putative gag protein homolog
22	putative groovin gene protein

Tabela 5. Lista de genes preditos no subclone de BAC MA4_106O017 - contigs 12 e 13 através da anotação automática RiceGaas.

#	Função dos genes predita – Contig 12
1	putative Phosphoglycerate kinase, chloroplast precursor
2	putative Taurine--pyruvate aminotransferase
3	putative shock protein SRC2
4	putative beta-galactosidase
5	putative glutaredoxin
6	putative YIBUD6
#	Função dos genes predita – Contig 13
1	hypothetical protein
2	putative Bromodomain adjacent to zinc finger domain 2A (Transcription termination factor-I interacting protein 5) (TTF-I interacting protein 5) (Tip5) (hWALp3)
3	putative Aminomethyltransferase, mitochondrial precursor (Glycine cleavage system T protein) (GCVT)
4	putative Amine oxidase flavin containing domain protein 2 (AOF2 protein) (BRAFF-HDAC complex protein BHC110)
5	putative galactosyltransferase family protein
6	putative Nonspecific lipid-transfer protein A (NS-LTP A) (Phospholipid transfer protein) (PLTP)
7	putative conjugal transfer protein Trb1
8	hypothetical protein
9	hypothetical protein
10	putative subtilisin-like proteinase homolog
11	putative nucleotide exchange factor RasGEF K
12	hypothetical protein
13	putative Auxin response factor 6
14	putative bZIP protein
15	hypothetical protein
16	putative BTB/POZ domain containing protein 4 (Zinc finger protein 340)
17	putative RING finger protein 12 (LIM domain interacting RING finger protein) (RING finger LIM domain-binding protein) (R-LIM) (NY-REN-43 antigen)
18	hypothetical protein
19	hypothetical protein
20	putative protein M7.2
21	putative COG1391: Glutamine synthetase adenylyltransferase
22	hypothetical protein
23	putative PRUNEM1
24	putative Structure Of The Conserved Domain Of Anac, A Member Of The Nac Family Of Transcription Factors
25	unknown protein
26	putative DnaJ homolog subfamily B member 12
27	hypothetical protein similar to WD-repeat protein RBAP1
28	hypothetical protein
29	putative salicylic acid-induced fragment 1 protein
30	putative proteophosphoglycan, membrane-associated
31	putative COG0477: Permeases of the major facilitator superfamily
32	putative glutaredoxin

Tabela 6. Lista de genes preditos no subclone de BAC MA4_111B014 através da anotação automática RiceGaas.

#	Função dos genes predita
1	putative tRNA-dihydrouridine synthase A
2	putative glucuronyltransferase-P, long form
3	putative 1-aminocyclopropane-1-carboxylate oxidase 2 (ACC oxidase 2) (Ethylene-forming enzyme) (EFE)
4	putative novel S-100VICaBP type calcium binding domain, EF hand domain and Filaggrin domain containing protein
5	hypothetical protein
6	hypothetical protein
7	unknown protein
8	putative PB1 polymerase subunit
9	unknown protein
10	putative proteophosphoglycan, membrane-associated
11	putative Nmr Solution Structure Of The U Box Domain From Atpub14, An Armadillo Repeat Containing Protein From Arabidopsis Thaliana
12	putative retinitis pigmentosa GTPase regulator
13	putative Crystal Structure Of Archaelhodopsin-1
14	putative Phytosulfokine receptor precursor (Phytosulfokine LRR receptor kinase)
15	putative Ubiquinol-cytochrome c reductase iron-sulfur subunit, mitochondrial precursor (Rieske iron-sulfur protein) (RISP)
16	hypothetical protein
17	putative Mutator-like transposase
18	putative dehydrogenase subunit alpha
19	putative COG2202: FOG: PAS/PAC domain
20	putative RTX protein

Referências Bibliográficas

EWING, B.; HILLIER, L.; WENDL, M. C.; GREEN, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. **Genome Research**, v. 8, n. 3, p. 175-185, 1998.

FAO. **FAOstat 2004**. Disponível em: < <http://faostat.fao.org/default.aspx> >. Acesso em: 2005.

GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Research**, v. 8, p. 195-202, 1998.

GREEN, P. PHRAP. 1999. Disponível em: < <http://bozeman.genome.washington.edu/phrap/docs/phrap.html> >. Acesso em: 2005.

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

INTERNATIONAL NETWORK FOR THE IMPROVEMENT OF BANANA AND PLANTAIN. Disponível em: < www.inibap.org >. Acesso em: 2005.

SAKATA, K.; NAGAMURA, Y.; NUMA, H.; ANTONIO, B. A.; NAGASAKI, H.; IDONUMA, A.; WATANABE, W.; SHIMIZU, Y.; HORIUCHI, I.; MATSUMOTO, T.; SASAKI, T.; HIGO, K. "RiceGAAS: an automated annotation system and database for rice genome sequence". **Nucleic Acids Research**, Oxford, v. 30, p. 98-102, 2002.

TELLES, G. P.; SILVA, F. L. da. Trimming and clustering sugarcane ESTs. **Genetic Molecular Biology**, Ribeirão Preto, Brazil, 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.