

**UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO MULTI-INSTITUCIONAL
EM BIOTECNOLOGIA**

**ESTUDO DE GENES EXPRESSOS EM FRUTOS DE CAMU-
CAMU: SEQÜENCIAMENTO DE ESTs**

MARCICLEIDE LIMA DA SILVA

Manaus - Amazonas
2006

**UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO MULTI-INSTITUCIONAL
EM BIOTECNOLOGIA**

MARCICLEIDE LIMA DA SILVA

**ESTUDO DE GENES EXPRESSOS EM FRUTOS DE CAMU-
CAMU: SEQÜENCIAMENTO DE ESTs**

Tese apresentada ao Programa de Pós-graduação Multi-institucional em Biotecnologia da Universidade Federal do Amazonas - UFAM, como requisito parcial para a obtenção do título de Doutor em Biotecnologia, área de concentração Agroflorestal.

Orientador: Dr. Kaoru Yuyama - INPA

Co-orientadores: Dr. Spartaco Astolfi Filho - UFAM

Dra. Elionor R. Almeida - EMBRAPA/CENARGEN

Manaus - Amazonas
2006

Silva, Marcicleide Lima da

S586e Estudo de genes expressos em frutos de camu-camu:
seqüenciamento de ESTs / Marcicleide Lima da Silva. -
Manaus: UFAM, 2006.

157 f.; il.

**Tese (Doutorado Multi-institucional em
Biotecnologia) – Universidade Federal do Amazonas,
2006.**

Orientador: Dr. Kaoru Yuyama

1. Camu-camu 2. *Myrciaria dúbia* 3. Sequência
Expressa de DNA I.Título

CDU 634.32(043.2)

MARCICLEIDE LIMA DA SILVA

ESTUDO DE GENES EXPRESSOS EM FRUTOS DE CAMU-CAMU:
SEQÜENCIAMENTO DE ESTs

Tese apresentada ao Programa de Pós-graduação Multi-institucional em Biotecnologia da Universidade Federal do Amazonas - UFAM, como parte do requisito para a obtenção do título de Doutor em Biotecnologia, área de concentração Agroflorestal.

Aprovada 22 de junho de 2006.

BANCA EXAMINADORA

Prof^o. Dr. Kaoru Yuyama, Presidente
Instituto Nacional de Pesquisa da Amazônia

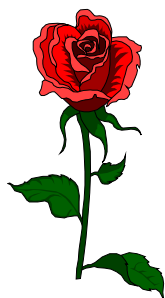
Prof^o. Dr. Luiz Antônio Serrão Contim, membro
Centro Universitário Nilton Lins

Dra. Nelcimar Reis Souza, membro
Empresa Brasileira de Pesquisa Agropecuária

Prof^a. Dra. Maria Teresa Gomes Lopes, membro
Universidade Federal do Amazonas

Prof^o, Dr. Charles Roland Clement, membro
Instituto Nacional de Pesquisas da Amazônia

Dedicatória



Cabeça de luz
Cabeça bela
Sempre buscando o céu
Corpo sofrido
Marcas
Corpo de espinhos
Que testam
Protegem
Ajudam a crescer
E o contraste é belo
Mas não vale tocar
É o natural
Talvez um dia não terá espinhos
E será como um anjo de luz
Em que se misturam beleza e brilho
E o caule?
Sempre flexível
E sem espinhos
Mas eternamente ROSA.

Dedico esta tese a Deus e a Rosa do Carmo Lima da Silva j, homenagem póstuma a minha amada mãe e agradecimento pelo amor, dedicação e estímulo para que eu continuasse a ir em busca dos meus sonhos.

Agradecimentos

Acima de tudo dou graças a Deus, por ter me concedido vida, a inteligência e a capacidade de buscar novos conhecimentos. Inicialmente a Superintendência da Zona Franca de Manaus e posteriormente ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de doutorado no Programa de Pós-graduação Muti-institucional em Biotecnologia. A Universidade Federal do Amazonas, ao Instituto Nacional de Pesquisas da Amazônia e ao Centro de Recursos Genéticos e Biotecnologia - CENARGEN que me concederam a oportunidade para a realização dos estudos e pesquisas que compõem este trabalho. Ao Dr. Kaoru Yuyama pela orientação, amizade e confiança em mim depositada. Ao Dr. Spartaco Astolfi Filho e a Dra. Elionor Rita Pereira de Almeida pela co-orientação segura em todos os aspectos e pela confiança em mim depositada. Ao meu esposo Demóstenes do Espírito Santo pelo amor, compreensão e apoio dado durante o desenvolvimento deste trabalho. À técnica do laboratório de tecnologia do DNA da UFAM, Enedina Assunção, pela cordialidade e presteza, em nome de quem agradeço a todos os membros do laboratório que contribuíram para a realização deste trabalho.

Resumo

O camu-camu (*Myrciaria dúbia* (H.B.K.) McVaugh) é uma espécie nativa da região Amazônica, cujo fruto apresenta elevado teor de ácido ascórbico (vitamina C). O estudo do genoma funcional em frutos de camu-camu tem como base o seqüenciamento de fragmentos de seqüências expressas - ESTs (*Expressed Sequence Tags*). Diante do exposto a presente tese pretende analisar e identificar genes expressos em frutos de camu-camu por meio de seqüenciamento de ESTs. O RNA total foi extraído a partir da casca-polpa. O seqüenciamento da extremidade 5' de insertos de cDNA foi realizado tanto no Laboratório de Tecnologia do DNA da UFAM e como na Plataforma de Seqüenciamento da EMBRAPA/CENARGEN. As seqüências ESTs obtidas foram submetidas ao Sistema Genoma, programa de anotação genômica que integra programas de gerenciamento de análise e visualização de seqüências nucleotídicas. Os resultados obtidos foram o desenvolvimento de um procedimento eficiente para extração de RNA total de frutos de camu-camu que possibilitou a obtenção de mRNAs de qualidade, utilizados na confecção de cDNAs de tamanhos variados (500pb a 4Kb). A partir do seqüenciamento foram obtidas 3196 ESTs válidas, sendo formados 1546 *singletons* e 358 *contigs*, resultando num total de 2586 seqüências ESTs com similaridade a seqüências encontradas no banco de genes. A análise da clusterização da biblioteca revelou um índice de 81% de novidade e 33% de redundância. Cerca de 90% dos *contigs* apresentaram baixa redundância (2-4 *reads* por *contigs*). Os dados da categorização das proteínas identificadas destacaram a categoria modificação pós-traducional, proteína *turnover*, chaperonas (13,2%). A partir do levantamento das espécies com maior número de ESTs com similaridade a seqüências de camu-camu destacou-se *Arabidopsis thaliana* com 49%. Cerca de 10 *uniques* apresentaram altíssima similaridade (*e-value* 0.0) a genes conhecidos. Os ESTs mais abundantemente expresso em frutos de camu-camu codificam a glutathione s-transferase. Foram observados cerca de 3% de seqüências (97 ESTs) com baixa similaridade (*e-value* > e^{-10}) e 15% não apresentaram similaridade com nenhuma seqüência contida no banco de genes. Foram identificadas 138 seqüências ESTs (4,3%) que codificam chaperonas moleculares com destaque à família sHSP que representa 33% das chaperonas expressas. ESTs relacionados ao metabolismo do ácido ascórbico também foram identificados, sendo nove relacionados a síntese e seis voltados para conversão e reciclagem do ácido ascórbico. ESTs relacionados ao amadurecimento e mecanismos de defesa do fruto também foram destacados.

Abstract

The camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh) is a native sort of the Amazonian region, whose fruit presents elevated content of ascorbic acid (vitamin C). The study of the functional genome in camu-camu fruits has like base the expressed sequence tags sequencing - ESTs. Faced with the displayed the present thesis is going to analyze and identify express genes in camu-camu fruits by means of ESTs sequencing. The total RNA was extracted from the shell-pulp. The extremity 5' sequencing' of cDNA insert was carried out so much in the Technology of the DNA Laboratory (UFAM) and in the Sequencing Platform (EMBRAPA/CENARGEN). The ESTs sequences obtained were submitted to the System Genome, program of genomic annotation that integrates analysis management programs and viewing of nucleotides sequences. It developed an efficient procedure for total RNA extraction of camu-camu fruits that enabled the obtaining of mRNAs of quality, utilized in the making of cDNAs of sizes varied (500pb to 4Kb). From the sequencing were obtained 3196 ESTs valid, being formed 1546 singletons and 358 contigs, resulting of 2586 ESTs sequences in total with similarity the sequences found in the gene bank. The analysis library clusterization revealed an index of 81% novelty and 32,54% redundancy. Around 90% of the contigs presented decrease redundancy (2-4 reads by contigs). The facts of the categorization of the proteins identified detached the posttranslational modification, protein turnover, chaperones (13,2%) category. From the hoist of the species with bigger number of ESTs with similarity the camu-camu sequences detached itself *Arabidopsis thaliana* with 49%. Around 10 uniques presented very high similarity (and-value 0.0) to known genes. The ESTs more abundantly express in camu-camu fruits encode to glutathione s-transferase. They were observed around 3% sequences (97 ESTs) with decrease similarity (e-value > e-10) and 15% did not they present similarity with no contained sequence in the gene bank. They were identified 138 ESTs sequences (4,3%) that they encode molecular chaperones with prevalence of the sHSP family that represents 33% of express chaperones. ESTs related to the ascorbic acid metabolism also were identified, being nine related the synthesis and six come back for ascorbic acid conversion and recycling. ESTs related to the ripening and mechanisms of defense of the fruit also were noticeable.

LISTA DE FIGURAS

Introdução geral	
Figura 01 - Aspecto geral da planta de camu-camu	02
Figura 02 - Aspecto geral dos ramos e frutos de camu-camu	03
Capítulo 1	
Figura 01 - Aspectos gerais de frutos de camu-camu	14
Figura 02 - Aspecto da polpa e casca de frutos de camu-camu	17
Figura 03 - Esquema do protocolo de síntese de cDNA e ligação no vetor pSPORT 1 ..	21
Figura 04 - Mapa do plasmídeo pSPORT 1, as setas indicam o local de inserção dos cDNAs	22
Figura 05 - Perfil eletroforético do RNA Total extraído da polpa e casca de frutos de camu-camu. As setas indicam RNAs ribossomos	26
Figura 06 - Perfil eletroforético de fracionamento do cDNA obtidos a partir de fruto de camu-camu	27
Figura 07 - Perfil eletroforético dos plasmídios extraídos a partir de clones transformados com sistema de ligação contendo cDNA de frutos de camu- camu	28
Figura 08 - Perfil eletroforético de produtos de PCR a partir de plasmídios contendo insertos de cDNA de frutos de camu-camu (Marcador de 1Kd)	28
Capítulo 2	
Figura 01 - Categorização funcional das prováveis proteínas identificadas a partir da biblioteca de ESTs de frutos de camu-camu	46
Figura 02 - Nível de similaridade encontrada na biblioteca de cDNA de frutos de camu-camu (3196 ESTs válidos)	51
Figura 03 - Percentagem de EST com similaridade a seqüências de diferentes organismos vegetais	53
Capítulo 3	
Figura 01 - Esquema da estratégia de anotação de chaperonas molecular da biblioteca de ESTs de camu-camu armazenada no programa Sistema Genoma (BIOFOCO)	70
Figura 02 - Percentual de genes potenciais codificantes de famílias de chaperonas, anotadas de seqüências ESTs de frutos de camu-camu.....	72
Capítulo 4	
Figura 01- Esquema da estratégia de anotação de possíveis genes relacionados ao metabolismo do ácido ascórbico da biblioteca de ESTs de camu-camu armazenada no programa Sistema Genoma (BIOFOCO)	93
Figura 2 - Biossíntese, degradação e reciclagem do ácido ascórbico (A.As.) em plantas	98

LISTA DE TABELAS

Introdução geral	
Tabela 01 - Teores de ácido ascórbico do camu-camu e da acerola analisados por diversos autores	04
Capítulo 2	
Tabela 01 - Relação de organismos vegetais cujos transcriptomas foram analisados via sequenciamento de ESTs	39
Tabela 02 - Sumário do sequenciamento de EST de cDNA de frutos de camu-camu	43
Tabela 03 - Relação de seqüências ESTs não identificadas e não similares a seqüências do banco de genes	50
Tabela 04 - As 20 seqüências formadas a partir de ESTs de frutos de camu-camu com alta similaridade a seqüências do banco de genes	52
Tabela 05 - Os 20 genes mais expressos na biblioteca de frutos de camu-camu, representados pelo maior número de ESTs relacionado	54
Capítulo 3	
Tabela 01 - Anotação de genes potenciais codificantes de chaperonas molecular em frutos de camu-camu	74
Capítulo 4	
Tabela 01 - Relação de possíveis genes identificados na biblioteca de ESTs de camu-camu com similaridade a proteínas relacionadas ao metabolismo do ácido ascórbico	96
Outros genes de interesse biotecnológico e perspectivas futuras	
Tabela 01 - Relação de genes de interesse biotecnologia em frutos	106

Sumário

Introdução Geral	1
1. Camu-camu	1
2. Justificativa	5
3. Objetivos Geral	7
Objetivos específicos	7
4. Referências Bibliográficas	8
Capítulo 1 – Desenvolvimento de protocolo de extração de RNA total e construção da biblioteca de cDNA de frutos de camu-camu de plantas cultivadas em terra-firme	11
Resumo	12
<i>Abstract</i>	13
1. Introdução	14
2. Materiais e Métodos	17
2.1. Material biológico	17
2.2. Extração de RNA total	18
2.3. Construção da Biblioteca de cDNA	20
2.4. Seqüenciamento de cDNAs	23
3. Resultado e Discussão.....	25
4. Conclusões	30
5. Referências Bibliográficas	31
Capítulo 2 – Anotação funcional da biblioteca de ESTs (<i>Expressed Sequence Tags</i>) de frutos de camu-camu - <i>CAMUEST</i>	33
Resumo	34
<i>Abstract</i>	35
1. Introdução	36
2. Materiais e Métodos	41
3. Resultado e Discussão	43
3.1. Análise da Biblioteca	43
3.2. Categorização funcional das proteínas identificadas	44
3.3. Análise da similaridade das seqüências ESTs	49
4. Conclusões	57
5. Referências Bibliográficas	58
Capítulo 3 – Identificação de potenciais genes codificantes de chaperonas	

em frutos de camu-camu	62
Resumo	63
<i>Abstract</i>	64
1. Introdução	65
2. Materiais e Métodos	70
3. Resultados e Discussão	71
4. Conclusões	79
5. Referências Bibliográficas	80
Capítulo 4 – Potenciais genes relacionados ao metabolismo do ácido ascórbico em frutos de camu-camu	85
Resumo	86
<i>Abstract</i>	87
1. Introdução	88
2. Materiais e Métodos	93
3. Resultado e Discussão	94
3.1. Biossíntese de ácido ascórbico	94
3.2. Conversão e reciclagem de ácido ascórbico	95
4. Conclusões	100
5. Referências Bibliográficas	101
Outros genes de interesse biotecnológico e perspectivas futuras	105
Referências Bibliográficas	112
Apendice – Relação geral de prováveis genes identificados expressos em frutos de camu-camu	113

PREFÁCIO

No Brasil o estudo genômico com plantas teve início o seqüenciamento da cana-de-açúcar (*Saccharum officinarum*) e na Amazônia tem se destacado as atividades da Rede da Amazônia Legal de Pesquisas Genômicas – REALGENE que vem estudando por meio de ESTs o genoma funcional de frutos do guaraná (*Paullinia cupana* Kunth).

O Instituto Nacional de Pesquisas da Amazônia juntamente com a Universidade Federal do Amazonas e o Centro de Recursos Genéticos de Biotecnologia da EMBRAPA, em 2002, iniciaram o estudo genômico funcional do camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh), em função do seu grande potencial econômico, por apresentar frutos ricos em ácido ascórbico (Vitamina C), com variabilidade ampla entre 0,845 a 6,1g de ácido ascórbico/100g de polpa.

Esta tese é parte desses estudos visando analisar e identificar genes expressos em frutos de camu-camu, formando uma coleção de genes que poderão ser utilizados como base para a domesticação e no melhoramento genético da planta.

A idéia de escrever em capítulos, de acordo com o que é apresentado, procura trazer uma maior dinâmica à leitura e facilitar a busca por informações específicas.

No capítulo 1 será abordado o estabelecimento de um protocolo de extração de RNA total de frutos de camu-camu. Não existe um protocolo de extração de RNA total para frutos de camu-camu, o elevado teor de polissacarídeos dificulta a extração por protocolos convencionais, descrevemos o método que desenvolvemos para extração de RNA total. O capítulo também tratará da confecção da primeira biblioteca de cDNA de frutos de camu-camu de plantas cultivadas em terra-firme e do seqüenciamento de ESTs (*Expressed Sequence Tags*).

No Capítulo 2 será apresentada uma avaliação geral das seqüências obtidas na biblioteca de ESTs de frutos de camu-camu, por meio de ferramentas de

bioinformática, onde serão mostrados dados como número de *contigs* e *singletons*, redundância, os potenciais genes mais expressos e com maior similaridade a seqüências contidas no banco de genes nr/Blast do NCBI; além da relação dos organismo com proteínas mais similares a do camu-camu, e a categorização das prováveis proteínas identificadas e seu provável papel em frutos.

As chaperonas foi o grupo de genes com maior representação em termos de número de ESTs obtidos na biblioteca de camu-camu, em função disto o Capítulo 3 foi dedicado a avaliação destas proteínas que parecem ter grande importância no desenvolvimento de frutos de camu-camu cultivado em terra-firme.

O camu-camu tem por seu principal atrativo o elevado teor de vitamina C, os resultados da avaliação de potenciais genes, presentes na biblioteca de ESTs de frutos de camu-camu, relacionados a biossíntese, degradação e regeneração do ácido ascórbico foram discutidos no Capítulo 4 desta tese, onde também foram apresentados os atuais avanços no estudo do metabolismo do ácido ascórbico em plantas.

