

TRANSCRIPTOMA DE *Musa acuminata* NO DATAMusa

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

Diretores Executivos

José Geraldo Eugênio de França

Kepler Euclides Filho

Tatiana Deane de Abreu Sá

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 109

TRANSCRIPTOMA DE *Musa acuminata* NO DATAMusa

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

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) 3448-4600 Fax:

(61) 3340-3624

<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)

T 772 Transcriptoma de *Musa acuminata* no DATAMusa / Manoel Teixeira Souza Júnior ... [et al.]. – Brasília: Embrapa Recursos Genéticos e Biotecnologia, 2005.

21 p. – (Boletim de Pesquisa e Desenvolvimento / Embrapa Recursos Genéticos e Biotecnologia, 1676 – 1340; 109)

1. *Musa acuminata* – Transcriptoma. 2. DATAMusa - banco de dados - genômica de banana. 3. DATAMusa - banco de dados - genômica estrutural. 4. DATAMusa - banco de dados – transcriptoma. 5. DATAMusa - banco de dados – análogos de genes de resistência I. Souza Júnior, Manoel Teixeira. II. Série.

584.39 – CDD 21.

TRANSCRIPTOMA DE *Musa acuminata* NO DATAMusa

Manoel Teixeira Souza Júnior¹
Candice Mello Romero Santos²
Natália Florêncio Martins³
Felipe Rodrigues da Silva⁴
Roberto Coti Togawa⁵
Lisângela Pinheiro Cassiano⁶
Elionor Rita Pereira de Almeida⁷
Marly Catarina Felipe Coelho⁸
Alexandre Rodrigues Caetano⁹
Ana Yamaguishi Ciampi¹⁰
Pietro Piffanelli¹¹
Robert Neil Gerald Miller¹²

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 - Parco Tecnológico Padano / Lodi -Italy

¹² PhD - Universidade Católica de Brasília(UCB)

SUMÁRIO

RESUMO	4
ABSTRACT.....	5
INTRODUÇÃO	6
MATERIAL E MÉTODOS.....	7
RESULTADO	10

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 ao transcriptoma no DATAMusa é resultado da produção e caracterização de sete bibliotecas de cDNA, e teve o intuito de descobrir os genes expressos em folhas submetidas a estresse de temperatura, folhas infectadas com o patógeno *Mycosphaerella fijiensis* em estágio precoce e tardio de infecção, flor masculina, raízes de plantas *in vitro* e casca verde de *Musa acuminata*. No total foram produzidas 17.113 seqüências, sendo que após a análise de qualidade e limpeza das seqüências restaram 13.297 seqüências de alta qualidade. 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 homologia com as seqüências depositadas nos bancos de dados públicos.

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 Catholic University of Brasília (UCB) and the French Center of Agricultural Research for International Development (CIRAD), between February 2002 and 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 transcriptome data in the *DATAMusa* 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 total of 17,113 sequences were produced out of the seven cDNA libraries. After trimming the sequences, 13,297 sequences of high quality were kept for further analysis. The assembling of these sequences allowed the identification of 5,317 clusters (being 1,647 contigs and 3,570 singlets), which were denominated *Musa acuminata* Assembled EST Sequences (MaAES). Out of the 5,317 MaAES, 1,975 (37.2%) showed no matches with existing sequences in public databases.

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 (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 *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. 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 deverá ser alcançado mediante um maior entendimento da genética e do genoma deste gênero, permitindo elaborar novas estratégias de melhoramento genético e de transgenia direcionada.

O estudo da genômica funcional ou transcriptoma da bananeira permitirá a elucidação detalhada da função de milhares de genes envolvidos nos processos de diferenciação e desenvolvimento e/ou processos envolvidos nas respostas às alterações do ambiente biótico/ou abiótico. Visando entender o perfil de

expressão de diferentes genes, algumas metodologias podem ser utilizadas, tal como o sequenciamento em larga escala de clones de cDNA, gerando os chamados "Expressed Sequence Tags" (ESTs).

Esta metodologia permite identificar os transcritos específicos dentro de uma biblioteca de cDNA, gerando um perfil quantitativo e qualitativo de diferentes tecidos, tipos de células e estágios de desenvolvimento, sendo considerada uma ferramenta interessante para estudos de expressão gênica, e mapeamento genético (BOGUSKI e SCHULER, 1995; BRENDEL et al., 2002; RONNING et al., 2003; RUDD, 2003).

O presente boletim de pesquisa relata as atividades de pesquisa em transcriptoma 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. Bibliotecas de cDNA: 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 1).

Duas bibliotecas de cDNA de folha de *Musa acuminata* ssp. *burmannicoides* var. Calcutta 4 (AA) submetidas a estresse de temperatura (calor e frio) foram produzidas (SANTOS et al., 2005). 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).

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. *burmannioides* 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 Grupo 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.

2. Extração de RNA total e purificação de poli (A) + RNA: 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 masculina foram isolados de acordo com Chang et al. (1993). Para todas bibliotecas, o RNA total foi submetido à purificação de poli (A) + RNA utilizando o kit “Micro FastTrack 2.0 mRNA Isolation” (Invitrogen Life Technologies, USA), seguindo protocolo apresentado pelos fabricantes.

3. Construção das bibliotecas de cDNA: As bibliotecas foram construídas utilizando-se o kit “Creator Smart cDNA library” e o vetor pDNR-LIB (Clontech Laboratories, Inc., USA).

4. Sequenciamento: As extremidades 5' dos clones de cDNA foram seqüenciadas na plataforma de sequenciamento de DNA da Embrapa Recursos Genéticos e Biotecnologia (<http://www/laboratorios/psd/psd.html>) utilizando-se o oligonucleotídeo iniciador “M13 forward” (5` - TGT AAA ACG ACG GCC AGT - 3`) 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 e tamanho mínimo de seqüência de 140 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 e value $< 10^{-5}$, verificou-se a similaridade entre os "MaAES" e as seqüências no banco de dados GenBank nr (BENSON et al., 2002), MIPS *Arabidopsis thaliana* (SCHOOFF et al., 2002) e SwissProt (GASTEIGER et al., 2001), e para prever 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

O total de 17.113 seqüências foram obtidas a partir das sete bibliotecas de cDNA (Tabela 1), e estas seqüências apresentaram tamanho médio de 913 bases. Destas 17.113 seqüências, 3.816 (22,3%) foram descartadas após a análise de qualidade. Estas seqüências foram descartadas por serem ribossomais (42 seqüências ou 0,2%), por derrapagem (666 seqüências ou 3,9%), por tamanho do inserto (2.520 seqüências ou 14,7%), e por qualidade

(588 seqüências ou 3,4%). As seqüências restantes (13.297 ou 77,7%), após a limpeza, apresentaram número médio de bases com Phred acima de 20, por read bom, igual ou superior a 570 nucleotídeos.

As 13.297 seqüências de alta qualidade foram agrupadas utilizando-se a ferramenta CAP3 (HUANG e MADAN, 1999), tendo produzido 5.317 "*Musa acuminata* Assembled EST Sequences" (MaAES). Destes, 1.647 MaAES eram contigs (agrupamentos formados por duas ou mais seqüências) e 3.670 eram singlets (agrupamento formado por uma seqüência) (Figura 1).

A freqüência de distribuição das seqüências após o agrupamento é apresentada na Figura 2. Dos 5.317 MaAES, 14 são compostos por 50 ou mais seqüências, e o mais populoso foi resultado da junção de 502 seqüências. A maioria dos MaAES (95,1% ou 5.057 MaAES) são compostos por 1-5 seqüências, e 69 % destes foram compostos por um transcrito.

As seqüências consenso dos 5.317 MaAES foram inicialmente submetidas ao Blastx (e value $< 10^{-5}$) (ALTSCHUL et al., 1997) contra o banco de dados GenBank *nr* (BENSON et al., 2002). Entre estas, 79 (1,48%) apresentaram homologia com proteínas conhecidas de banana (Tabela 2), e 1.593 (30%) não apresentaram homologia com nenhuma proteína neste banco de dados.

Os MaAES mais abundantes, que continham 413 e 502 transcritos, apresentaram homologia com Anx1 e Ribulose biphosphate carboxylase small chain, respectivamente (Tabela 3).