Por fim, além das perspectivas futuras, outros potenciais genes de interesse biotecnológico foram destacados no último tópico da tese, entre eles: os da biossíntese de antocianina, um importante corante natural; genes relacionados ao estresse oxidativo, que tem papel na defesa celular contra espécies reativas de oxigênio; e genes relacionados aos fatores de transcrição importante na regulação da expressão gênica. Estes genes são apenas uma pequena representação do potencial da biblioteca de ESTs de frutos de camu-camu, em termos de utilização biotecnológica.

INTRODUÇÃO GERAL

1. CAMU-CAMU

O camu-camu (*Myrciaria dubia* (H.B.K.) Mc Vaugh) pertence à família Myrtaceae, ao gênero *Myrciaria*, e apresenta várias sinonímias botânicas que incluem: *Myrciaria paraensis* Berg (McVAUGH, 1969); *Myrciaria divaricata* (Bertham) O. Berg; *Myrciaria spruceana* O. Berg; *Psidium dubium* H.B.K. (VILLACHICA, 1996).

No Brasil é popularmente conhecido por camu-camu, araçá-d'água, caçari, araçá de igapó, azedinha; no Peru, camu-camu; na fronteira entre a Colômbia e Venezuela, guayabo; e na Venezuela por guayabito (FERREIRA, 1986; SUAREZ MERA, 1987; CHÁVEZ FLORES, 1988; CAVALCANTE, 1991).

A planta adulta é um arbusto que pode alcançar de 3 a 8 m de altura (RUIZ, 1994; CORREA, 2000) (Figura 01). As raízes são profundas (pivotantes) e do tipo cônica, com muitas raízes secundárias horizontais e pêlos absorventes (RUIZ, 1994; CORREA, 2000). Os ramos são glabros, cilíndricos, lisos, de coloração marrom clara e as cascas se desprendem de forma natural (FERREYRA, 1959; YUYAMA, 1999). As folhas são opostas, simples, pecioladas (5 a 8mm), ovaladas-elípticas até lanceoladas (FERREYRA, 1959; CORREA, 2000), ápice acuminado e base arredondada, borda inteira e ligeiramente ondulada, nervura central aplanada na superfície adaxial e proeminente na abaxial das folhas (RUIZ, 1994).

O fruto é do tipo baga (CAVALCANTE, 1991), globoso de superfície lisa e brilhante, de cor roxa a vermelha escura até preta púrpura ao amadurecer (CORREA, 2000) (Figura 02). As sementes têm formato reniforme em número de 1 a 4 por fruto, sendo conspicuamente aplanadas, de coloração roxa a avermelhada, cobertas por uma camada fina de pêlo (FERREYRA, 1959; CALZADA BENZA, 1980).



Figura 01 - Aspecto geral da planta de camu-camu

A inflorescência é axilar em forma de racemos (FALCÃO et al., 1993). As flores individuais de camu-camu são hermafroditas, de cor branca com 4 pétalas e 4 sépalas (RUIZ, 1994); o ovário é ínfero; estames em número de 125 com comprimentos semelhantes ao do estilete, aproximadamente 1 cm (FERREYRA, 1959; CORREA, 2000).

O camu-camu apresenta inúmeras variantes morfológicas. Há indícios da ocorrência de diferentes ecótipos em toda a bacia do rio Solimões/Amazonas, chamando a atenção para a existência de um tipo arbustivo e outro arbóreo, com *status* taxonômico ainda não definido (TEIXEIRA et al., 2004).

A caracterização populacional de camu-camu na região Amazônica foi estudada por meio de isoenzimas de esterase e esterase-D que demonstraram que ocorrem diferenças genéticas entre as populações camu-camu oriundas de Iquitos (Perú), Uatumã (Amazonas) e de Boa Vista (Roraima) (TEIXEIRA et al., 2004).

Quanto a amplitude de dispersão, o camu-camu abrange uma grande extensão da região norte do Brasil e em países distintos da Amazônia (YUYAMA, 1999). McVaugh (1969) estabeleceu a área de distribuição do camu-camu, estendendo-se de leste a oeste da região central do Estado do Pará, passando pelo meio e alto Amazonas até a parte oriental do Peru. Ao norte chega ao rio Casiquiare e às partes altas e média da bacia do rio Orinoco.



Figura 02 - Aspecto geral dos ramos e frutos de camu-camu

O camu-camu é encontrado naturalmente nas margens dos rios, lagos e igapós de toda a região amazônica, predominante nas águas pretas, em substratos aluviais, de textura limosa, argilosa, limo-argilosa, limo arenosa, e em solos drenados (CALZADA BENZA, 1980; SUAREZ MERA, 1987). Segundo Falcão et al. (1993), ocorre também, naturalmente e sem grande distinção, nas margens de águas barrentas e brancas.

Em relação a composição nutricional, o fruto apresenta elevado teor de vitamina C (ácido ascórbico) com variabilidade ampla entre 0,845g a 6,112g de ácido

ascórbico/100g de polpa (ZAPATA & DUFOUR, 1993; YUYAMA, K. et al., 2002), superior à da acerola (1,79g/100g de polpa), até então considerada como a fruta mais rica nesta vitamina (MATSUURA et al., 1998; SANCHES et al., 1998) (Tabela 1).

O conteúdo de proteínas varia entre 400mg a 890mg/100g de polpa e o de carboidratos entre 3500mg a 4700mg/100g de polpa de camu-camu (ROCA, 1965; ANDRADE, 1991; JUSTI et al., 2000). A concentração de fibra alimentar varia entre 0,43g a 6,18g por 100g de fruto apresentando uma adequação na ordem de 11,5% da quantidade de fibra recomendada (YUYAMA, L. et al., 2002)

FRUTO	TEOR DE ÁCIDO ASCÓRBICO*	AUTORES
Camu-camu	2880	ROCA (1965)
Camu-camu	845 - 939	ZAPATA E DUFOUR (1993)
Camu-camu	2489 - 3133	ANDRADE et al. (1995)
Camu-camu	1910 - 2061	ALVES et al. (2000)
Camu-camu	1380 - 1490	JUSTI et al. (2000)
Camu-camu	5737 - 6112	YUYAMA, K. et al. (2002)
Acerola	1038 - 2318	MATSUURA et al. (1998)
Acerola	1558 - 2560	SANCHES et al. (1998)

*(mg/100g de polpa)

Tabela 01 - Teores de ácido ascórbico do camu-camu e da acerola analisados por diversos autores.

A composição mineral, em 100g de polpa de camu-camu, inclui o sódio (27 a 111,3mg); potássio (532 a 838,8mg), cálcio (62 a 157,3mg), magnésio (47 a 123,8mg), manganês (1,4 a 21,1mg), zinco (1,2 a 3,6mg), cobre (0,5 a 2,0mg) e Ferro (1,3 a 1,8mg) (ZAPATA & DUFOUR, 1993; JUSTI et al., 2000). Os ácidos graxos presentes nos frutos de camu-camu são os tridecanoicos, palmiticos, esteáricos, oléicos, linoléicos, γ -linoléico, α -linoléico, eicosadienóico, EPA e trisanóico (JUSTI et al., 2000).

Os principais pigmentos encontrados no fruto do camu-camu são as antocianinas responsáveis pela coloração vermelho-púrpura do fruto maduro, que podem alcançar, em 113 dias após a antese, 0,8mg em 100g de polpa de frutos, sendo do tipo cianidina-3-glicose (ANDRADE, 1993; ZANATTA et al, 2005); e os carotenóides, que em geral

decrecem com a maturação do fruto, alcançando aos 113 dias, menos que 0,2mg em 100g de polpa (ANDRADE, 1993). Estudos cromatográficos com HPLC-DAD e HPLC-MS, revelaram que a luteína é o carotenóide mais abundante em frutos de camu-camu, seguido pelo β -caroteno e a zeaxantina. Os carotenóides neoxantina, β -criptoxantina, 5,6-epóxido e o *cis*- β -caroteno se apresentam em baixos níveis (AZEVEDO-MELEIRO & RODRIGUEZ-AMAYA, 2004).

No estudo de compostos voláteis em frutos de camu-camu, observou-se os seguintes componentes: acetato de etil, α -pineno, α -fenqueno, butirato de etil, canfeno, car-3-eno, α -felandreno, α -terpineno, eucalyptol, terpinoleno, fencol, 4-terpineno, humuleno, γ -terpineno, β -pineno, β -myrceno, β -felandreno, d-limoneno, p-cimeno, β -cariofileno (FRANCO E SHIABAMOTO, 2000).

Ueda et al. (2004) isolaram do camu-camu o ácido elágico e dois derivados, o ácido 4-O-metilelágico e o ácido 4-(α -ramnopiranosil) elágico como inibidores de aldose redutase (AR). A aldose redutase cataliza glicose em sorbitol na via poli-ol do metabolismo da glicose. Em paciente com *diabetes melitos* esta via é acelerada, formando sorbitol em tecidos insensíveis a insulina. Inibidores de AR são capazes de prevenir a redução da glicose a sorbitol e reduzindo complicações nos diabéticos (MIZUNO et al., 1999).

2. JUSTIFICATIVA DA TESE

A descoberta de uma espécie amazônica com elevado teor de vitamina C surpreendeu a comunidade científica brasileira. Apesar do camu-camu ser pouco conhecido pelas comunidades amazônicas brasileira ele já vem sendo bastante consumido em diferentes países, não apenas como complemento nutricional mas como

um fitoterápico no tratamento da depressão, no fortalecimento do sistema imune e com grande potencial na fabricação de cosméticos.

O mercado consumidor do camu-camu está em constante crescimento. A demanda pela polpa congelada de camu-camu em Manaus e em outras cidades do país vem despertando o interesse na domesticação desta espécie e ampliando a viabilidade econômica do cultivo desta espécie em terra-firme, uma vez que os frutos ainda são coletados de plantas nativas (várzea e igapó).

Obter plantas com o teor de vitamina C determinado é de grande importância e tem sido o principal objetivo dos pesquisadores. O desenvolvimento de material melhorado tem sido um desafio em função da grande variabilidade genética das populações de camu-camu.

O estudo dos genes expressos em frutos de camu-camu, abordagem desta tese, é um importante passo para compreender os fatores genéticos determinantes no acúmulo e na variação das concentrações de vitamina C em camu-camu de diferentes populações. Adicionalmente, a identificação desses genes poderá auxiliar na elucidação da via biossintética da vitamina C em plantas.

Não existem estudos sobre a expressão de genes em frutos de camu-camu sendo este nosso trabalho pioneiro. Neste sentido, propusemos na presente tese realizar uma abordagem genômica, na qual utilizamos o seqüenciamento de fragmentos de seqüências expressas - ESTs (*Expressed Sequence Tags*), técnica atualmente bastante utilizada em muitos projetos genoma.

Esta técnica permite não apenas a identificação de genes, mas também, mostrar o nível de complexidade e a influência dos fatores ambientais na expressão de genes em tecidos-específicos; além de poder ser utilizados para caracterização da variabilidade genética de populações, e mais recentemente, por meio de marcadores

moleculares do tipo microssatélites presentes nestas seqüências (STERKY & LUNDEBERG, 2000; CORDEIRO et al., 2001; VARSHNEY et al., 2005).

Além do mais a análise dos ESTs poderá contribuir para um maior conhecimento sobre o camu-camu, informações que poderão auxiliar em programas de melhoramento genético e conseqüente domesticação da espécie, pelo uso como marcadores funcionais para seleção de plantas com características desejáveis.

3. OBJETIVO GERAL

- Analisar e identificar genes expressos em frutos de camu-camu de plantas cultivadas em terra-firme, por meio de seqüenciamento de ESTs (*Expressed Sequence Tags*).

Objetivos Específicos

- Estabelecer protocolo de extração de RNA total a partir da casca-polpa de frutos de camu-camu;
- Confeccionar biblioteca direcional de populações de cDNAs de frutos de camu-camu;
- Sequenciar 3000 ESTs;
- Analisar as seqüências ESTs por meio de ferramentas de bioinformática para a identificação dos genes expressos nesses frutos.

4. REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, R. E.; BORGES, M. F.; MOURA, C. F. H. Camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh). In: ALVES, R. E.; FIGUEIRAS, H. A. C.; MOURA, C. F. H. Caracterização de frutas nativas da América Latina. Jaboticabal. FUNEP, pp. 23-26, 2000.
- AZEVEDO-MELEIRO, Cristiane H.; RODRIGUES-AMAYA, Delia B. Confirmation of the identity of the carotenoids of tropical fruits by HPLC-DAD and HPLC-MS. *J. Food Comp. Anal.*, v. 17, pp. 385-396, 2004.
- ANDRADE, Jerusa S.; GALEZI, Maria A. M.; ARAGÃO, Carlos G.; CHÁVEZ-FLORES, Wanders B. Valor nutricional do camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh) cultivado em terra firme da Amazônia Central. *Rev. Bras. Frutic.*, Cruz das Almas, v.13, n.3, pp. 307-311, 1991.
- ANDRADE, Jerusa. S. Uso de atmosfera modificada e refrigeração para manutenção da qualidade pós-colheita do camu-camu (*Myrciaria dubia* (H.B.K.)McVaugh), 50p., 1993. (Tese)
- ANDRADE, Jerusa C.; GALEZI, Maria A. M.; ARAGÃO, Carlos G.; FERREIRA, Sidney A. N. Changes in the concentration of total vitamin C during maturation and ripening of camu-camu (*Myrciaria dubia* (H.B.K.) Mc Vaguh) fruits cultivated in the upland of brasilian central amazon. *Acta Horticulture*, v. 370, pp. 177-180, 1995.
- CALZADA BENZA, J. C. 143 Frutales nativos. La Molina: Librería El Estudiante., 314 p., 1980.
- CAVALCANTI, P. B. Frutas cosmetíveis da Amazônia. 5ª. ed. Belém:CUJUP, 279p., 1991.
- CHÁVEZ FLORES, Wanders B. A importância econômica do camu-camu. Toda fruta, São Caetano do Sul. v. 3, n. 27, pp. 37-7, 1988.
- 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*, v. 160, pp. 1115-1123, 2001.
- CORREA, S.I. Cultivo de camu-camu *Myrciaria dubia* H.B.K. en la region loreto. Lima-Perú:INIA, 32p., 2000. (Série: Manual n°. 01-00)
- FALCÃO, Martha A.; FERREIRA, Sidney A. N.; CHÁVEZ-FLORES, Wanders B.; CLEMENT, Charles R. Aspectos fenológicos e ecológicos do "camu-camu" (*Myrciaria dubia* (H.B.K.) Mc Vaugh) na terra firme da Amazônia Central. In: *Aspéctos fenológicos, ecológicos e de produtividade de algumas na Amazônia brasileira: araçá-boi (Eugenia stipitata Mc Vaugh), Biribá (Rollinia mucosa (Jacq.) Baill), camu-camu (Myrciaria dubia (H.B.K.) Mc Vaugh), cupuaçu (Theobroma grandiflorum (Willd. Ex Spreng.) Schum) e graviola (Annona muricata L.)*, pp. 57-65, 1993.
- FERREIRA, Sidney A. N. Camu-camu. Campinas. *Informativo da Sociedade Brasileira de Fruticultura*, v.5, n. 2, pp. 11-12, 1986.
- FERREYRA, H.R. El "camu-camu" nueva fuente natural de vitamina C. *Bol. Exp. Agropecuária*. Lima. v.7, n. 4, pp. 28-31, 1959.
- FRANCO, Maria R. B.; SHIABAMOTO, Takayushi. Volatile composition of some Brazilian Fruits: Umbu-caja (*Spondias citherea*), camu-camu (*Myrciaria dubia*), araçá-boi (*Eugenia stipitata*) and cupuaçu (*Theobroma grandiflorum*), *J. Agric. Food Chem*, v. 48, pp. 1263-1265, 2000.
- JUSTI, Karin C.; VISENTAINER, Jesui V.; SOUZA, Nilson E.; MATSUSHITA, M. Nutritional composition and vitamin C stability in stored camu-camu (*Myrciaria dubia*) pulp. *Archivos Latinoamericanos de Nutricion*, v. 50, n. 4, pp. 405-408, 2000.