Para predizer a função das proteínas codificadas pelos 5.317 MaAES, realizou-se o Blastx (e value $< 10^{-5}$) contra o banco de dados KOG (Eukaryotic Orthologous Groups) (TATUSOV et al., 2003), o qual contém grupos de proteínas ortológicas de sete genomas de eucariontes. Esta comparação resultou em 2.451 (46,1%) MaAES com homologia a proteínas com função conhecida e 2.866 (53,9%) sem homologia com as seqüências depositadas neste banco.

Os MaAES que apresentaram homologia com o banco de dados KOG foram categorizados em 22 classes funcionais, dos quais 15,7% correspondem a genes envolvidos com modificação pós-tradução, "turnover" de proteínas e função chaperona; 12,8% com a predição de função geral; 12,1% com genes

envolvidos com a tradução; 7,5% com genes com função não conclusiva; 6,1% com genes envolvidos com a transdução de sinal; e 5,4% com genes com função desconhecida (Figura 3).

Quando comparados com a base de dados MIPs *Arabidopsis thaliana*, 3.479 MaAES (65,4%) apresentaram homologia com seqüências descritas naquele banco; enquanto que 2.425 (45,6%) apresentaram homologia com seqüências protéicas depositadas na base de dados SwissProt.

É importante salientar que 1.549 MaAES (29,1%) não apresentaram homologia com as seqüências protéicas depositadas nos bancos de seqüências testados, utilizando-se Blastx com e value $< 10^{-5}$. Portanto, aproximadamente 1/3 dos MaAES caracterizados nesta biblioteca de cDNA são de genes ainda não descritos em nenhuma das quatro bases de dados utilizadas para comparação.

Todas as seqüências consenso dos 5.317 MaAES foram depositadas no DATAMusa e estão disponíveis para acesso pelos interessados, mediante assinatura de acordo de confidencialidade e de transferência de material, no endereço <http://genoma.embrapa.br/musa>.

Tabela 1. Número total de seqüências de ESTs das bibliotecas de cDNA geradas.

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 <i>in vitro</i>	2.155
Casca	2.420
Folhas de bananeira infectadas com <i>Mycosphaerella fijiensis</i> em estágio inicial de infecção	3.902
Folhas de bananeira infectadas com <i>Mycosphaerella fijiensis</i> em estágio avançado de infecção	3.812
Total	17.113

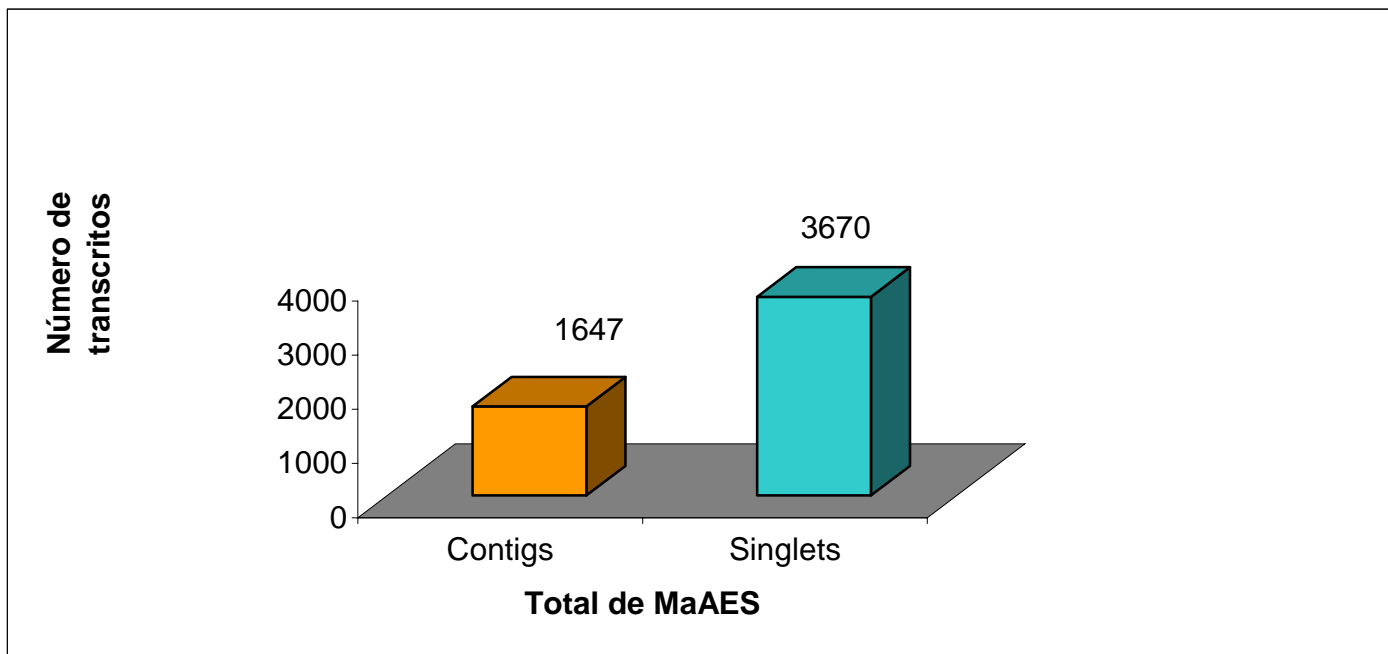


Figura 1. Distribuição dos MaAES após agrupamento com a ferramenta CAP3 das seqüências classificadas como de alta qualidade.