- MATSUURA, F.C.A.U.; CARDOSO, R.L.; OLIVEIRA, J.R.P.; OLIVEIRA, J.A.B.; SANTOS, D.B. Determinações físico-químicas de frutos de genótipos de acerola (*Malpighia glabra* L.). Congresso Brasileiro de Fruticultura, 15. Poços de caldas, 18-23/10-1998. *Resumos*. Lavras, SBF, p.65, 1998.
- McVAUGH, R. Botany of the Guyana highland. Part. VIII. *Memoirs of the New York Botanical Garden*. New York. v.18, n. 2, pp. 55-286, 1969.
- ROCA, N.A. *Estudio químico bromatológico de la Myrciaria paraensis* Berg. Lima, Universidade Nacional Mayor de San Marcos, Facultad de Química. 56p., 1965. (Tese)
- RUIZ, R.R. *Manejo e industrialización de los frutales nativos en la amazonia peruana*. INIA – Pucallpa, pp. 9-18, 1994.
- SANCHES, J.; KANESIRO, M.A.B.; DURIGAN, J.F. Efeito do tempo de armazenamento na qualidade de polpa de acerola. Congresso Brasileiro de Fruticultura, 15. Poços de caldas, 18-23/10-1998. *Resumos*. Lavras, SBF, p.65, 1998.
- STERKY, Fredrik; LUNDEBERG, Joakim. Sequence analysis of genes and genomes. *J. Biotech.*, v. 76, pp. 1-31, 2000.
- SUAREZ MERA, P.A. Camu-camu *Myrciaria dubia* (H.B.K.) McVaugh. In: PRANCE, G.T. *Botânica econômica de algumas espécies amazônicas*. Manaus, INPA/FUA, s.p., 1987.
- TEIXEIRA, Aylton S.; CHAVES, Lena da S.; YUYAMA, Kaoru. Esterases no exame da estrutura populacional de Camu-camu (*Myrciaria dubia* (Kunth) McVaugh-Myrtaceae). *Acta Amazônica*, v. 34, n. 1, pp. 89-96, 2004.
- UEDA, H.; KUROIWA, E.; TACHIBANA, Y.; KAWANISHI, K.; AYALA, F.; MORIYASU, M. Aldose reductase inhibitors from the leaves of *Myrciaria dubia* (H.B. & K.) McVaugh. *Phytomedicine*, v. 11, pp. 652-656, 2004.
- VARSHNEY, Rajeev K.; SIGMUND, Ralf; BÖRNER, Andreas; KORZUN, Víctor; STEIN, Nils; SORRELLS, Mark E.; LANGRIDGE, Peter; GRANER, Andreas. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science*, v. 168, pp. 195-202, 2005.
- VILLACHICA, H. *El cultivo del camu-camu (Myrciaria dubia (H.B.K.) Mc Vaugh) en la Amazonia Peruana*. Tratado de Cooperación Amazonia, Lima-Peru, 95p., 1996.
- YUYAMA, Kaoru. Banco ativo de germoplasma de camu-camu (*Myrciaria dubia* (H.B.K.) Mc Vaugh). In: Ferreira, F. R. (ed.). *Recursos genéticos de espécies frutíferas no Brasil*. Brasília, EMBRAPA/CENARGEN, pp. 90-93, 1999.
- YUYAMA, Kaoru; AGUIAR, Jaime P. L.; YUYAMA, Lúcia K. O. Camu-camu: um fruto fantástico como de vitamina C. *Acta Amazonica*, v. 32, n. 1, pp. 169-174, 2002.
- YUYAMA, Lúcia K. O.; BARROS, Solimar E.; AGUIAR, Jaime P. L.; YUYAMA, Kaoru; SILVA FILHO, Danilo F. quantificação de fibra alimentar em algumas populações de cubiu (*Dolanum sessiliflorum* Dunal), camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh) e açaí (*Euterpe oleracea* Mart.), *Acta Amazonica*, v. 32, n.3, pp. 491-497, 2002.
- ZAPATA, Sérgio M.; DUFOUR, Jean-Pierre. Camu-camu *Myrciaria dubia* (H.B.K.) McVaugh: chemical composition of fruit. *J. Sci. Food Agric.*, v.61, pp. 349-351, 1995.
- MIZUNO, K., KATO, N., MAKINO, M., SUZUKI, T., SHINDO, M. Continuous inhibition of excessive polyol pathway flux in peripheral nerves by aldose reductase inhibitor fidarestat leads to improvement of diabetic neuropathy. *J. Diabet Complications*, v. 13, pp. 141-150, 1999.

Capítulo 1

CONSTRUÇÃO DA BIBLIOTECA DE EST (*EXPRESSED SEQUENCE TAGs*) DE FRUTOS DE CAMU-CAMU DE PLANTAS CULTIVADAS EM TERRA-FIRME

RESUMO

Para o estudo de genoma funcional (genes expressos) em frutos de camu-camu é necessária a construção de Bibliotecas de cDNA. O presente trabalho teve como objetivo desenvolver um procedimento para a extração de RNA total de frutos de camu-camu para, subsequentemente, construir e seqüenciar biblioteca de cDNAs de frutos de camu-camu de plantas cultivadas em terra-firme. Para a extração do RNA total foi modificado o protocolo de Asif et al. (2000). O RNA foi extraída da casca-polpa de frutos. O mRNA poliadenilado foi purificado utilizando o "*Kit Message Maker mRNA Isolation System* (GIBCO)". A síntese e clonagem de cDNA foi conduzida segundo o "*Kit SuperScript Plasmid System with Gateway Synthesis and Cloning* (Invitrogen)". O seqüenciamento dos cDNAs foi realizado no Laboratório de Tecnologia do DNA do Centro de Apoio Multidisciplinar da Universidade Federal do Amazonas (UFAM), e na Plataforma de Seqüenciamento da EMBRAPA – CENARGEN. Foi desenvolvido um procedimento eficiente para extração de RNA total de frutos de camu-camu, método que tem como componente base do tampão de extração CTAB e β -mercaptoetanol, e que apresenta também, uma segunda limpeza com SDS 0,5%. Este método possibilitou a obtenção de mRNAs de qualidade, utilizados na confecção de cDNAs de tamanhos variados (500pb a 4Kb), armazenados em plasmídio pSPORT1 e transformados em *E. coli* XL1Blue. O seqüenciamento da extremidade 5' de cDNA com tamanho entre 500pb a 1000pb possibilitou a formação da primeira biblioteca de ESTs de frutos de camu-camu, até o presente momento, com cerca de 3000 seqüências.

ABSTRACT

For the functional genome study (express genes) in camu-camu fruits is necessary of Libraries of cDNA construction. The present work had like aim develop a procedure for the RNA Total extraction of camu-camu fruits for, subsequently, build and sequencing library of cDNAs of camu-camu fruits of cultivated plants in *terra-firme*. For the RNA Total extraction of the was modified the protocol of Asif et al. (2000). The RNA was extracted of the frutis shell-pulp. The mRNA poliadenilado was purified utilizing the Message Maker mRNA Isolation System (GIBCO). The synthesis and cloning of cDNA was driven second the SuperScript Plasmid System with Gateway Synthesis and Cloning (Invitrogen). The cDNA sequencing was carried out in the Technology of the DNA Laboratory of the Federal University of the Amazon (UFAM), and in the Sequencing Platform of the EMBRAPA – CENARGEN. It was developed an efficient procedure for RNA total extraction of camu-camu fruits, approach that has like component base of the plug of CTAB and β -mercaptoetanol extraction, and that presents also, a second cleaning with 0.5% SDS. This approach enabled the obtaining of mRNAs of quality, utilized in the making of sizes varied cDNA (500pb to 4Kb), stored in pSPORT1 plasmid and *E. coli* XL1Blue transformed. The sequecing of the 5' extremity of cDNA with size between 500pb to 1000pb enabled the formation of the first ESTs library of camu-camu fruits, to the present moment, with around 3000 sequences.

1. INTRODUÇÃO

O camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh), espécie Amazônica com grande potencial econômico, apresenta fruto rico em ácido ascórbico (Vitamina C), com variabilidade ampla entre 0,845 a 6,1g de ácido ascórbico/100g de polpa (CALZADA BENZA, 1980; ANDRADE et al., 1991; ZAPATA & DUFOUR, 1995; JUSTI et al., 2000; YUYAMA, K. et al., 2002). O camu-camu é fonte potencial de fibra alimentar total, insolúvel e solúvel numa concentração média de 2,88g (YUYAMA, L. et al., 2002, p. 491), cuja casca apresenta antocianina, importante corante natural (ANDRADE, 1993) (Figura 1).



Figura 1 – Aspectos gerais de frutos de camu-camu

Fonte: <http://www.inpa.gov.br/cpca/areas/camu-camu.html>

Para o estudo de genoma funcional (genes expressos) faz-se necessária a construção de Bibliotecas de cDNA que contém as seqüências codificantes dos genes que são expressos em um dado organismo ou mesmo em uma dada célula ou tecido.

Para o estudo da expressão de genes é necessário, primeiramente, a obtenção de RNA total, este é composto pelo RNA mensageiro (mRNA) - que representa de 1 à 5% do RNA total em células eucarióticas, pelo RNA ribossomal (rRNA), pelo RNA de transferência (tRNA) e por outros pequenos RNAs (STERKY & LUNDEBERG, 2000).

A primeira dificuldade na obtenção do RNA é que eles são muito sensíveis a degradação por ribonucleases endógenas e exógenas (RNAses), por isso todo material deve ser cuidadosamente limpo antes do uso, sendo necessário o uso de inibidores de RNAses (STERKY & LUNDEBERG, 2000).

Existem no mercado "kits" específicos para a extração do RNA total, entretanto, alguns tecidos vegetais necessitam de protocolo de extração específico, para melhor eficiência na extração do RNA. No caso de frutos, vários compostos orgânicos podem dificultar a extração do RNA afetando sua quantidade e/ou qualidade, como polissacarídios, polifenóis (ASIF et al., 2000; GASIC et al., 2004), terpenos (WU et al., 2002), flavonóides e outros metabólitos secundários (GAO et al., 2000 e 2001). Tao et al. (2004) indicam o conteúdo de água e outros compostos na polpa de frutos maduros como agentes que podem dificultar a extração do RNA.

Os mRNAs maduros possuem uma cauda com aproximadamente 40-200 adeninas consecutivas, importantes para o seu isolamento e para sua conversão em cDNA, para o qual são utilizadas algumas enzimas, como a transcriptase reversa, utilizada para a síntese da primeira fita; a Rnase H, que degrada a fita de RNA do híbrido DNA-RNA; a DNA polimerase I que sintetiza a segunda fita de DNA, formando o cDNA dupla fita (STERKY & LUNDEBERG, 2000).

Para iniciar a síntese da primeira fita de cDNA, é necessário um "primer" oligo (dT) que se anela a cauda poli (A+); podem ser usados também, primers randômicos hexâmeros ou primers específicos para um determinado gene (STERKY & LUNDEBERG, 2000).

Atualmente, um dos métodos de síntese de cDNA para clonagem direcional dos insertos no vetor, tem sido feita incluindo um sítio de restrição (*Not* I) no primer oligo (dT), usado na síntese da primeira fita do cDNA. Após a síntese da segunda fita um "adaptador" (*Sal* I ou freqüentemente *Eco* RI) é ligado em ambas extremidades do fragmento, posteriormente é realizada digestão dupla com as enzimas de restrição *Not* I e *Sal* I para tornar a extremidade coesivas facilitando a ligação ao vetor (INVITROGEN). Os fragmentos de cDNA de fita dupla são então clonados num vetor apropriado criando uma população de clones chamadas de Bibliotecas de cDNAs.

A princípio a posição das extremidade 5'e 3' do cDNA no vetor é conhecida, em função disto, as extremidade dos vetores podem ser utilizadas como iniciadores (*primers*) para o seqüenciamento de curtas seqüências de cDNAs, com cerca de 100pb-300pb, chamadas de EST (*Expressed Sequence Tag*) (FRISTENSKY et al., 1999). Os ESTs são, portanto, seqüências parciais de cDNAs, reversamente transcritos a partir de um mRNA e representam uma seqüência gênica codificada, livre de *introns*.

A possibilidade de produção orientada de ESTs, e do seqüenciamento das duas extremidades do cDNA, pode fornecer a seqüência completa de um gene (JONGENEEL, 2000). Em muitos projetos genoma, o seqüenciamento dos EST tem sido bem sucedido. Estes contêm seqüências parciais de muitos mRNAs presentes em vários tecidos usados na construção de bibliotecas de cDNA (ADAMS et al., 1992).

O trabalho aqui descrito teve como objetivo desenvolver um procedimento para a extração de RNA total da casca-polpa de frutos de camu-camu com, subsequentemente, construção e seqüenciamento de uma biblioteca de cDNAs de frutos de camu-camu de plantas cultivadas em terra-firme.

2. MATERIAIS E MÉTODOS

2.1. Material Biológico

Os frutos foram coletados de plantas de camu-camu, cultivadas em terra-firme na Fazenda Yurican (Rio Preto da Eva, AM) provenientes de germoplasma oriundos de Iquitos, Peru. Essa coleção foi construída e se encontra sob coordenação do Instituto Nacional de Pesquisas da Amazônia – INPA (Manaus, AM).

As etapas que se seguem foram realizadas no Laboratório de Nutrigenômica do Centro Nacional de Recursos Genéticos - CENARGEN da Empresa Brasileira de Pesquisa Agropecuária EMBRAPA (Brasília, DF). Após coletados, os frutos foram congelados inteiros, em gelo seco, e levados ao laboratório onde foi feita a separação dos componentes: casca-polpa e sementes (Figura 02). A casca-polpa foi utilizada para a extração de RNA total.



Figura 2 - Aspecto da polpa e casca de frutos de camu-camu

2.2. Extração do RNA total

Não há descrição de um protocolo de extração de RNA total para frutos de camu-camu, vários métodos convencionais de extração de RNA foram testados (método Trizol; método fenol-clorofórmio/cloreto de lítio; método fenol quente), entretanto, sem nenhum

sucesso, provavelmente em função da grande quantidade de polissacarídeo presente no fruto (ANDRADE et al., 1995).

Buscou-se na literatura protocolos que possibilitassem a extração de RNA total de frutos com elevado teor de polissacarídios. Testou-se então o protocolo descrito por Asif et al. (2000), que embora tenha possibilitado a extração do RNA total de frutos de camu-camu este se apresentava com pouca quantidade e baixa qualidade e ainda um significativo teor de contaminação com polissacarídeos. Este protocolo foi modificado para melhorar a qualidade do RNA extraído.

As soluções e reagentes utilizados foram os seguintes:

1. Tampão de Extração (TENC): 200 mM Tris-HCl pH 8,2; 1,4 M NaCl; 20 mM EDTA (pH 8,0); e 2% CTAB (*hexadecyltrimethylammonium bromide*)
2. β -mercaptoetanol
3. Cloreto de Lítio 10M
4. Acetato de Sódio 3M pH 5,2
5. Água MiliQ autoclavada e tratada com DEPC
6. Etanol Absoluto e Etanol 70%
7. Fenol:Clorofórmio (1:1)(clorofane)
8. Clorofórmio:alcool isoamílico (24:1) (clorofil)
9. tRNA de levedura (GIBCO)
10. SDS 0,5% (dodecil sulfato de sódio)

PROTÓCOLO

1. Usar 10 mL do tampão TENC por grama de tecido. Adicionar 1 μ L de β -mercaptoetanol por mL de Tampão antes de usar;

2. Ao tecido congelado adicionar o TENC pré-aquecido a 65 °C e homogeneizar 2X por 30 s. Usar tubo corex de 150 mL (5g de tecido - 50 mL de TENC);
3. Incubar a 65 °C por 1 hora (banho-maria). Vortexar fortemente a cada 15 min.;
4. Adicionar igual volume de clorofórmio: álcool isoamílico (24:1). Misturar vigorosamente até as duas fases formarem uma emulsão. Abrir o tubo durante a mistura para liberar os vapores;
5. Centrifugar 9000 rpm por 25 min. a temperatura ambiente (T.A.) rotor GSA sorvall;
6. Coletar a fase aquosa e re-extrair com igual volume de clorofórmio: álcool isoamílico (24:1);
7. Centrifugar 9000 rpm por 25 min. a T. A., rotor GSA sorvall;
8. Transferir a fase aquosa para outro tubo e adicionar 10 M LiCl para uma concentração final de 3 M. Adicionar a solução 40 µg/mL tRNA levedura. Armazenar a 4 °C durante a noite;
9. Transferir e distribuir a solução para tubos corex 30 mL e remover o RNA por centrifugação a 9000 rpm de 40 a 60 min a 4 °C, rotor SS34 sorvall;
10. Retirar o excesso da solução, permanecendo apenas 500 µL para dissolver o sedimento e transferi-lo para tubos de 1,5 mL (eppendorf).
11. Centrifugar a 13000 rpm por 10 min. a 4 °C;
12. Descartar o sobrenadante e dissolver o pellet em 100 µL de SDS 0,5%;
13. Adicionar igual volume de fenol:clorofórmio (1:1), vortexar vigorosamente;
14. Centrifugar a 13000 rpm por 10 min. a 4 °C;
15. Transferir a fase aquosa para outro tubo eppendorf e adicionar igual volume de clorofórmio: álcool isoamílico (24:1), vortexar vigorosamente;
16. Centrifugar a 13000 rpm por 10 min. a 4 °C;
17. Transferir a fase aquosa para outro tubo eppendorf e adicionar acetato de sódio 3 M pH 5.2 para a concentração final de 0,3 M e 3 volumes de etanol 100%;

18. Manter o RNA precipitando a $-70\text{ }^{\circ}\text{C}$ por 3 horas ou durante a noite a $-20\text{ }^{\circ}\text{C}$;
19. Centrifugar a 13000 rpm por 20 min. a $4\text{ }^{\circ}\text{C}$;
20. Lavar o sedimento com etanol 70% (a $-20\text{ }^{\circ}\text{C}$) e secar a T. A.;
21. Ressuspender o sedimento em 20 à 40 μL de H_2O -DEPC autoclavada.

O RNA total foi observado em gel de agarose 2% desnaturante contendo formaldeído e tampão MOPS e posteriormente submetido à purificação de mRNA e confecção de cDNA para estudo dos genes expressos em camu-camu.

2.3. Construção da biblioteca de cDNAs

A biblioteca foi construída no Laboratório de Nutrigenômica EMBRAPA/CENARGEN (Brasília, DF). O mRNA poliadenilado foi purificado utilizando o "*Kit Message Maker mRNA Isolation System* (GIBCO)" seguindo as instruções contidas no "Kit".

A síntese e clonagem de cDNA foi conduzida segundo o "*Kit SuperScript Plasmid System with Gateway Synthesis and Cloning* (Invitrogen)", resumido na figura 03. Para a síntese do cDNA foram necessárias $2\mu\text{g}$ de mRNA. Um primer-adaptador oligodT-Not I é utilizado para síntese da primeira fita de cDNA, este se anela à cauda poli(A) do mRNA, esse tipo de adaptador é utilizado para direcionar a futura ligação do cDNA ao plasmídeo.

Após a síntese da primeira fita de cDNA a enzima RNase H degrada o RNA do híbrido RNA-cDNA, a seguir a DNA polimerase sintetiza a segunda fita de cDNA usando a primeira como molde e utilizando pedaços do mRNA como iniciadores num processo análogo ao "*nick translation*".

Para aumentar a eficiência de ligação do inserto ao vetor as extremidades *blunt end* do cDNA são convertidas a extremidades colantes (*stick end*), pela adição de

adaptadores *Sal I* pela enzima T4 DNA ligase. Após a digestão com *Not I*, os cDNAs foram fracionados por cromatografia em coluna para retirada de resíduos de adaptadores que pudessem impedir a ligação do cDNA ao vetor (Figura 4) e reduzir a quantidade de pequenos insertos (< 500pb).