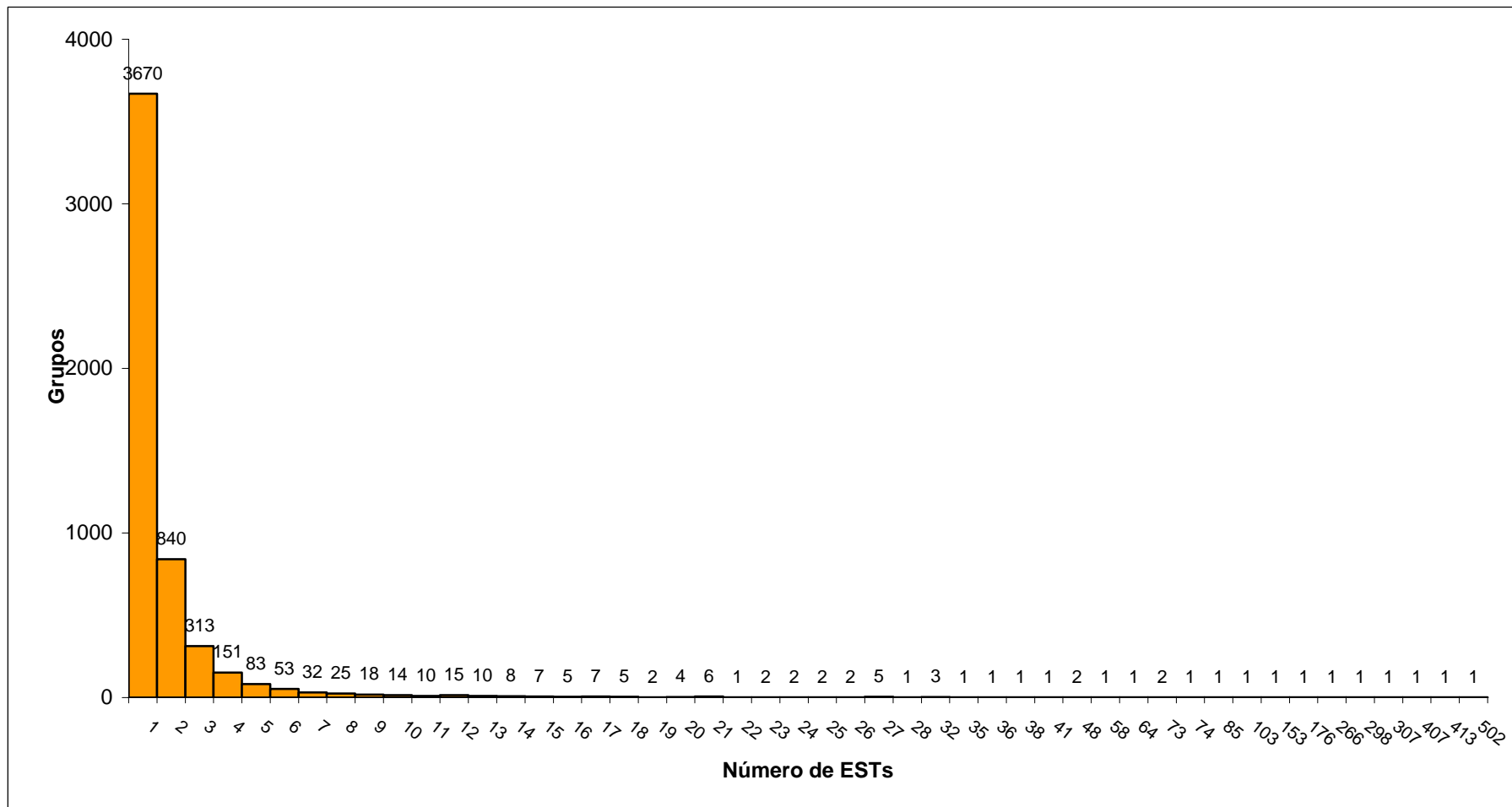


Figura 2. Freqüência de distribuição das seqüências de todas as bibliotecas de cDNA agrupadas mediante uso da ferramenta CAP3.

Tabela 2. Homologia das seqüências consenso dos MaAES com as seqüências de *Musa acuminata* depositadas no banco de dados Genbank *nr*.

Gene	E-value
gb AAL05886.1 germin-like protein [<i>Musa acuminata</i>]	e-100
gb AAM48480.1 lectin [<i>Musa acuminata</i>]	1e-74
gb AAM48480.1 lectin [<i>Musa acuminata</i>]	5e-65
gb AAL05886.1 germin-like protein [<i>Musa acuminata</i>]	e-102
gb AAK54834.1 profilin [<i>Musa acuminata</i>]	1e-70
sp O24045 RBS_MUSAC Ribulose biphosphate carboxylase small chain	5e-79
gb AAT35818.1 abscisic stress ripening protein-like protein [<i>Musa acuminata</i>]	3e-25
gb AAK54834.1 profilin [<i>Musa acuminata</i>]	9e-56
sp O24045 RBS_MUSAC Ribulose biphosphate carboxylase small chain	8e-54
gb AAQ14245.1 actin [<i>Musa acuminata</i>]	e-104
gb AAO15471.1 ethylene receptor-like protein [<i>Musa acuminata</i>]	9e-61
gb AAT67244.1 BTF3b-like transcription factor [<i>Musa acuminata</i>]	2e-67
sp O24045 RBS_MUSAC Ribulose biphosphate carboxylase small chain	3e-72
emb CAC81811.1 putative chitinase [<i>Musa acuminata</i>]	6e-25
gb AAG44759.1 metallothionein-like protein [<i>Musa acuminata</i>]	5e-13
gb AAR88777.1 putative membrane protein type I [<i>Musa acuminata</i>]	1e-15
gb AAG44759.1 metallothionein-like protein [<i>Musa acuminata</i>]	8e-13
sp O24045 RBS_MUSAC Ribulose biphosphate carboxylase small chain	1e-96
gb AAM69298.1 unknown [<i>Musa acuminata</i>]	1e-15
gb AAB82774.1 ripening-associated protein [<i>Musa acuminata</i>]	3e-33
sp O24045 RBS_MUSAC Ribulose biphosphate carboxylase small chain	4e-74
gb AAV70659.1 glyceraldehyde-3-phosphate dehydrogenase [<i>Musa acuminata</i>]	e-174
gb AAG44759.1 metallothionein-like protein [<i>Musa acuminata</i>]	6e-13
gb AAL05886.1 germin-like protein [<i>Musa acuminata</i>]	2e-88
gb AAG44757.1 metallothionein-like protein [<i>Musa acuminata</i>]	6e-27
gb AAT67244.1 BTF3b-like transcription factor [<i>Musa acuminata</i>]	6e-64
gb AAT35818.1 abscisic stress ripening protein-like protein [<i>Musa acuminata</i>]	6e-20
emb CAC81811.1 putative chitinase [<i>Musa acuminata</i>]	4e-77
gb AAB82778.1 ripening-associated protein [<i>Musa acuminata</i>]	4e-92
gb AAG44757.1 metallothionein-like protein [<i>Musa acuminata</i>]	2e-27
gb AAV65145.1 sucrose phosphate synthase [<i>Musa acuminata</i>]	3e-25
gb AAB71138.1 S-adenosyl-L-methionine synthetase homolog [<i>Musa acuminata</i>]	2e-88
gb AAB71138.1 S-adenosyl-L-methionine synthetase homolog [<i>Musa acuminata</i>]	e-115

gb AAB82772.2	beta-1, 3-glucanase [<i>Musa acuminata</i>]	3e-45
gb AAB71138.1	S-adenosyl-L-methionine synthetase homolog [<i>Musa acuminata</i>]	7e-63
gb AAD28755.1	cytosolic fructose-1,6-bisphosphatase [<i>Musa acuminata</i>]	e-143
gb AAO11776.1	alpha-amylase precursor [<i>Musa acuminata</i>]	3e-28
gb AAT67244.1	BTF3b-like transcription factor [<i>Musa acuminata</i>]	4e-64
gb AAT67244.1	BTF3b-like transcription factor [<i>Musa acuminata</i>]	1e-66
gb AAB82774.1	ripening-associated protein [<i>Musa acuminata</i>]	4e-35
gb AAK82991.1	actin [<i>Musa x paradisiaca</i>]	e-106
gb AAB82776.2	ripening-associated protein [<i>Musa acuminata</i>]	3e-34
gb AAF19195.1	pectate lyase 1 [<i>Musa acuminata</i>]	6e-24
gb AAO21213.1	cell wall invertase [<i>Musa acuminata</i>]	2e-56
gb AAL05885.1	endochitinase [<i>Musa acuminata</i>]	9e-66
gb AAL05886.1	germin-like protein [<i>Musa acuminata</i>]	4e-81
gb AAM10554.1	ribulose bisphosphate carboxylase large subunit [<i>Musa acuminata</i>]	9e-51
gb AAR96011.1	hypothetical protein [<i>Musa acuminata</i>]	4e-35
gb AAG44759.1	metallothionein-like protein [<i>Musa acuminata</i>]	7e-13
gb AAB82772.2	beta-1, 3-glucanase [<i>Musa acuminata</i>]	8e-66
gb AAG44759.1	metallothionein-like protein [<i>Musa acuminata</i>]	5e-11
emb CAA64856.1	1-aminocyclopropane-1-carboxylate oxidase [<i>Musa acuminata</i>]	2e-58
gb AAT67244.1	BTF3b-like transcription factor [<i>Musa acuminata</i>]	2e-37
gb AAL73122.1	putative O-deacetylbaocatin III-10-O-acetyl transferase-like	6e-09
gb AAL14240.1	proteinase inhibitor [<i>Musa acuminata</i>]	6e-16
gb AAR96010.1	calmodulin-like protein [<i>Musa acuminata</i>]	3e-52
gb AAN31756.1	expansin1 [<i>Musa acuminata</i>]	3e-70
emb CAC81811.1	putative chitinase [<i>Musa acuminata</i>]	3e-85
gb AAL27555.1	hypothetical protein [<i>Musa acuminata</i>]	4e-41
gb AAQ14245.1	actin [<i>Musa acuminata</i>]	6e-47
gb AAL06348.1	submergence induced protein-like protein [<i>Musa acuminata</i>]	8e-07
gb AAM48480.1	lectin [<i>Musa acuminata</i>]	4e-46
gb AAR96006.1	4-nitrophenylphosphatase-like protein [<i>Musa acuminata</i>]	8e-46
gb AAB82774.1	ripening-associated protein [<i>Musa acuminata</i>]	7e-30
emb CAC81811.1	putative chitinase [<i>Musa acuminata</i>]	1e-25
gb AAT35818.1	abscisic stress ripening protein-like protein [<i>Musa acuminata</i>]	2e-11
gb AAR00513.1	1-aminocyclopropane-1-carboxylate synthase [<i>Musa acuminata</i>]	e-117
gb AAR96008.1	ARIADNE-like protein [<i>Musa acuminata</i>]	5e-67
gb AAT35818.1	abscisic stress ripening protein-like protein [<i>Musa acuminata</i>]	2e-12
gb AAB82776.2	ripening-associated protein [<i>Musa acuminata</i>]	2e-41

gb AAG44758.1 metallothionein-like protein [<i>Musa acuminata</i>]	1e-19
gb AAL05886.1 germin-like protein [<i>Musa acuminata</i>]	1e-83
gb AAL82597.1 aspartate aminotransferase 2 [<i>Musa acuminata</i>]	9e-53
gb AAM69298.1 unknown [<i>Musa acuminata</i>]	5e-20
gb AAG44759.1 metallothionein-like protein [<i>Musa acuminata</i>]	1e-11
gb AAK07428.2 malate synthase [<i>Musa acuminata</i>]	7e-21
emb CAC81812.1 putative chitinase [<i>Musa acuminata</i>]	2e-08
SP O24045 RBS_MUSAC Ribulose bisphosphate carboxylase small chain, chloroplast	2e-65
SP O24045 RBS_MUSAC Ribulose bisphosphate carboxylase small chain, chloroplast	4e-18