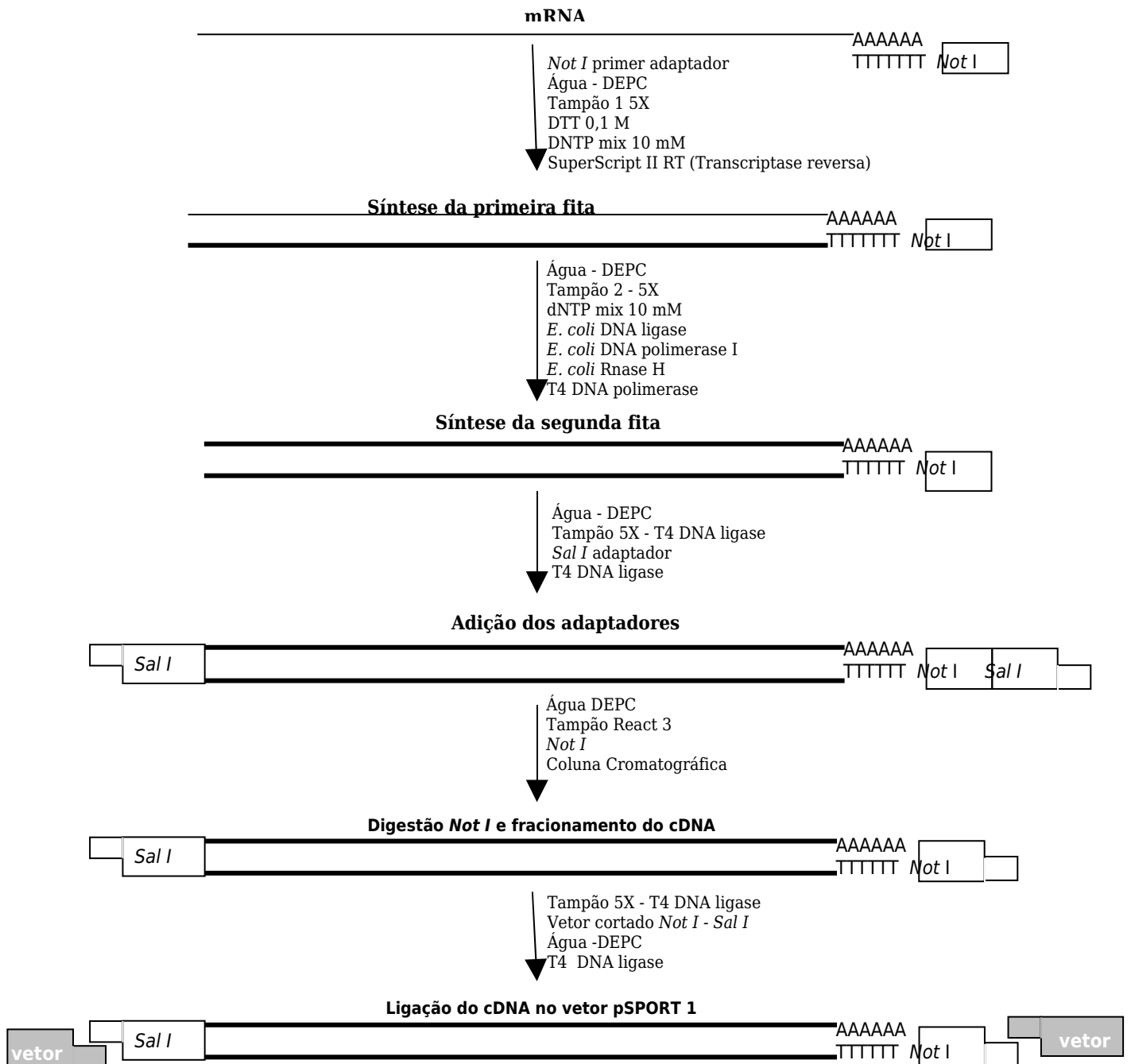


Figura 3 - Esquema do procedimento de síntese de DNA e ligação do vetor pSPORT 1

Os cDNAs obtidos foram separados de acordo com o tamanho (aproximadamente de 500pb a 1000pb; 1000pb a 2000pb; e 2000pb a 3000pb) formando três "pools" que foram então ligados, cada um, ao plasmídeo pSPORT 1, previamente digerido com *Sal* I e *Not* I (Figura 4).

Os sistemas de ligação foram utilizados para transformar *E. coli* XL1-blue por eletroporação (voltagem 1900 e resistência 2500). As colônias transformadas foram preservadas em meio LB com glicerol 16% e armazenadas em freezer -70° C.

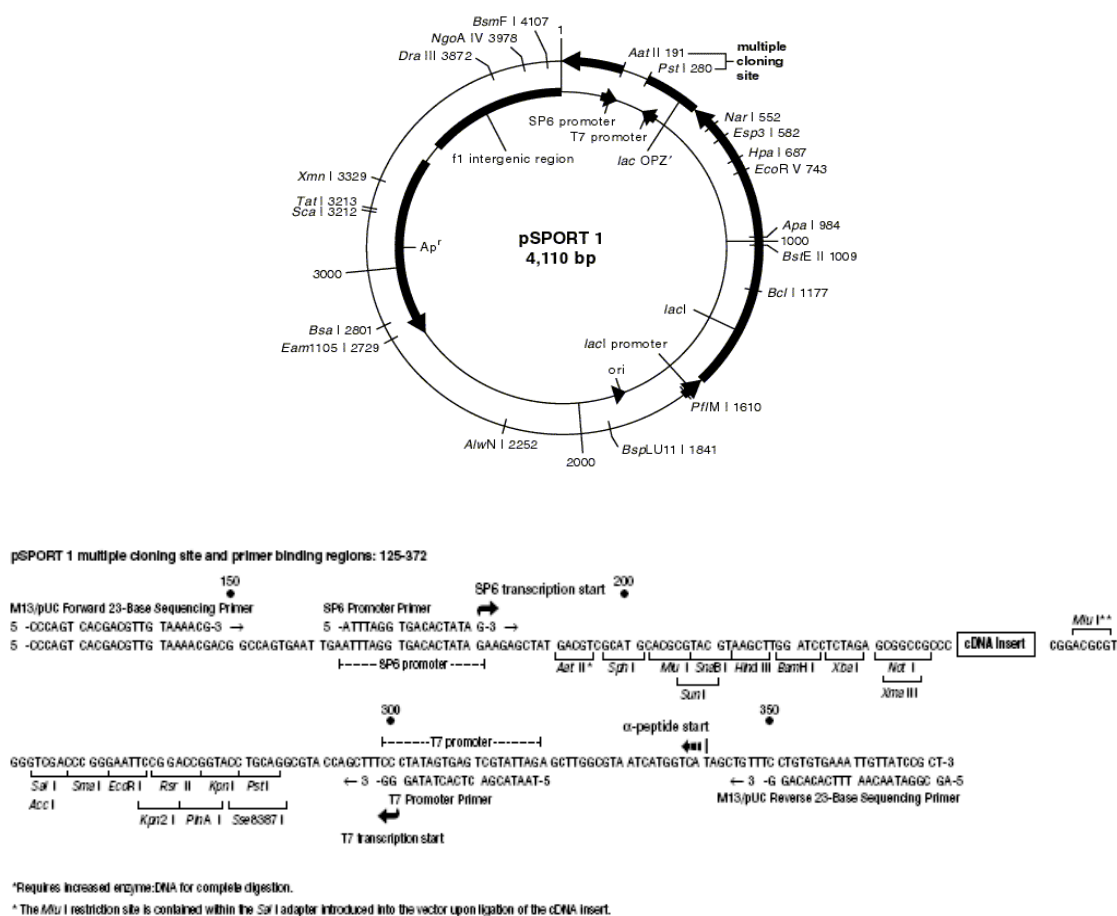


Figura 4 - Mapa do plasmídeo pSPORT 1, as setas indicam o local de inserção dos cDNAs
 Fonte: <http://www.invitrogen.com/content/sfs/manuals/18248.pdf>

A validação da biblioteca cDNA foi feita por meio da extração de plasmídeo (lise alcalina) e reação de PCR de algumas colônias do *pool* 3 (500pb e 100pb) para confirmação da presença e do tamanho dos insertos.

2.4. Seqüenciamento dos cDNAs

O seqüenciamento dos cDNAs foi realizado tanto no Laboratório de Tecnologia do DNA do Centro de Apoio Multidisciplinar da Universidade Federal do Amazonas (UFAM), como na Plataforma de Seqüenciamento da EMBRAPA – Recursos Genéticos e Biotecnologia (CENARGEN).

Na UFAM, para o seqüenciamento, primeiramente, isolou-se os plasmídios, em seguida foram retiradas alíquotas de 1µL de cada clone recombinante preservado a -70° C e adicionados em 1,5 mL de meio LB líquido em placas do tipo "deep well", com ampicilina (100 µg/mL). As culturas foram incubadas com agitação (37 °C à 150 rpm) por cerca de 22 horas. Os plasmídios foram isolados e purificados pelo método de lise alcalina (mini preparação) (SAMBROOK et al., 1989). A análise dos plasmídios isolados foi feita em eletroforese e gel de agarose 0,8%.

Para a reação de seqüenciamento foi utilizado o "Kit DYEnamic ET Dye Terminator" e o "*primer T7" reverse* (3' CTAATACGACTCACTATAGGG 5'), realizada em termociclador. O programa utilizado para a reação de seqüenciamento foi: 95° C - 15 seg.; 95° C; 55° C - 20 seg.; 60° C - 1 min.; 4° C – infinito com 35 ciclos.

O seqüenciamento foi realizado no MegaBACE 1000, seqüenciador automático da *Amersham Biosciences*, que utiliza eletroforese capilar para separar fragmentos de DNA. Na EMBRAPA–Recursos Genéticos e Biotecnologia (Cenargen) foram preparadas placas contendo os clones transformantes isolados, que em seguida foram encaminhados a Plataforma de seqüenciamento, que executou todas as etapas descritas nas páginas anteriores. O *primer* utilizado foi o "*M13 -20 reverse*" (3'CGACGTTGTAAAACGACGG CCAGT5') e o seqüenciador ABI3700 automático da *Applied Biosystems*.

3. RESULTADOS E DISCUSSÃO

Quanto ao desenvolvimento de um procedimento para a extração de RNA total de frutos de camu-camu, as principais alterações realizadas no protocolo foram as seguintes: acrescentou-se o uso do tRNA levedura e SDS 0,5%, a concentração do Tris-HCl do tampão de extração foi aumentada para 200mM e foi excluído o uso de fenol hidratado, mantendo-se apenas o fenol:clorofórmio após o SDS 0,5% (ASIF et al., 2000).

Para obtenção do RNA total foi necessária à utilização de no mínimo 5g de polpa-casca, que corresponde ao peso de um fruto sem semente (Figura 01). Verificou-se que o CTAB adicionado de β -mercaptoetanol foram eficientes na separação de grande parte dos polissacarídeos. O aumento da concentração de Tris-HCl, no TENC, visou equilibrar o pH do sistema de extração em função da elevada acidez do fruto (JUSTI et al., 2000). Segundo Asif et al. (2000), o uso do EDTA 20 mM e do β -mercaptoetanol no tampão de extração também auxilia na remoção de polifenóis.

A adição do tRNA de levedura ao cloreto de lítio contribuiu na precipitação do RNA total, pelo aumento de concentração do mesmo. A utilização do SDS 0,5% foi necessária para eliminação de resíduos de polissacarídios e proteínas ainda existentes no RNA precipitado e conseqüentemente obtendo-se um maior grau de purificação.

As precipitações com feno/clorofórmio (1:1) e posteriormente com clorofórmio:alcool isoamílico foram realizadas para extrair as proteínas e também para retirar resíduos de fenol e do SDS do RNA precipitado para que os mesmos não interferissem na síntese dos cDNAs.

Para obtenção da quantidade de mRNA necessária à síntese de cDNA foram extraídos RNA total de cerca de 6 frutos de camu-camu (em estágio inicial de maturação e maduros), totalizando entorno de 30g de casca-polpa. O RNA total isolado

era translúcido e quando visualizado em gel de agarose com formaldeído observou-se as bandas do RNAs ribossômicos intactas e claramente visíveis, indicando que o RNA não estava degradado (Figura 05).

O tempo médio de extração de RNA total com este protocolo foi de 48 horas, relativamente longo, principalmente em função do tempo de precipitação do RNA no cloreto de lítio e em acetato de sódio.

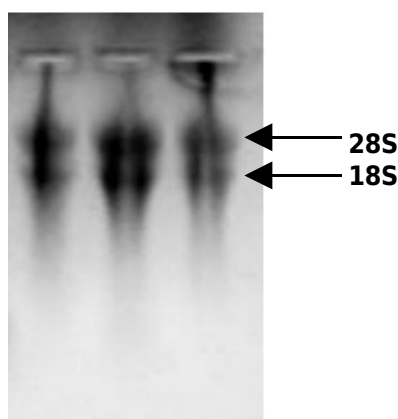


Figura 05 – Perfil eletroforético do RNA Total extraído da polpa-casca de três diferentes frutos de camu-camu visualizado em gel de agarose 2% desnaturante com MOPS e formaldeído.

Aproximadamente 2 μg de mRNA foram usados para a síntese do cDNA dupla fita. O equivalente a 10 $\eta\text{g}/\mu\text{L}$ de insertos e 1 μg de pSPORT1 foram utilizados no sistema de ligação (volume final de 50 μL) para produção da biblioteca, para transformação de *E. coli* XL1-Blue utilizou-se 1 μL do sistema de ligação.

Todo processo de síntese de cDNA foi feita sem marcação radioativa, diferente do proposto pelo fabricante do kit, dessa forma para confirmação da síntese do cDNA avaliou-se o perfil eletroforético das frações obtidas (Figura 06).

A biblioteca produzida apresentou excelente qualidade e cDNAs de tamanhos variados, com insertos entre 500 pb a 4kb (Figura 06). O título da biblioteca de cDNA

foi estimado em 10^6 clones/ μg de cDNA, que equivale a um número de 5000 clones recombinantes independentes por 10^{-5} μg de cDNA.

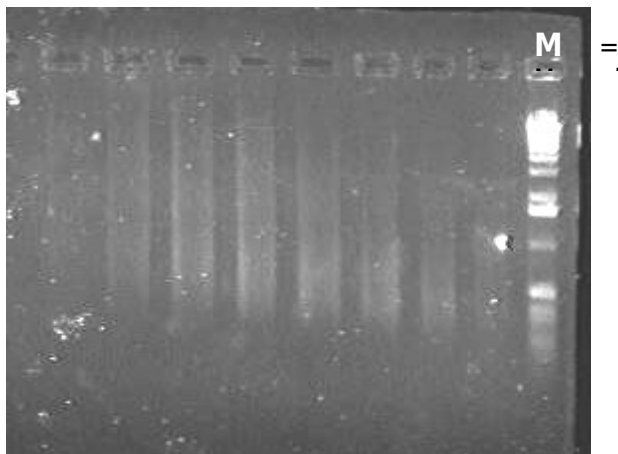


Figura 06 – Perfil eletroforético das frações do cDNA obtido a partir de polpa e casca de camu-camu. M= marcador.

Para verificação do processo de clonagem dos fragmentos e confirmação dos tamanhos relativos dos insertos clonados, alguns clones foram utilizados no processo de validação da biblioteca de cDNA, que permitiu avaliar a qualidade da biblioteca construída antes do seqüenciamento dos clones.

Nesse processo, algumas colônias do *pool* 3 (500pb a 1000pb) foram analisadas e confirmou-se por meio do perfil eletroforético dos plasmídios o sucesso no processo de transformação (Figura 7) e a partir da análise dos produtos da PCR, com os plasmídios isolados, que os insertos apresentam tamanhos variados dentro da faixa esperada (500pb e 1000pb) (Figura 08).

Verificou-se que alguns clones apresentaram mais de uma banda amplificada, isso se deve provavelmente ao fato de ter entrado mais de um tipo de plasmídio na célula competente (*E. coli* XL-Blue). Esses tipos de clones podem gerar problema no seqüenciamento causando, por exemplo, *no hits* (seqüências sem similaridades nos banco de genes) e portanto, não confiáveis, necessitando, primeiramente a separação dos dois insertos dos clones.

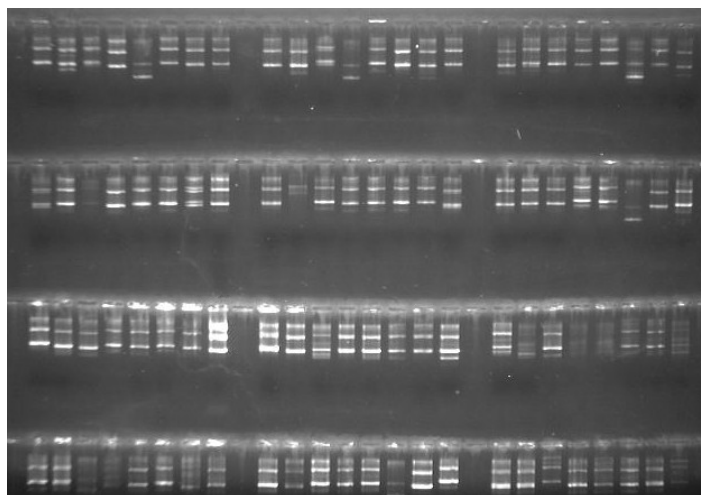


Figura 07 – Perfil eletroforético dos plasmídios extraídos a partir de clones transformados com sistema de ligação contendo cDNA de frutos de camu-camu.

A partir do seqüenciamento da extremidade 5' de 2000 insertos na UFAM e 3000 insertos na Plataforma de seqüenciamento do CENARGEN de cDNAs de frutos de camu-camu obteve-se cerca de 64% de seqüências com boa qualidade, percentual satisfatório em função de possíveis problemas na reação de seqüenciamento, principalmente nos clones com mais de um inserto.