Tabela 3. MaAES mais populosos e suas similaridades com as seqüências protéicas depositadas na base de dados Genebank *nr*.

Número de Seqüências	Seqüência no Genebank <i>nr</i>
41	dbj BAB19963.1 putative ripening protein [<i>Calystegia soldanella</i>]
48	gb AAP23943.1 CCR protein [x <i>Citrofortunella mitis</i>]
48	gb AAN60205.1 photosystem II 10 kDa protein [<i>Xerophyta humilis</i>]
58	gb AAB82774.1 ripening-associated protein [<i>Musa acuminata</i>]
64	gb AAG44759.1 metallothionein-like protein [<i>Musa acuminata</i>]
73	gb AAG44759.1 metallothionein-like protein [<i>Musa acuminata</i>]
73	emb CAB52585.1 metallothionein-like protein [<i>Elaeis guineensis</i>]
74	emb CAB52585.1 metallothionein-like protein [<i>Elaeis guineensis</i>]
85	gb AAF35186.1 lipid transfer protein precursor [<i>Gossypium hirsutum</i>]
103	gb AAT35818.1 abscisic stress ripening protein-like protein [<i>Musa acuminata</i>].
153	No Hits
176	dbj BAD83479.1 hypothetical protein [<i>Nicotiana tabacum</i>]
266	emb CAB52585.1 metallothionein-like protein [<i>Elaeis guineensis</i>]
298	gb AAN60205.1 photosystem II 10 kDa protein [<i>Xerophyta humilis</i>]
307	emb CAB52585.1 metallothionein-like protein [<i>Elaeis guineensis</i>]
407	gb AAK30204.1 endoxyloglucan transferase [<i>Daucus carota</i>]
413	gb AAR13288.1 Anx1 [<i>Gossypium hirsutum</i>]
502	sp O24045 RBS_MUSAC Ribulose bisphosphate carboxylase small chain