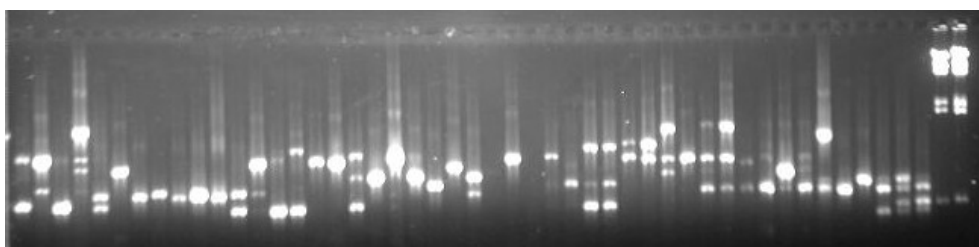


Figura 08 – Perfil eletroforético de produtos de PCR a partir de plasmídios contendo insertos de cDNA de frutos de camu-camu (Marcador de 1Kd)

O percentual de seqüências com boa qualidade correspondeu a 3196 seqüências ESTs validadas e armazenadas no Programa Sistema Genoma (<http://condor.genoma.cenargen.embrapa.br/genoma/classes/JFrameGenoma.jnlp>), sendo criada a primeira biblioteca de ESTs de frutos de camu-camu, chamada de *CAMUEST*, posteriormente submetida a anotação funcional.

4. CONCLUSÕES

- Foi desenvolvido o procedimento eficiente para extração de RNA total de frutos de camu-camu, método que tem como componente base do tampão de extração CTAB e β -mercaptoetanol e uma segunda limpeza com SDS 0,5%.
- Este método possibilitou a obtenção de mRNAs de qualidade, utilizados na confecção de cDNAs de tamanhos variados (500pb a 4Kb), armazenados em plasmídio pSPORT1 e transformados em *E. coli* XL1Blue.
- O seqüenciamento da extremidade 5' de cDNA com tamanho entre 500pb a 1000pb possibilitou a formação da primeira biblioteca de ESTs de frutos de camu-camu, até o presente momento, com cerca de 3000 seqüências armazenadas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMS, M. D. DUBNICK, M.; KERLAVAGE, A. R.; MORENO, R.; FIELDS, C.; VERTER, J. C. Sequence identification de 2,375 human brain genes (see comments) *Nature* v. 355, pp. 632-4, 1992.
- ANDRADE, Jerusa. S. Uso de atmosfera modificada e refrigeração para manutenção da qualidade pós-colheita do camu-camu (*Myrciaria dubia* (H.B.K.)McVaugh), 50p., 1993. (Tese)
- ANDRADE, Jerusa S.; GALEZI, Maria A. M.; ARAGÃO, Carlos G.; CHÁVEZ-FLORES, Wanders B. Valor nutricional do camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh) cultivado em terra firme da Amazônia Central. *Rev. Bras. Frutic.* Cruz das Almas. v.13, n.3, pp. 307-311, 1991.
- ANDRADE, Jerusa C.; ARAGÃO, Carlos G.; GALEAZZI, Maria A. M.; FERREIRA, Sidney A. N. Changes in the concentration of total vitamin C during maturation and ripening of camu-camu (*Myrciaria dubia* (H.B.K.) Mc Vaguh) fruits cultivated in the upland of brasilian central amazon. *Acta Horticulture* v. 370, pp. 177-180, 1995.
- ASIF, Mehar H.; DHAWAN, Puneet; NATH, Pravendra. A simple procedure for the isolation of high quality RNA from ripening banana fruit., *Plant Mol. Biol. Repr.*, v. 18, pp. 109-115, 2000.
- CALZADA BENZA, J.C. *143 Frutales nativos*. Librería El Estudiante. La Molina, 314p., 1980.
- FRISTENSKY, Brian; BALCERZAK, Margaret; HE, Daiffen; ZHANG, Peijun. Expressed sequence tags from the defense response of ***Brassica napus*** to ***Leptosphaeria maculans***. *Mol.Plant Pathol.On-Line* [<http://www.bspp.org.uk/mpol/1999/0301FRISTENSKY>] acesso em 01/07/2002.
- GAO, J.W.; LIU, J.Z.; LI, B.; FENG, B.S.; YU, G.Q.; LI, Z.S. Preliminary study on pigments in aleurone layer of blue-grained wheat. *Acta Biotanica Boreali-occidentalia*, v. 20, n.6, pp. 936-941, 2000.
- GAO, Jianwei; LIU, Jianzhong; LI, Bin; LI, Zhensheng. Isolation and purification of functional Total RNA from blue-grained wheat endosperm tissues containing high levels of starches and flavonoids. *Plant Mol. Biol. Reporter*, v. 19, pp. 185a-185i, 2001
- GASIC, Ksenija; HERNANDEZ, Alvaro; KORBAN, Schuyler. S. RNA Extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Plant Mol. Biol. Reporter* v. 22, pp. 437a-437g, 2004.
- INVITROGEN Life Technologies. SuperScript Plasmid System with Gateway Technology for cDNA Synthesis and cloning, pp.38. (Cat. No. 18248047)
- JUSTI, Karin C.; VISENTAINER, Jesui V.; SOUZA, Nilson E.; MATSUSHITA, M. Nutritional composition and vitamin C stability in stored camu-camu (*Myrciaria dubia*) pulp. *Archivos Latinoamericanos de Nutricion*, v. 50, n. 4, pp. 405-408, 2000.
- SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor, v.3, 1989.
- STERKY, Fredrik.; LUNDEBERG, Joakin. Sequence analysis of genes and genomes. *Journal of Biotechnology*, v.76, pp. 1-31, 2000
- TAO, Nengguo; CHENG, Yunjiang; XU, Juan; XU, Qiang; DENG, Xiuxin. Na effective protocol for the isolation of RNA from the pulp of ripening citrus fruits. *Plant Mol. Biol. Repr.*, v.22, pp. 305a-f, 2004.
- WU, Yingru; LLEWELLYN, Danny J.; DENNIS, Elizabeth S. A quick and easy method for isolating good-quality RNA from cotton (*Gossypium hirsutum* L.) tissues. *Plant Mol. Biol. Reporter*, v. 20, pp. 213-218, 2002.
- YUYAMA, Kaoru.; AGUIAR, Jaime P. L.; YUYAMA, Lucia K. O. Camu-camu: um fruto fantástico como de vitamina C. *Acta Amazonica*, v. 32, n. 1, p. 169-174, 2002.
- YUYAMA, Lúia. K. O.; BARROS, S. E.; AGUIAR, Jaim. P. L.; YUYAMA, Kaoru; SILVA FILHO, Danilo F. quantificação de fibra alimentar em algumas populações de cubiu (*Dolanum sessiliflorum* Dunal), camu-

camu (*Myrciaria dubia* (H.B.K.) McVaugh) e açaí (*Euterpe oleracea* Mart.), *Acta Amazonica* , v. 32, n.3, pp. 491-497, 2002

ZAPATA, Sergio M.; DUFOUR, Jean-Pierre. Camu-camu *Myrciaria dubia* (H.B.K.) McVaugh: chemical composition of fruit. *J. Sci. Food Agric.*, v.61, p. 349-351, 1995.

Capítulo 2

ANOTAÇÃO FUNCIONAL DA BIBLIOTECA DE ESTs DE FRUTOS DE CAMU-CAMU - *CAMUEST*

RESUMO

A partir do seqüenciamento de 5000 ESTs da extremidade 5' de insertos de cDNA da biblioteca de frutos de camu-camu, 3196 ESTs foram validados no programa Sistema Genoma, sendo formados 1546 *singletons* e 358 *contigs*, resultando num total de 2586 seqüências ESTs com similaridade a seqüências encontradas no banco de genes. A análise da clusterização da biblioteca revelou um índice de 81% de novidade e 32,54% de redundância para a biblioteca de cDNA. Cerca de 90% dos *contigs* apresentaram baixa redundância (2-4 *reads* por *contigs*). As espécies com seqüências com similaridade às de camu-camu foram *Arabidopsis thaliana* (49%), *Oryza sativa* (20%) e *Glycine max* (4%). Dentre as categorias de proteínas com função conhecida destacaram-se as seguintes: modificação pós-traducional, proteína *turnover*, chaperonas (13,2%), metabolismo e transporte de carboidratos (8,8%) e tradução, biogênese e estrutura ribossomal (5,7%). Foram observados 3% de seqüências (97 ESTs) com baixa similaridade ($e\text{-value} > e^{-10}$) e 15% não apresentaram similaridade com seqüências contida no banco de genes. Dentre as seqüências ESTs com similaridade no *GenBank*, 348 foram classificados como proteínas desconhecidas. As seqüências ESTs de frutos de camu-camu com maior similaridade ($e\text{-value} 0.0$) foram relacionadas as proteínas: Proteossoma 26S subunidade 4; Alcool deidrogenase 2; Alfa tubulina 1; catalase; Pirofosfatase H⁺; duas proteínas de choque térmico 83; 70KDa proteína de choque térmico 1. Os genes abundantemente expressos em frutos de camu-camu e seus respectivos níveis de expressão foram os seguintes: glutathiona S-transferase (1,04%); Alergênica maior *Mal d 1.06A02* (0,73 %); Proteína de choque térmico cognata 70 kDa 1 (0,58%); proteína como extensina (0,58%); fator de alongação da tradução 1A-9 (0,50%).

ABSTRACT

From of 5000 ESTs sequencing of 5' extremity 5' cDNA insert of camu-camu fruits library, 3196 ESTs were validated in the Genome System program, being formed 1546 singletons and 358 contigs, resulting in a total of 2586 ESTs sequences with similarity the gene bank sequences. The clursterization analysis of the library revealed an index of novelty 81% and redundancy 32,54% of cDNA library. Around 90% of the contigs presented decrease redundancy (2-4 reads by contigs). The species with sequences with similarity to the of camu-camu were *Arabidopsis thaliana* (49%), *Oryza sativa* (20%) and *Glycine max* (4%). Among the protein categories with known function detached itself the following: posttranslational modification, protein turnover, chaperones (13,2%), carbohydrate transport and metabolism (8,8%) and translation, ribosomal structure and biogenesis (5,7%). They were observed 3% of sequences (97 ESTs) with decrease similarity (and-value> and-10) and 15% did not they present similarity with gen bank sequences. Among the sequences ESTs with similarity in the GenBank, 348 were classified like unknown proteins. The ESTs sequences of camu-camu fruits with bigger similarity (and-value 0.0) were related the proteins: 26S proteosoma subunity 4; alcohol dehydrogenase 2; tubulin Alpha 1; catalase; H+ pyrophosphatase; two heat shock proteins 83; heat shock protein 70KDa 1. The abundantly expresss genes in camu-camu fruits and his respective levels of expression were the following: glutathione S-transferase (1,04%); Major alergenec Mad d 1.06A02 (0,73%); heat chock cognate 70 kDa protein 1 (0,58%); extensina like protein (0,58%); translation elongation factor 1A-9 (0,50%).

1. INTRODUÇÃO

A biologia molecular defende que os genes, codificados no DNA, são transcritos em mRNA e traduzidos em proteínas funcionais. Em vista disso, a expressão dos genes passou a ser estudada pelo isolamento de mRNA, posteriormente transcritos em cDNA (DNA complementar). Estes representam fragmentos de genes individuais que podem ser isolados, clonados e sequenciados (GERHOLD & CASKEY, 1996, p. ; FRISTENSKY et al., 1999). Na década de 80, o advento do seqüenciamento automático facilitou a obtenção de várias seqüências de cerca de 100pb a 300pb das extremidades 3' e 5' de muitos clones de cDNA através dos projetos genomas. Estas curtas seqüências de DNA, chamadas de EST (*Expressed Sequence Tag*), tem sido uma ótima ferramenta na caracterização de genes (FRISTENSKY et al., 1999).

Com a obtenção destas seqüências ESTs a pergunta imediata que surge é: qual seria a função celular destas seqüências nucleotídicas? A resposta pode ser obtida através da utilização de programas computacionais que são capazes de decifrar os pares de bases de um fragmento genômico.

Em função deste fato, a análise computacional destas seqüências tornou-se um exercício de suma importância para viabilizar a extração das informações pertinentes e realmente poder concretizar as potencialidades oriundas de se possuir o genoma completo de um organismo. Considerando sua relevância surgiu um novo ramo da ciência biológica voltado para o tratamento deste problema: a Bioinformática (LUSCOMBE et al., 2001).

A Bioinformática consiste no uso de sistemas computacionais na análise e caracterização molecular (DNA e proteínas). Técnicas de informática são aplicadas para entender e organizar a informação associada com moléculas em larga escala, sendo portanto, um sistema de gerenciamento para biologia molecular (LUSCOMBE et

al., 2001). Programas de alinhamento como o BLAST são usados para comparar a similaridade de seqüências protéicas ou nucleotídicas alvo com os bancos de dados de proteínas e DNA, com seqüências de DNA frequentemente sofrendo tradução conceitual antes da comparação ser executada (ALTSCHUL et al., 1997).

Entre os bancos de dados existentes está em destaque o GenBank, que é formado pela união de diferentes bancos e que disponibiliza suas seqüências publicamente (www.ncbi.nlm.nih.gov). Outro banco importante é o dbEST, que consiste no conjunto de ESTs depositadas no GenBank, que torna a análise de ESTs mais direta.

A liberação do acesso ao banco de dados de ESTs (*Espressed Sequence Tags*), em 1994, (dbEST - http://ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) via depósito no GenBank tem possibilitado o uso dos ESTs para três aplicações: explorar famílias de genes descritos, conhecer a posição de clones e localizar mapas gênicos com STS (GERHOLD & CASKEY, 1996).

Outros bancos de seqüências, também, vêm sendo organizados como: o *Gene2EST* ([http:// woody.embl-heidelberg.de/gene2est/](http://woody.embl-heidelberg.de/gene2est/)) (GEMÜND et al 2001); o *PEDE* (*Pig EST Data Explorer* - <http://pede.dna.affrc.go.jp/>) (UENISHI, 2004); *Genome Sequence Data Base* (GSDB) (<http://scop.wehi.edu.au/gsdb/>); *European Molecular Biology Laboratory Nucleotide Sequence Database* (EMBL) (<http://www.ebi.ac.uk/embl/>); e *DNA Data Bank of Japan* (DDBL) (<http://ddbj.nig.ac.jp/>).

A identificação de ESTs tornou-se popular em função da possibilidade da descoberta de genes e do perfil de expressão de genes tecido-específico. Em vegetais, existem inúmeros trabalhos onde a análise de EST tem sido aplicada, como por exemplo, numerosos tecidos de *Arabidopsis* (HÖFTE et al., 1993; COOKE et al., 1996), arroz (*Oryza sativa*) (SASAKI et al., 1994), milho (*Zea mays*) (EWIRG et al., 1999), e tomate (*Lycopersicum esculentum*) (OLSON et al., 1991); botão de flor de repolho chinês (*Brassica campestris* L. ssp. *pekinensis*) (LIM et al., 1996); tecido de madeira

(*Populus tremula* e *P. trichocarpa*) (STERKY et al., 1998); xilema de pinho (*Pinus taeda* L.) (ALLONA et al., 1998); sementes de uva (*Vitis vinifera*) (ABLETT et al, 2000); cultivares de cana-de-açúcar (*Saccharum officinarum*) (VETTORE et al., 2001), entre outros (Tabela 01).

Segundo Sterky e Lundeberg (2000), as seqüências de EST são extremamente valiosas para identificação de novos genes. Elas podem mostrar o nível de complexidade da expressão de genes em tecidos vegetais, a influência de fatores ambientais na expressão tecido-específico de genes, determinar os limites taxonômicos da espécie, a posição do gene no cromossomo, além de contribuir para o entendimento de intrôns e limites de exon. A descoberta de polimorfismos de nucleotídeos (SNPs) pode ser uma ferramenta importante para caracterização da variabilidade genética e na construção de mapas metabólicos.

O estudo genômico no Brasil teve como marco inicial o seqüenciamento da bactéria *Xylella fastidiosa*, causadora da doença “amarelinho do *Citrus*”, abrindo caminho para outros projetos serem realizados, como o da cana-de-açúcar (*Saccharum officinarum*) (VETTORE et al., 2001). Na Amazônia, no campo da estudo genômico, destaca-se as atividades da Rede da Amazônia Legal de Pesquisas Genômicas – REALGENE que tem possibilitado o desenvolvimento de estudos genômicos em diferentes espécies, entre eles o estudo de ESTs em frutos do guaraná (*Paullinia cupana* Kunth).

O Instituto Nacional de Pesquisas da Amazônia e a Universidade Federal do Amazonas, com a colaboração da Centro de Recursos Genéticos de Biotecnologia da EMBRAPA, iniciaram, em 2004, o estudo de ESTs em camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh), em função do seu grande potencial econômico, por apresentar frutos ricos em ácido ascórbico (Vitamina C), com variabilidade ampla entre 0,845 a 6,1g de ácido ascórbico/100g de polpa.

ORGANISMO	ÓRGÃO	Nº. DE ESTs	FONTE
<i>Arabidopsis thaliana</i>	Sementes, botão de flor, cultura de células, folhas, plantas verdes, inflorescência, óvulo, raiz, flores e hipocótilo.	227.670	HÖFTE et al., 1993; COOKE et al., 1996; QUACKENBUSH et al., 2000 e 2001; EWIRG et al., 1999
<i>Bixa orellana</i>	Sementes	870	JAKO et al., 2002
<i>Brassica campestris</i>	Botão de flor	1216	LIM et al., 1996
<i>Glossypium</i>	Fibra, óvulo, sementes, antera, embrião, folhas, raízes e hipocótilo.	52.818	TIGR*; BURKE & O'MAHONY, 2001.
<i>Glycine max</i>	Meristema apical, botão, cotilédone, hipocótilo, meristema floral, flores, folhas, plumula, vagem, raiz, sementes, embrião somático e caule.	333.481	QUACKENBUSH et al., 2000, 2001
<i>Lotus japonicus</i>	Flores, botões, nódulos, raiz, vagem, sementes e plantas jovens.	109.618	TIGR*
<i>Lycopersicon esculentum</i>	Antera, <i>callus</i> , ovário, cotilédone, flores, frutos, pericarpo, folhas, meristema, pólen, raiz, sementes, estômatos, pétalas e tricomas.	155.317	OLSON et al., 1991 QUACKENBUSH et al., 2000 e 2001 RONNING et al., 2003.
<i>Medicago truncatula</i>	Flores, folhas, cotilédones, vagem, raiz, sementes e caule.	189.714	QUACKENBUSH et al., 2001 SUZUKI et al., 2002 SHELTON et al., 2002
<i>Melaleuca alternifolia</i>	Parte aérea de plantas jovens	454	
<i>Nicotiana benthamiana</i>	Folhas e raiz	18.832	TIGR*
<i>Nicotiana tabacum</i>	Antera, cultura de células, flores, folhas, tecidos regenerados, raízes e estigma.	9.998	TIGR*
<i>Oryza sativa</i>	Antera, <i>callus</i> , cariópse, embrião endosperma, flores, folhas, sementes, floema, pistilo e raiz.	272.567	SASAKI et al., 1994; YAMAMOTO & SASAKI, 1997; QUACKENBUSH et al., 2000 e 2001
<i>Pinus taeda.</i>	Parte aérea, pinha macho, pólen, embrião, raiz, caule e xilema.	125.061	ALLONA et al., 1998 TIGR*
<i>Saccharum officinarum</i>	Meristema apical, <i>callus</i> , inflorescência, casca da haste, flores, folhas, raízes e sementes.	291.689	VETTORE et al., 2001 TIGR*
<i>Solanum tuberosum</i>	Epiderme, folhas, pecíolo, microtubolo, perimedula, raiz, parte aérea, folhas e botões.	157.197	QUACKENBUSH et al., 2000 e 2001 RONNING et al., 2003.
<i>Vitis vinifera</i>	Folhas e semente	4917	ABLETT et al., 2000.
<i>Zea mays</i>	Raiz, antera, folhas, embrião, inflorescência, endosperma, meristema, sementes, mesófilo, ovário, pedicelo, pericarpo, pólen, raiz, esperma e esporos.	109.618	QUACKENBUSH et al., 2000 e 2001 EWIRG et al., 1999

*TIGR - <http://www.tigr.org/plantProjects.shtml>

Tabela 01 - Relação de organismos vegetais cujos transcriptomas foram analisados via sequenciamento de ESTs.

Conhecer os genes expressos no fruto do camu-camu é um passo importante para tentar compreender seu metabolismo e a partir daí tentar inferir os possíveis fatores determinantes na expressão de metabólitos, como por exemplo, na determinação do acúmulo de vitamina C em concentração tão elevada e variadas em frutos de diferentes populações; bem como, demais vias biosintéticas existentes no fruto.

Além disso, o seqüenciamento de ESTs poderá contribuir para um maior conhecimento sobre esta espécie e auxiliar na sua domesticação. A utilização destes dados poderiam auxiliar os programas de melhoramento genético bem como estimular ainda mais o cultivo em terra-firme. O conhecimento da biologia do camu-camu poderá favorecer sua comercialização na região, tornando-se também uma importante fonte de renda para as comunidades locais e para o Estado.

Diante do exposto objetivou-se analisar as seqüências ESTs de frutos de camu-camu de plantas cultivadas em terra-firme, com o uso de tecnologias genômicas e por meio de ferramentas de bioinformática.

2. MATERIAL E MÉTODOS

A partir do seqüenciamento da extremidade 5' de cDNAs de frutos de camu-camu oriundos de plantas cultivadas em terra firme da fazenda Yuricam no Rio Preto da Eva (AM) obteve-se cerca de 5000 seqüências ESTs que foram submetidas ao Sistema Genoma, programa de anotação genômica produzido no Laboratório de Bioinformática da Rede de Bioinformática do Centro-Oeste – BIOFOCO, da EMBRAPA - Recursos Genéticos e Biotecnologia, (<http://condor.genoma.cenargen.embrapa.br/genoma/classes/JFrameGenoma.jnlp>).

O Sistema Genoma integra programas de gerenciamento, análise e visualização de seqüências nucleotídicas e cromatogramas produzidos em sequenciadores automáticos (Megabace, ABI3700) (www.genoma.embrapa.br).

O *pipeline* de análise do sistema genoma consiste em realizar uma filtragem pela qualidade e extensão da seqüência nucleotídica pelos programas PHRED (EWING et al., 1998), seguido da limpeza dos segmentos com vetores e adaptadores pelo programa CROSSMATCH (<http://www.phrap.org/phredphrap/general.html>). Após a análise de qualidade dos cromatogramas e eliminação de seqüências de vetores a serem selecionados, as seqüências armazenadas no banco de dados através da hierarquia de projeto-biblioteca-placa-clone.

Após armazenadas as seqüências são submetidas à fase de montagem através do programa TGICL (<http://www.tigr.org/tdb/tgi/software/>), que gera os agrupamentos por similaridade e por qualidade. O resultado desta fase é a construção dos grupos de contigs e singletos. O conjunto total de seqüências agrupadas é então comparada com diferentes bancos de dados disponíveis com o programa de alinhamento BLAST e BLASTx.

Os bancos disponíveis são espelhos dos bancos mantidos no GenBank e são atualizados uma vez por mês para os servidores do Laboratório de Bioinformática da Embrapa-Recursos Genéticos e Biotecnologia. Dentre os bancos disponíveis estão: nr, o genoma de *Arabidopsis*, o genoma do milho (versão de setembro de 2005), proteínas das vias de carotenóides, ESTs de outros organismos, swissprot, genoma de arroz (versão de maio de 2005), *Gene Ontology* e aminoácidos de mitocôndrias.

Os resultados das buscas foram analisados quanto à similaridade usado como parâmetros o índice estatístico *e-value* menor/igual a e^{-10} pela redundância da biblioteca, para o número de *contigs* e *singletons*; além de característica funcionais das proteínas (categorização) identificadas com grupos órtogos (GO). Outras ferramentas foram associadas à anotação como a possibilidade de identificação de microssatélites (Troll), *Open Reads Frames* (ORF Finder) e de mapas metabólicos (KEGG).

3. RESULTADOS E DISCUSSÃO

3.1. Análise da Biblioteca

Das 5000 seqüências ESTs obtidas 36% foram consideradas inválidas em função do tamanho e/ou qualidade da seqüência, bem como do percentual de vetor presente na mesma. Para ser considerada válida, pelo Sistema Genoma, as seqüências deveriam ter boa qualidade, ter baixo percentual de vetor e com tamanho mínimo de 150pb (sem o vetor).

A partir das 3196 ESTs de frutos de camu-camu validadas no programa Sistema Genoma, sendo formados 1546 *singletons* (seqüências que não apresentam similaridade com nenhuma outra) e 358 *contigs* (seqüências idênticas ou contendo regiões similares formando um contig, ou um unigene, que seria a representação virtual do RNAm na célula) (Tabela 2).

Biblioteca de ESTs	
No. de ESTs sequenciados	5000
No. seqüências válidas	3196
No. de ESTs identificados*	2586
Redundância (%)**	32,54
Novidade (%)***	81
Agrupamento das seqüências ESTs	
No. de <i>singletons</i>	1546
No. de <i>contig</i>	358
No. de <i>reads</i> em <i>contigs</i>	1040
No. de seqüências reunidas (<i>Unique</i>)	1904
Distribuição dos ESTs nos <i>contigs</i>	
2 – 4 ESTs (%)	89,4
5 – 10 ESTs (%)	9,2
> 10 ESTs (%)	1,4

* Total de reads em contigs + singletons

Redundância = número de ESTs em *contigs100/número total de ESTs

***Percentual fornecido pelo programa Sistema Genoma

Tabela 02 – Sumário do seqüenciamento de EST de cDNA de frutos de camu-camu

Um total de 1040 seqüências ESTs foram agrupadas formando os *contigs*, que unidos aos *singletons* somam 2586 seqüências ESTs com similaridade a seqüências encontradas no banco de genes, destes foram identificados 1904 *uniques* - conceituado como provável seqüência que codificam um gene, seja ela uma *singleton* ou um *contig* (Tabela 2).

A análise da clusterização da biblioteca revelou um índice de 81% de novidade e 32,54% de redundância para a biblioteca de cDNA de camu-camu, estes índices mostram que a biblioteca ainda pode ser bastante explorada para a obtenção e identificação de novos genes (Tabela 2).

A distribuição das *reads* entre os *contigs* confirma a baixa redundância da biblioteca (Tabela 2). Baseado nas categorias descritas por Lee et al. (2005), observou-se que: cerca de 89% dos ESTs que formaram *contigs* apresentaram baixa redundância (2-4 *reads* por *contigs*), dos quais 218 *contigs* apresentaram duas *reads*, 70 *contigs* apresentaram 3 *reads* e 32 *contigs* apresentaram 4 *reads*; 9,2% com média redundância (5-10 *reads* por *contigs*); e apenas 1,4% dos ESTs apresentaram alta redundância (> 10 *reads* por *contigs*), dentre os quais apenas um *contig* apresentou o maior número de *reads* (19 *reads*). Parece existir um consenso de que os *contigs* formados com até 4 *reads* são considerados não redundantes, entretanto, com relação as demais categorias isso não é observado.

3.2. Categorização funcional das proteínas identificadas

Baseado principalmente nas categorias funcionais estabelecidas para eucariotos pelo NCBI, os ESTs com provável identidade com proteínas foram classificados dentro de 24 grupos: 22 grupos de categorias funcionais propriamente ditas; um grupo

contendo proteínas com funções preditas e um grupo de proteínas com funções desconhecidas (Figura 01).

O maior percentual observado foi relacionado à proteínas com funções não determinadas, em torno de 34% (Figura 01). Este mesmo perfil, também é observado na avaliação dos ESTs das espécies *Melaleuca alternifolia* (SHELTON et al., 2002) *Lolium perene* (SAWBRIDGE et al., 2003) e *Saccharum officinarum* (CASU et al., 2003).

Dentre as categorias com proteínas com função conhecida destacaram-se as seguintes: modificações pós-traducional, proteínas *turnover* e chaperonas (13,2%), metabolismo e transporte de carboidratos (8,8%), tradução, biogênese e estrutura ribossomal (5,7%) (Figura 01).

Em cana-de-açúcar proteínas relacionadas ao metabolismo e transporte de carboidratos representaram 2,1% de genes expressos em talos imaturos e 2,4% de gene expressos em talos em amadurecimento, sugerindo que os transcritos envolvidos na acumulação de carboidratos não são abundantes nesta espécie (CASU et al., 2003). Em *Populus alba* x *P. tremula* var. *glandulosa* dos 8962 ESTs de células em suspensão, foram identificados 4,2% de proteínas relacionados com a transcrição e 5,2% a sinal transdução (LEE et al., 2005).

Relacionados à categoria modificações pós-traducionais, proteína *turnover* e chaperonas foram identificados os genes provavelmente codificantes de proteínas como: glutationa S-transferase, proteínas de choque térmico, tioredoxina, protease cisteína, proteinase aspartica, proteínas como calnexin, proteosoma alfa 20S subunidade 6, proteína disulfeto isomerase (vide apêndice 1). As chaperonas serão caracterizadas no capítulo 3.

As proteínas relacionadas as modificações pós-traducionais (*Posttranslational modification* – PTM) modulam a atividade de muitas enzimas eucarióticas. As PTM são

atividades coadjuvantes de mudanças de propriedades de proteínas pela clivagem proteolítica ou pela adição de um grupo químico (fosfato, glicose, manose, etc.) em um ou mais aminoácidos (MANN E JENSEN, 2003).

Com as PTMs há a possibilidade de regular o metabolismo enzimático, podendo afetar a atividade enzimática de uma proteína, a sua localização intracelular, as interações proteína-proteína e a sua estabilidade (HUBER & HARDIN, 2004).

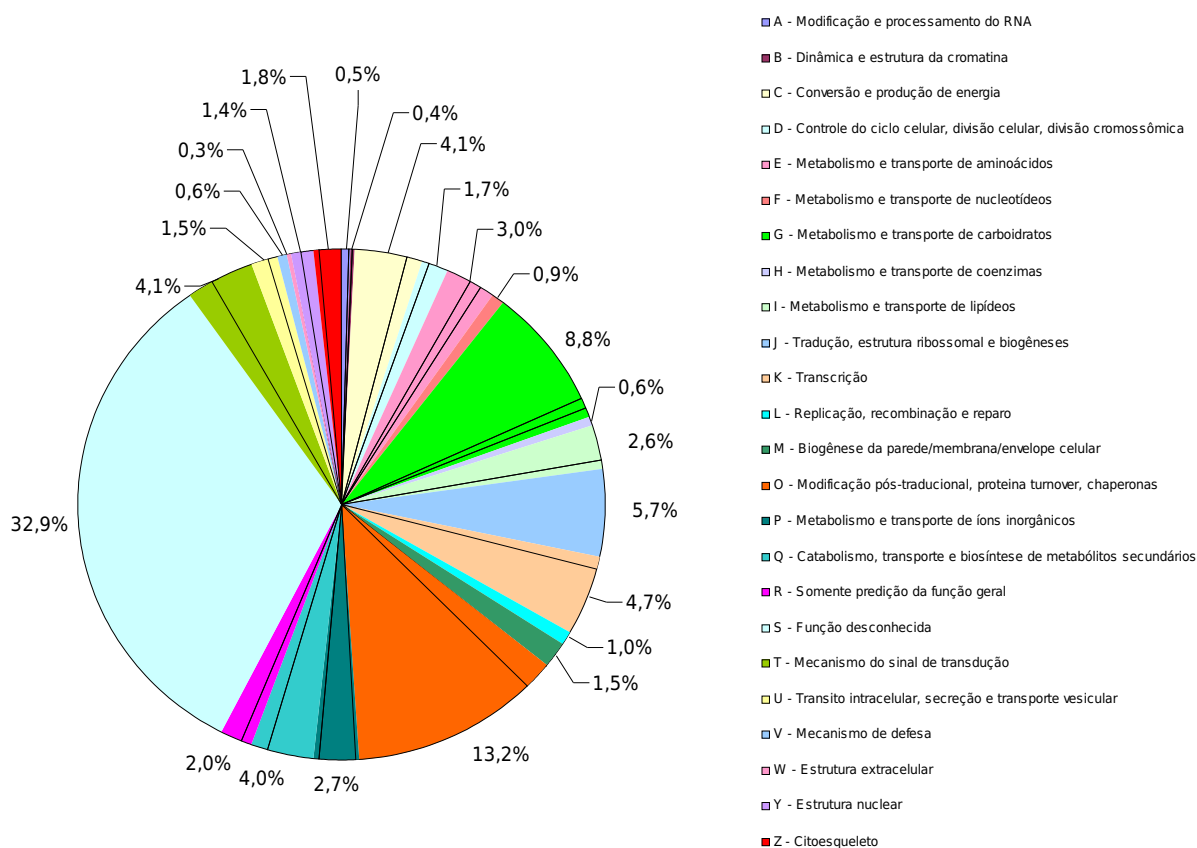


Figura 01 – Categorização funcional das prováveis proteínas identificadas a partir da biblioteca de ESTs de frutos de camu-camu.

Em relação à categoria metabolismo e transporte de carboidratos foram relacionados genes que codificam proteínas como gliceraldeído-3-fosfato deidrogenase, aquaporina, malato deidrogenase, enolase, β-amilase, fosfoglicerato mutase, celulose

síntese, ribulose-fosfato-3-epimerase, entre outras (vide apêndice 1). As enzimas relacionadas ao metabolismo do ácido ascórbico são descritas no capítulo 4.

O armazenamento de carboidratos é necessário para sustentar o desenvolvimento das plantas em períodos de estresse, durante a dormência, e muito importante no início de crescimento e frutificação.

Genes relacionados ao metabolismo dos carboidratos em cana-de-açúcar foram definidos como seqüências gênicas provavelmente codificantes de enzimas envolvidas na síntese, na modificação ou colapso de açúcares simples, nos fosfatos de açúcares, açúcares de nucleotídeo ou no transporte de açúcar (CASU et al., 2003).

Em sementes de uva observou-se uma maior expressão de genes relacionados ao metabolismo de açúcares e polissacarídeos, sugerindo elevada atividade no transporte e produção de carboidratos na semente (ABLETT et al., 2000).

Dentre as proteínas relacionadas à tradução, biogênese e estrutura ribossomal foram identificados ESTs prováveis codificantes de proteínas ribossomais, fator de iniciação eIF4A-15, provável leucil-tRNA síntase, proteína EF-1-alfa ligada ao GTP, constituinte estrutural de ribossomos, provável fator de iniciação tansducional eIF-1-A, entre outros (vide apêndice 1).

A tradução das informações genéticas em produtos protéicos é essencial para todo o sistema biológico. Uma etapa chave na expressão de genes é controlada pela interação entre o mRNA e os ribossomos. Nos organismos eucariotos, a tradução do mRNA pode ser dividida em três consecutivas etapas: iniciação, alongação e terminação (CHAPMAN & BROWN, 2004).

A fidelidade na síntese de proteínas é em parte dependente da correta troca do tRNA com seus aminoácidos cognatos. Este processo é catalizado pelas aminoacil-tRNA sintetases específicas para cada tRNA particular. As 20 aminoacil-tRNA podem

ser divididas em duas distintas classes baseada nas similaridades funcionais e estruturais (BULLARD et al., 2000)

A células vegetais requerem todas aminoacil-tRNA sintetase em todos os seus compartimentos (citosol, cloroplasto e mitocôndrias). Algumas seqüências de aminoacil-tRNA sintetases de *Arabidopsis thaliana* foram depositadas em banco de dados, em muitos casos como seqüências parciais (EST). O cDNA da asparagil-tRNA sintetase de *A. thaliana* foi caracterizado (AUBOURG et al.,1998).

A quarta categoria funcional protéica com maior percentagem foi relativa ao processo de transcrição, na qual foram identificadas, dentre outras, as seguintes proteínas: proteína MADS *box*, fator de transcrição como *myb 2*, provável fator de transcrição bHLH, provável proteína bZIP relacionado ao amadurecimento, fator de transcrição 16, WRKY17, fator de transcrição de choque térmico 29 (vide apêndice 1).

A transcrição consiste na síntese de RNAs, processo que ocorre em três etapas reguladas por fatores de transcrição. Os fatores de transcrição são seqüências específicas de proteínas ligadas ao DNA que interagem com a região promotora de genes alvo e modulam a rápida iniciação da síntese do mRNA pela RNA polimerase II (ENDT et al., 2002).