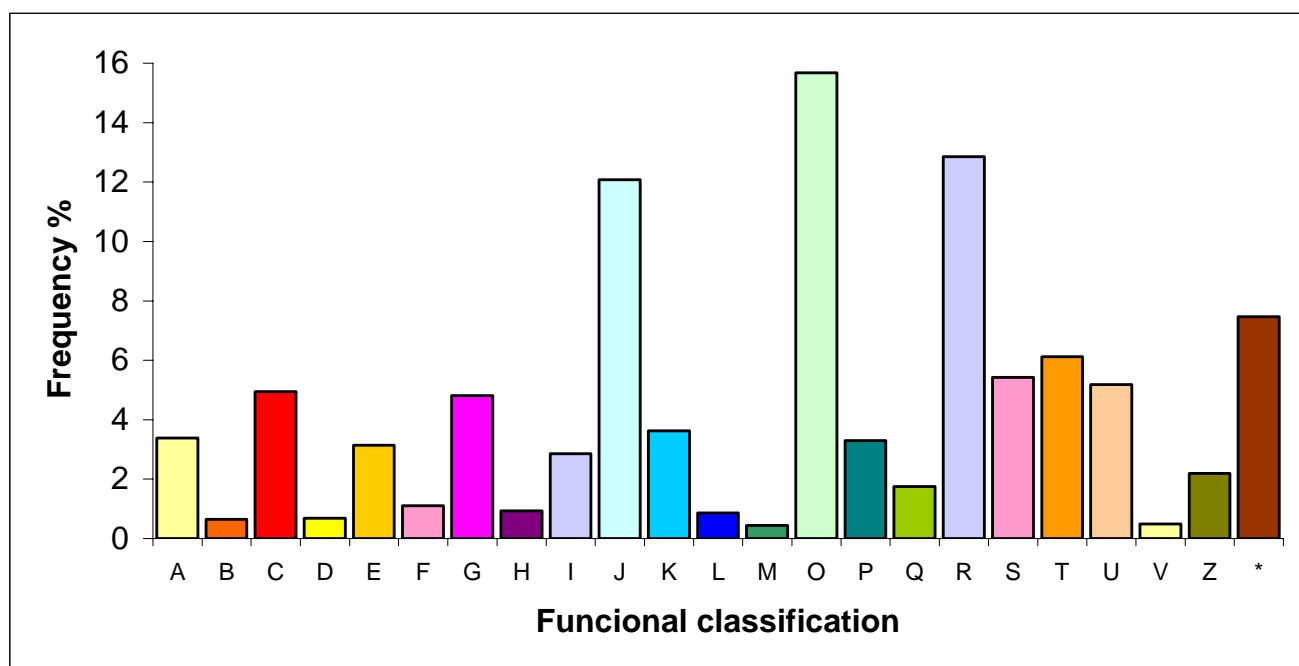


Figura 3. Classificação funcional e freqüência das classes de MaAES de acordo com Eukaryotic Clusters of Orthologs (KOG). Significado das categorias funcionais: *A*, Processamento e modificação do RNA, *B*, Estrutura e dinâmica da cromatina *C*, Produção e conversão de energia, *D*, Controle do ciclo celular e mitose, *E*, Metabolismo e transporte de aminoácido, *F*, Metabolismo e transporte de nucleotídeos, *G*, Metabolismo e transporte de carboidratos, *H*, Metabolismo de coenzimas *I*, Metabolismo de lipídeos, *J*, Tradução, *K*, Transcrição, *L*, Replicação e Reparo, *O*, Modificação pós-tradução, turnover de proteínas, função chaperona, *P*, Metabolismo e transporte de íons inorgânicos, *Q*, Biosíntese de metabólicos secundários, transporte e catabolismo, *R*, Somente predição de função geral, *S*, Função desconhecida, *T*, Transdução de sinal, *U*, Trânsito intracelular e secreção, *Z*, Citoesqueleto e * corresponde à função não conclusiva.

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. 3389-402, 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.

BOGUSKI, M. S.; SCHULER, G. D. Establishing a human transcript map. **Nature Genetics**, New York, v. 10, p. 369-371, 1995.

BRENDEL, V.; KURTZ, S.; WALBOT, V. Comparative genomics of *Arabidopsis* and maize: prospects and limitations. **Genome Biology. Reviews**, v. 3, p. 1005.1-1005.6, 2002.

CHANG, S.; PURYEAR, J.; CAIRNEY, J. A. Simple and efficient method for isolating RNA from Pine trees. **Plant Molecular Biology Reporter**, Athens, 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. Disponível em: <www.fao.org> . Acesso em: abr. 2004.

GASTEIGER, E.; JUNG, E.; BAIROCH, A. SWISS-PROT: connecting biological knowledge via a protein database. **Current Issues Molecular Biology**, 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. 868-77, 1999.

RONNING, C. M.; STEGALKINA, S. S.; ASCENZI, R. A.; BOUGRI, O.; HART, A. L.; UTTERBACH, T. R.; VANAKEN, S. E.; RIEDMULLER, S. B.; WHITE, J. A.; CHO, J.; PERTEA, G. M.; LEE, Y.; KARAMYCHEVA, S.; SULTANA, R.; TSAI, J.; QUACKENBUSH, J.; GRIFFITHS, H. M.; RESTREPO, S.; SMART, C. D.; FRY, W. E.; VAN DER HOEVEN, R.; TANKSLEY, S.; ZHANG, P.; JIN, H.; YAMAMOTO, M. L.; BAKER, B. J.; BUELL, C. R. Comparative analyses of potato expressed sequence tag libraries. **Plant Physiology**, Minneapolis, v. 131, n. 2, p. 419-29, 2003.

RUDD, S. Expressed sequence tags alternative or complement to whole genome sequences? **Trends in Plant Science**, Oxford, GB, v. 8, p. 321-329, 2003.

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 JÚNIOR, 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. 91-3, 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. **Bionformactics**, v. 4, p. 1-14, 2003.

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