Uma variedade de fatores de transcrição já foram identificados em diferentes espécies vegetais, como o bZIP de *Saccharum officinarum* (VINCENTZ et al., 2001); MADS *box* de *Solanum tuberosum* (KANG E HANNAPPEL, 1995), *Eucalyptus globulus* (DECROOCQ et al.,1999) e *Dianthus caryophyllus* L. cv Scania (BAUDINETTE et al., 2000); WRKY e bHLH de *Arabidopsis thaliana* (KALDE et al., 2003; PENFIELD et al., 2005).

Em relação aos mecanismos de transdução de sinal foram identificadas ESTs provavelmente codificantes das seguintes proteínas: provável proteína leucina contendo *zipper*, proteína associada ao microtúbulo 65kD, quinase MAP, calmodulina 8,

provável CBF1 interagindo com o co-repressor CIR, como copina III, entre outros (vide apêndice 1).

A transdução de sinal é o processo no qual um receptor interage com um ligante na superfície da célula, transmitindo um sinal biológico para direcionar uma determinada rota metabólica. Em cana-de-açúcar foram identificados vários componentes para transdução de sinais entre eles, calmodulin, a quinase MAP e ACC sintase (SOUZA et al., 2001), quinase MAP também foram identificados em *Nicotiana tabacum* (SHIBATA et al., 1995).

Quanto a categoria de conversão e produção de energia foram identificadas ESTs relacionadas às proteínas anidrase carbônica, glicolato oxidase/oxidoreductase, serina o-acetiltransferase 1, ribulose-1,5-bifosfato carboxilase/oxigenase ativase, proteína cloroplastídica do fotossistema II 10kDa, provável NADH deidrogenase, entre outras (vide apêndice 1).

Na categoria catabolismo, transporte e biossíntese de metabólitos secundários foram identificadas ESTs possivelmente codificadoras das proteínas citocromo P450, proteína transportadora de acil, homologa a 2-hidroxi isoflavona/diidroflavonol redutase, 4-coumatato: coenzima A ligase, chalcone síntase, cinamil álcool deidrogenase, provável diidroflavonol redutase, provável flavonol síntase, giberilina 2 oxidase, leucoantocianidina dioxigenase, entre outras (vide apêndice 1).

3.3. Análise da Similaridade das Seqüências ESTs.

Das 3196 seqüências ESTs validadas pelo Sistema Genoma 15% não apresentaram similaridade com nenhuma seqüência contida no banco de genes (nr/Blastx) (Figura 7), chamadas de *no hits*, desta 337 formaram *singletons* e 118 ESTs foram agrupados em 44 *contigs* (Tabela 3).

Quanto aos *no hits* alguns aspectos devem ser ressaltados, essas seqüências tem grande importância pois podem representar genes nunca identificados, por outro lado podem também, indicar a presença de clones com mais de um tipo de inserto ou simplesmente problemas técnicos durante a reação de seqüenciamento, em se tratando dos *singletons*. Contudo, levando-se em consideração as seqüências que formaram *contigs*, a margem de erro diminui consideravelmente e em função do número de seqüências expressas obtidas, estas sim, podem ser de grande importância no metabolismo do camu-camu.

Foram observados 3% de seqüências (97 ESTs) com baixa similaridade (*e-value* > e^{-10}) chamadas de *low hits* (Figura 02), destas 30 seqüências estão distribuídos em 15 *contigs* e 67 ESTs foram identificados como *singletons* (Tabela 3).

Definição	No. EST		
	<i>Singleton</i>	<i>Contig</i>	<i>Reads em contigs</i>
	<i>s</i>	<i>s</i>	
<i>Low hits</i>	67	15	30
<i>No hits</i>	337	44	118
Proteínas desconhecidas	222	29	68
Proteínas hipotéticas	33	1	2
Prováveis proteínas	20	-	-
Produto protéico não nomeado	3	-	-

**No hits* – proteínas que não apresentam similaridade com as seqüências dos bancos de dados utilizados para a análise.

***Low hits* – seqüências que apresentam baixa similaridade (*e-value* > e^{-10})

Tabela 03 – Relação de seqüências ESTs não identificadas e não similares a seqüências do banco de genes

Essa baixa similaridade pode ser atribuída à distância genética entre os organismos com seqüências gênicas depositadas no *GenBank* e o camu-camu. No NCBI, apenas uma única espécie do gênero *Myrciaria*, ao qual pertence o camu-camu, possui seqüências depositadas, entretanto com número insignificante, duas seqüências protéicas.

Cerca de 82% apresentaram *e-value* $\leq e^{-10}$, o que indica a obtenção de seqüências de genes potenciais codificantes de proteínas (Figura 02). Entretanto, 348 seqüências ESTs apresentaram similaridade com seqüências que codificam proteínas

ainda não identificadas, destas 222 *singletons* e 29 *contigs* foram classificados como proteínas desconhecidas, 33 *singletons* e 1 *contig* como proteínas hipotéticas, 20 ESTs como prováveis proteínas e 3 ESTs como produto protéico não nomeado (Tabela 3).

Dezoito por cento das seqüências apresentaram altíssima similaridade (e-value < e^{-71} à 0.0) à uma variedade de diferentes genes conhecidos ou família de genes (Figura 02).

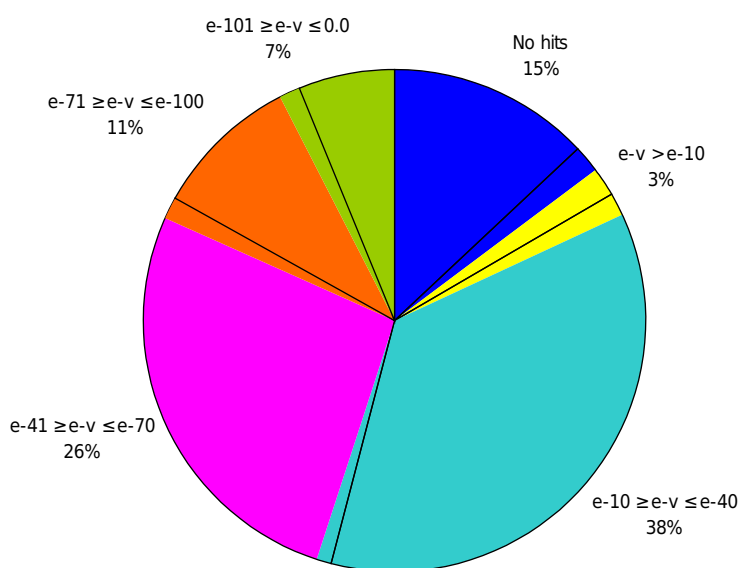


Figura 02 – Nível de similaridade encontrada na biblioteca de cDNA de frutos de camu-camu (3196 ESTs válidos)

As 20 seqüências de frutos de camu-camu com maior similaridade a genes conhecidos, estão relacionados na tabela 4. Sessenta e duas seqüências ESTs distribuídas em 10 *contigs* foram consideradas idênticas (*e-value* 0.0) a genes codificantes das seguintes proteínas: Proteossoma 26S subunidade tipo 4; álcool desidrogenase 2; alfa tubulina 1; catalase; H⁺ - pirofosfatase; duas proteínas de choque térmico 83, 70 kDa proteína de choque térmico 1; SNF1 relatada como proteína quinase; e fator de alongação na tradução 1A-9 (Tabela 4).

Clusters	No. ESTs	Definição	Organismo	GI	e-value
CL205					
contig1	2	Proteossoma 26S subunidade tipo 4	<i>Solanum tuberosum</i>	77745479	0.0
CL12 contig1	8	Alcool desidrogenase 2	<i>Vitis vinifera</i> <i>Pseudotsuga menziesii</i> var. <i>menziesii</i>	18027092	0.0
CL64 contig1	4	Alfa tubulina 1	<i>Prunus persica</i>	56481497	0.0
CL13 contig1	9	Catalase	<i>Vitis vinifera</i>	32526568	0.0
CL18 contig1	8	H ⁺ -pirofosfatase	<i>Ipomoea nil</i>	7739779	0.0
CL36 contig1	5	Proteína de choque térmico 83	<i>Ipomoea nil</i>	1708314	0.0
CL70 contig1	4	Proteína de choque térmico 83	<i>Ipomoea nil</i>	1708314	0.0
CL3 contig1	15	70 kDa cognata de choque térmico 1	<i>Vigna radiata</i>	45331281	0.0
CL197					
contig1	2	Proteína SNF1-relatada como quinase	<i>Cucumis sativus</i>	1743009	0.0
CL24 contig1	5	Fator de alongação na tradução 1 ^a -9	<i>Gossypium hirsutum</i>	74486744	0.0
CL139					
contig1	3	Fosfoglicerato quíase 1 citosólica	<i>Populus nigra</i>	3738257	3E-173
CL25 contig1	6	Proteína desconhecida	<i>Ricinus communis</i>	1621268	7E-171
CL32 contig1	6	Fator de alongação na tradução 1A-9	<i>Gossypium hirsutum</i>	74486744	3E-166
CL171					
contig1	3	HSP70 tipo DnaK – arroz	<i>Oryza sativa</i>	62733235	2E-161
CL177		GSK1 (GSK3/proteína quinase tipo Shaggy 1);			
contig1	2	quinase	<i>Arabidopsis thaliana</i>	42571361	1E-158
CL231					
contig1	2	Beta-D-glucosidase	<i>Gossypium hirsutum</i>	33391721	2E-156
CL90 contig1	4	Precursor da proteína tipo LYTB	<i>Adonis palaestina</i>	10716963	1E-154
CL149					
contig1	3	Fenilalanina amônio liase	<i>Camellia sinensis</i>	1171998	3E-154
CL22 contig1	8	Provável cisteína proteinase tipo papaína	<i>Gossypium hirsutum</i>	57282617	5E-153
CL450		ATP-ligante/ hidrolase/nucleosídeo-trifosfatase/			
contig1	2	nucleotídeo-ligante	<i>Arabidopsis thaliana</i>	15231775	2E-152
CL469					
contig1	2	Gene RNA ribossomal 18S	<i>Heteropyxis natalensis</i>	7595447	1E-150
CL89 contig1	3	Alil álcool desidrogenase	<i>Nicotiana tabacum</i>	6692816	3E-150

Tabela 04 – As 20 seqüências formadas a partir de ESTs de frutos de camu-camu com alta similaridade a seqüências do banco de genes.

Grande parte das 2586 seqüências ESTs de camu-camu (*reads* em *contigs* e *singletons*) apresentaram similaridade com seqüências que codificam proteínas oriundas de diferentes organismos vegetais. As 20 espécies com maior número de ESTs com similaridade a seqüências de camu-camu estão destacadas na figura 03.

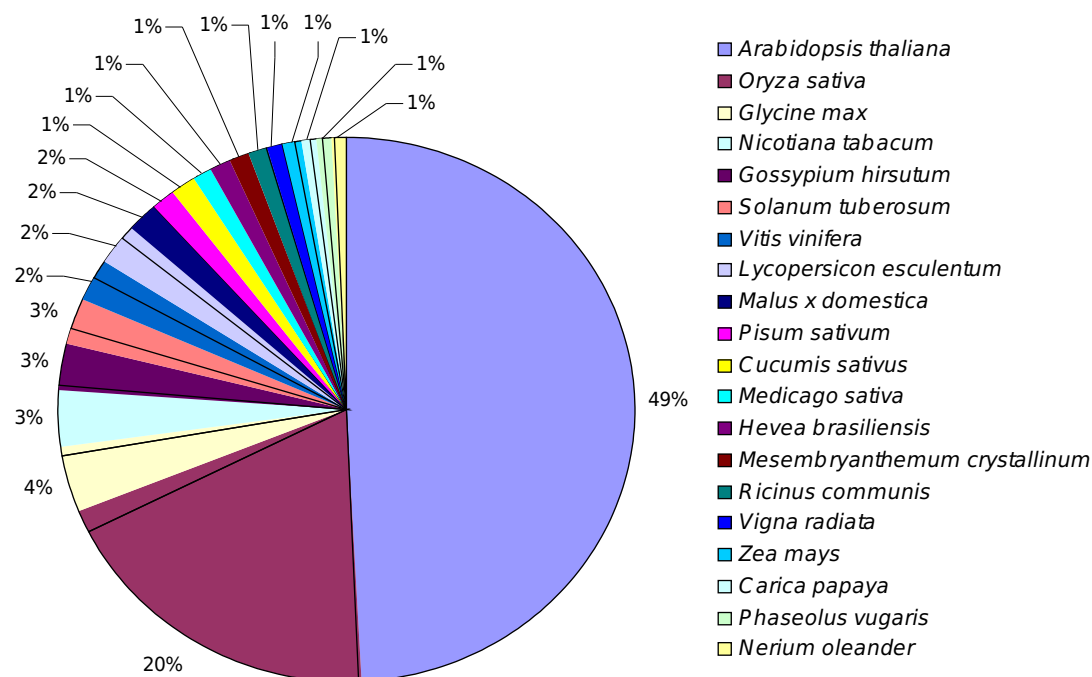


Figura 03 – Percentagem de EST com similaridade a seqüências de diferentes organismos vegetais.

A espécie *Arabidopsis thaliana* se destaca com 49% de seqüências similares, seguida da *Oryza sativa* com 20%, *Glycine max* com 4% e *Nicotiana tabacum*, *Gossypium hirsutum* e *Solanum tuberosum* ambas com 3% de seqüências similares cada. Este percentual de similaridade com as espécie *A. thaliana* e *O. sativa* é justificado pelo elevado número de seqüências depositadas no banco de gene (vê NCBI) (Figura 03).

Os vinte mais abundantes genes expressos em frutos de camu-camu das 2586 seqüências ESTs analisadas estão destacados na tabela 5. O gene mais freqüentemente

expresso codifica a proteína glutatona S-transferase, relacionado a 27 ESTs distribuídos em *contigs* e *singletons*, com nível de expressão de 1,04% e representando 2,06% dos genes identificados (Tabela 05).

Rank	No. ESTs	Definição	Nível de expressão (%)	Unique (%)
1	27	Glutaciona S-transferase	1,04	2,06
2	19	Alergênica maior Mal d 1.06A02	0,73	1,45
3	15	Cognata de choque térmico 70 kDa proteínas 1	0,58	1,15
4	15	Proteína tipo extensina	0,58	1,15
5	13	Fator de alongação de tradução 1 ^a -9	0,50	0,99
6	12	Giberelina 2 oxidase	0,46	0,92
7	12	(S)-norcoclaurina síntase	0,46	0,92
8	11	RNA ligante/ácido nucléico ligante	0,43	0,84
9	11	Metalotioneína	0,43	0,84
10	11	Provável proteína rica em glicina ligada ao RNA	0,43	0,84
11	11	ELIP	0,43	0,84
12	10	Catalase	0,39	0,76
13	9	Provável fator de transporte nuclear 2	0,35	0,69
14	9	Proteína de choque térmico 83	0,35	0,69
15	9	Gliceraldeído 3-fosfato desidrogenase	0,35	0,69
16	8	Provável proteínase cisteína tipo papaína	0,31	0,61
17	8	Proteínas resposta a baixa temperatura e sal	0,31	0,61
18	8	Proteína de choque térmico 17.7 kDa	0,31	0,61
19	8	H ⁺ -pirofosfatase	0,31	0,61
20	8	Alcool desidrogenase 2	0,31	0,61

Nível de expressão = no. EST*100/total de EST identificados

Unique = seqüência que codificam um gene, seja ela uma *singleton* ou um *contig* (= no. EST*100/total de Uniques)

Tabela 05 – Os 20 genes mais expressos na biblioteca de frutos de camu-camu, representados pelo maior número de ESTs relacionado.

As glutatona S-transferase (GSTs, E.C. 2.5.1.18) são enzimas que catalisam a conjugação do tripeptídeo glutatona (GSH) com uma variedade de substratos hidrofóbicos, electrofílicos e citotóxicos. As GSTs expressas em vegetais foram as primeiras a serem identificadas e intensivamente estudada por causa de sua capacidade de desintoxicação de herbicidas (MARRS, 1996).

Recentemente, outras subclasses de GST de plantas foram implicadas em numerosas respostas ao estresse, como ataque de patógenos, estresse oxidativo e

toxicidade por metais pesado. Além do mais, as GSTs desempenham um papel na resposta celular a auxinas e no metabolismo de produtos secundários como antocianinas e o ácido cinâmico (MARRS, 1996).

Dezenove ESTs foram relacionados a alergênica maior *Mal d 1.06A02* com nível de expressão de 0,73% e representando 1,45% dos *uniques genes* (Tabela 4). Vários cDNAs que codificam a proteína *Mal d 1* já foram clonados e a análise das seqüências revelou grande homologia com *Bet v 1* (proteína que causa alergia em várias pessoas) e outras proteínas intracelulares relacionadas a patógenos (IPR). Até agora a *Mal d 1* e o transcrito *Mal d 1* foram abundantemente achadas em frutos maduros (VANEK-KREBITZ et al., 1995; PÜHRINGER et al., 2000).

A proteína *Mal d 1* constitui um importante alergênico da maçã, e foi classificada como proteína 10 relacionada a patógenos (PR-10) por homologia de seqüência, embora sua indução por patógenos ou estresse não tenha sido demonstrada (PÜHRINGER et al., 2000).

Os prováveis genes da proteína cognata de choque térmico 70 kDa 1 e proteína tipo extensina estão relacionados a 15 ESTs cada (nível de expressão de 0,58% e 1,15% dos *uniques genes*), identificados em frutos de camu-camu, cada (Tabela 5).

A proteína cognata de choque térmico 70 kDa (HSC70) pertence à família Hsp70. Proteína citoplasmática constitutivamente expressa, presente em baixos níveis em células não estressadas (KIANG E TSOKOS, 1998; EFREMOVA et al., 2002). Estudos bioquímicos de fragmentos de HSC70 levaram ao mapeamento e caracterização de dois domínios, um fragmento de 44 kDa contém o domínio ATPase e um fragmentos de 18 kDa contém um domínio ligado ao peptídeo que une peptídios enovelados e desenovelados (ou desnaturados) (WANG et al., 1993; KAING E TSOKOS, 1998).

A extensina (glicoproteína rica em hidróxiprolina - HRGP) é um componente da parede celular primária e existe em duas formas: solúvel, que limita ionicamente a parede celular, e uma forma insolúvel que é depositada na parede celular durante a interação hospederio-patógeno por uma peroxidase/hidroperóxido. A extensina é importante no controle de crescimento, resistência a doenças e morfogênese, sua solubilidade pode ser regulada por fatores citoplasmáticos (ascorbato e/ou íons de hidrogênio), pois o meio ácido inibe a ligação com a peroxidase (COOPER & VARNER, 1984; BROWNLEADER et al., 1995).

Outros possíveis genes, também expressos em frutos de camu-camu foram: fator de alongação de tradução 1A-9 (13 ESTs - 0,50% expresso), giberelina 2 oxidase (12 ESTs - 0,46% expresso) e (S)-norcoclaurina síntase (12 ESTs - 0,46% expressos) (Tabela 4).

Os possíveis genes mais expressos em frutos de camu-camu estão, em geral, relacionados à defesa celular contra estresse de temperatura e estresse oxidativo e contra patógenos (Tabela 04). Isso se deve, provavelmente, ao material de origem das bibliotecas do camu-camu ser oriundo de terra-firme, fora de seu habitat natural, margens de rios e lagos da Amazônia (CALZADA BENZA, 1980; FALCÃO et al., 1993), onde durante o período de frutificação a planta encontra-se quase que totalmente submersa e conseqüentemente submetida a temperaturas mais baixas; o cultivo em terra-firme também favorece o surgimento de várias pragas na cultura (VILLACHICA, 1996; BARBOSA et al., 2004), o que provavelmente, estimulou o mecanismo celular de defesa contra patógenos.

4. CONCLUSÕES

- A biblioteca de ESTs de frutos de camu-camu contém até o presente momento 3196 ESTs validados no programa Sistema Genoma, com 2586 seqüências ESTs com similaridade a seqüências do banco de genes e um índice de 81% de novidade e 32,54% de redundância;
- As principais categorias funcionais nos quais os ESTs estão representados foram classificados em: modificação pós-traducional, proteína *turnover*, chaperonas (13,2%), Metabolismo e transporte de carboidratos (8,8%), Tradução, biogênese e estrutura ribossomal (5,7%), Transcrição (4,7%), Mecanismo de sinal transdução (4,1%), Conversão e produção de energia (4,1%), e Catabolismo, transporte e biossíntese de metabólitos secundários (4,0%);
- A *Arabidopsis thaliana* foi a espécie com maior número de ESTs com similaridade a seqüências de camu-camu (49%), seguida de *Oryza sativa* (20%) e *Glycine max* (4%);
- As seqüências com maior similaridade (*e-value* 0.0) a genes conhecidos, são codificantes de: proteosoma 26S subunidade 4; álcool desidrogenase 2, alfa tubulina 1, catalase, Pirofosfatase H⁺, duas proteínas de choque térmico 83; proteína de choque térmico 70 KDa 1; Proteína quinase relatada como SNF1; fator de alongação de transdução 1A-9.
- Os prováveis genes mais expressos em frutos de camu-camu foram: Glutathione S-transferase; Alergênica major Mad d 1.06A02; cognata de choque térmico 70 kDa proteína 1; proteína tipo extensina; fator de alongação de tradução 1A-9.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ABLETT, Effie; SEATON, George; SCOTT, Kirsten; SHELTON, Dale; GRAHAM, Michael W.; BAVESTOCK, Peter; LEE, L. Slade; HENRY, Robert. Analysis of grape ESTs: global gene expression patterns in leaf and berry. *Plant Sci.*, v. 159, pp. 87-95, 2000.

ALLONA, Isabel; QUINN, Michelle; SHOOP, Elizabeth; SWOPE, Kristi; STCYR, Sheila; CARLIS, John; RIEDL, John; RETZEL, Ernest; CAMPBELL, Malcon M; SEDEROFF, Ronald; WHETTEN, Ross W. Analysis of xylem formation in pine by cDNA sequencing. *PNAS*, v. 95, pp. 9693-9698, 1998.

ALTSCHUL, Stephen F. ; MADDEN, Thomas L.; SHÄFFER, Alejandro A.; ZHANG, Jinghui; ZHANG, Zheng; MILLER, Webb ; LIPMAN, David J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, v. 25, n. 17, pp. 3389-3402, 1997.

AUBOURG, S.; CHÉRON, A.; KREIS, M.; LECHARNY, A. Structure and expression of an asparaginyl-tRNA synthetase gene located on chromosome IV of *Arabidopsis thaliana* and adjacent to a novel gene of 15 exons. *Biochim. et Bioph. Acta*, v. 1398, pp. 225-231, 1998.

BARBOSA, Márcio L. L.; ACIOLI, Agno N. S.; OLIVEIRA, Arlem N.; SILVA, Nelito M.; CANTO, Sônia L. O. Ocorrência de *Tuthillia cognata* Hodkinson, Brown & Burckhardt, 1986 (Hemiptera: Homoptera, Psyllidae) em plantios experimentais de camu-camu *Myrciaria dúbia* (H.B.K.) McVaugh em Manaus (Amazonas, Brasil). *Acta Amazônica*, v. 34, n. 1, pp. 115-119, 2004.

BAUDINETTE, Stanley C.; STEVENSON, Trevor W.; SAVIN, Keith W. Isolation and characterisation of the carnation floral-specific MADS box gene, *CMB2*. *Plant Science*, v. 155, pp. 123-131, 2000.

BROWNLEADER, Michael D.; AHMED, Naziha; TREVAN, Michael; CHAPLIN, Martin F; DEY, Prakash M. Purification and partial characterization of tomato extensin peroxidase. *Plant Physiol.*, v. 109, pp. 1115-1123, 1995.

BULLARD, J.M.; CAI, Y.C.; SPREMULLI, L.L. Expression and characterization of the human mitochondrial leucyl-tRNA synthetase. *Biochim. et Bioph. Acta*, v. 1490, pp. 245-258, 2000.

BURKE, J.J.; O'MAHONY, J.O. Protective role in acquired thermotolerance of developmentally regulated heat shock protein in cotton seeds. *The Journal of Cotton Science.*, v. 5, pp. 174-183, 2001.

CALZADA BENZA, J. C. *143 Frutales nativos*. Librería El Estudiante. La Molina, 314p., 1980.

CASU, Rosanne E.; GROF, Christopher P. L.; RAE, Anne L.; MCINTYRE, Cynne L.; DIMMOCK, Christine M.; MANNERS, John M. Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. *Plant Mol. Biol.*, v. 52, pp. 371-386, 2003.

CHAPMAN, Bernice; BROWN, Chris. Translation termination in *Arabidopsis thaliana*: characterisation of three versions of release factor 1. *Gene*, v. 341, pp. 219-225, 2004.

COOKE, Richard et al. Further progress towards a catalogue of all *Arabidopsis* genes - analysis of a set of 5000 non-redundant ESTs. *Plant J.*, v. 9, pp. 101-124, 1996.

COOPER, Jaime B.; VARNER, Joseph E. Cross-Linking of Soluble Extensin in Isolated Cell Walls. *Plant Physiol.*, v.76, pp. 414-417, 1984.

DECROOCCQ, Veronique; ZHU, Xiaomei; KAUFFMAN, Maggie; KYOZYKA, Junko; PEACOCK, W. James; DENNIS, Elizabeth S.; LLEWELLYN, Danny J. A *TM3*-like MADS-box gene from *Eucalyptus* expressed in both vegetative and reproductive tissues. *Gene*, v. 228, pp. 155-160, 1999.

DUEK, Paula D.; FANKHAUSER, Christian. Transcriptional networks in plants: bHLH class transcription factors take centre stage in phytochrome signaling. *TRENDS in Plant Science*, v.10, n. 2, pp. 51-54, 2005.

EFREMOVA, Sofia M.; MARGULIS, Boris A.; GUZHOVA, Irina V.; ITSKOVICH, Valeria B.; LAUENROTH, Stephanie; MÜLLER, Werner E. G.; CHRÖDER, Heins C. Heat shock protein Hsp70 expression and DNA damage in Baikalian sponges exposed to model pollutants and wastewater from Baikalsk Pulp and Paper Plant. *Aquatic Toxicology*, v. 57, pp. 267–280, 2002.

ENDT, Débora V.; KIJNE, Jam W.; MEMELING, Joham Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry*, v. 61, pp. 107–114, 2002.

EWING, Brent; HILLIERM LaDeana; WENDL, Michael C.; GREEN, Phil. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, v. 8, pp.175-185, 1998.

EWING, R.; POIROT, O.; CLAVERIE, J.M. Comparative analysis of the *Arabidopsis* and rice expressed sequence tag (EST) sets. 1999. (*In press*)

FALCÃO, Martha A.; FERREIRA, Sidney A. N.; CHÁVEZ- FLORES, Wanders B.; CLEMENT, Charles R. Aspectos fenológicos e ecológicos do “camu-camu” (*Myrciaria dubia* (H.B.K.) Mc Vaugh) na terra firme da Amazônia Central. In: *Aspectos fenológicos, ecológicos e de produtividade de algumas na Amazônia brasileira: araca-boi (Eugenia stipitata Mc Vaugh), Biribá (Rollinia mucosa (Jacq.) Baill), camu-camu (Myrciaria dubia (H.B.K.) Mc Vaugh), cupuaçu (Theobroma grandiflorum (Willd. Ex Spreng.) Schum) e graviola (Annona muricata L.)*, pp. 57-65, 1993.

FRISTENSKY, Brian; BALCERZAK, Margaret; HE, Daiffen; ZHANG, Peijun. Expressed sequence tags from the defense response of ***Brassica napus*** to ***Leptosphaeria maculans***. *Mol. Plant Pathol. On-Line* [<http://www.bspp.org.uk/mppol/1999/0301FRISTENSKY>] acesso em 01/07/2002.

GERHOLD, D. & CASKEY, C.T. It's the genes! EST access to human genome content. *Bioessays*, v. 18, n.12, pp. 973-981, 1996.

GEMÜND, Christine; RAMU, Chenna; ALTERNBERG-GREULICH, Brigitte; GIBSON, Toby J. Gene2EST: a BLAST2 server for searching expressed sequence tag (EST) databases with eukaryotic gene-sized queries. *Nucleic Acids Research*, v.29, n.6, pp. 1272-1277, 2001.

HÖLFE, Herman; et al. An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* v. 4, pp. 1051-1061, 1993.

HUBER, Steven C.; HARDIN, Shane C. Numerous posttranslational modifications provide opportunities for the intricate regulation of metabolic enzymes at multiple levels. *Cur. Opin. Plant Biol.*, v. 7, pp. 318-322, 2004.

INVITROGEN - *Instruction Manual: SuperScript Plasmid System with Gateway Technology for cDNA Synthesis and Cloning*, 39p. (Cat. Series 18248047).

JAKO, Colette; COUTU, Catherine; ROEWER, Ingeborg; REED, Darwin W.; PELCHER, Lawrence E.; COVELLO, Patrick S. Probing carotenoid biosynthesis in developing seed coats of *Bixa orellana* (Bixaceae) through expressed sequence tag analysis. *Plant Science*, v. 163, pp. 141-145, 2002.

KALDE, Monika; BARTH, Meike; SOMSSICH, Imre E.; LIPPOK, Bernadette. Members of the Arabidopsis WRKY group III transcription factors are part of different plant defense signaling pathways. *Mol Plant Microbe Interact.*, v. 16, pp. 295-305, 2003.

KANG, Sang-Gu.; HANNAPEL, David J. Nucleotide sequences of novel potato (*Solanum tuberosum* L.) *MADS-box* cDNAs and their expression in vegetative organs. *Gene*, v. 166, pp. 329-330, 1995.

KIANG, Juliam G.; TSOKOS, George C. Heat Shock Protein 70 kDa: Molecular Biology, Biochemistry, and Physiology. *Pharmacol. Ther.*, v. 80, n.. 2, pp. 183–201, 1998.

LEE, Hyoshin; LEE, Jae-Soon; NOH, Eun-Woon; BAE, Eun-Kyung; CHOI, Yyoung-Im; HAN, Mu-Seok. Generation and analysis of expressed sequence tags from poplar (*Populus alba* x *P. tremula* var. *glandulosa*) suspension cells. *Plant Science*, v. 169, n. 6, pp. 1118–1124, 2005.

- LIM, Chae O.; KIM, Ho Y.; KIM, Min G.; LEE, Soo I.; CHUNG, Woo S.; PARK, Sung H.; HWANG, Inhwan H.; CHO, Moo J.; Expressed sequence tags of chinese cabbage flower bud cDNA *Plant Physiol.*, v. 111, pp. 577-588, 1996.
- LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. What is Bioinformatics? A Proposed Definition and Overview of the Field. *Method. Inform. Med.* v.40, pp. 346-58, 2001.
- MANN, Matthies; JENSEN, Ole N.; Proteomic analysis of post-translational modifications. *Nature Biotechnol.*, v. 21, pp. 255-261, 2003.
- MARRS, Kathleen A. The functions and regulation of glutathione s-transferases in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, v. 47, pp.127-58, 1996.
- NCBI – National Center for Biotechnology Information <<http://www.ncbi.nlm.nih.gov/>> acesso em 2004.
- OLSON, David C.; WHITE, Joseph A.; EDELMAN, Leonard; HARKINS, Richard N.; KENDE, Hans. Differential expression of two genes for 1-aminocyclopropane-1-carboxylate synthase in tomato fruits. *PNAS*, v.88, pp. 5340-5344, 1991.
- PENFIELD, Steven; JOSSE, Eve-Marie; KANNANGARA, Rubini; GILDAY, Alison D.; HALLIDAY, Karen J.; GRAHAM, Ian A. Cold and Light Control Seed Germination through the bHLH Transcription Factor SPATULA. *Curr. Biol.*, v. 15, pp. 1998-2006, 2005.
- PUHRINGER, Helene.; MOLL, Dieter; HOFFMANN-SOMMERGRUBER, Karin; WATILLON, Bernard; KATINGER, Hermann.; MACHADO, Margit L. C. The promoter of an apple *Ypr 10* gene, encoding the major allergen Mal d 1, is stress- and pathogen-inducible. *Plant Science*, v. 152, pp. 35-50, 2000.
- QUACKENBUSH, J.; LIANG, F.; HOLT, I.; PERTEA G.; UPTON, J. The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Research*, v.28, n. 1, pp. 141-145, 2000.
- QUACKENBUSH, J.; CHO, J.; LEE, D.; LIANG, F.; HOLT, I.; KARAMYCHEVA, S.; PARVIZI, B.; PERTEA G.; SULTANA, R.; WHITE, J. The TIGR gene indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Research*, v.29, n. 1, pp. 159-164, 2001.
- RONNING, C.M.; et al. Comparative analyses of potato expressed sequence tag libraries. *Plant Physiol.*, v. 131, pp. 419-429, 2003.
- SASAKI, Takuji; et al. Toward cataloguing all rice genes - large-scale sequencing of randomly chosen rice bud cDNAs from a callus cDNA library, *Plant J.*, v. 6, pp. 651-624, 1994.
- SAWBRIDGE, Tim; et al. Geration and analysis of expressed sequence tags in perennial ryegrass (*Lolium perenne* L.). *Plant Science*, v. 165, pp. 1089-1100, 2003.
- SHELTON, Dale; LEACH, David; BAVERSTOCK, Peter; HENRY, Robert. Isolation of genes involved in secondary metabolism from *Melaleuca alternifolia* (Cheel) using expressed sequence tags (ESTs). *Plant Science*, v. 162, pp. 9-15, 2002.
- SHIBATA, W.; BANNO, H.; ITO, Y.; HIRANO, K.; IRIE, K.; USAMI, S.; MACHIDA, C.; MACHIDA, Y. A tobacco protein kinase, NPK2, has a domain homologous to a domain found in activators of mitogen-activated protein kinases (MAPKKs). *Mol Gen Genet*, v. 246, pp. 401-410, 1995. (Abstract)
- SOUZA, Glaucia M.; SIMÕES, Ana C. Q.; OLIVERIA, Katia C.; GARAY, Humberto M.; FIORINI, Leonardo C.; GOMES, Felipe S.; NISHIYAMA-JUNIOR, Milton Y.; SILVA, Aline M. The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. *Genet. Mol. Biol.*, v. 24, n. 1-4, pp. 25-34, 2001.
- STERKY, Fredrik; et al. Gene discovery in the wood-forming tissues of poplar-analysis of 5.692 expressed sequence tags. *PNAS*, v. 95, pp. 13330-13335, 1998.
- SUSUZI, H.; ACHNINE L.; XU, R.; MATSUDA, S.P.T.; DIXON, R.A. A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *The Plant Journal*, v. 32, pp. 1033-1048, 2002.

UENISHI, Hirohide; EGUCHI, Tomoto; SUZUKI, Kohei; SAWAZAKI, Tetsuya; TOKI, Daisuke; SHINKAI, Hiroki; OKUMURA, Naohiko; HAMASIMA, Noriyuki; AWATA, Takashi. PEDE (Pig ESTS Data Explore): construction of a database for ESTs derived from porcine full-length cDNA libraries. *Nucleic Acids Research*, v. 32., pp. d484-d488, 2004.

WANG, T.F.; CHANG, J.H.; WANG, C. Identification of the peptide binding domain of hsc70 – 18 kilodalton fragment located immediately after atpase domain is sufficient for high affinity binding. *J. Biol. Chem.*, v. 268, n. 35, pp. 26049-26051, 1993.

VANEK-KREBITZ, Monika; HOFFMANN-SOMMERGRUBER, Karin; MACHADO, Margit L. C.; SUSANI, Markus; EBNER, Christof; KRAFT, Dietrich; SCHEINER, Otto; BREITENEDER, Heimo. Cloning and sequencing of Mal d 1, the major allergen from apple (*Malus domestica*) and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem. Biophys. Resear. Comm.*, v.214, n. 2, pp.538-551, 1995.

VETTORE, A.L.; DA SILVA, F.R.; KEMPER, E.L.; ARRUDA, P. The Library that made Sucest . *Gen. Mol. Biol.*, v.24, n.1-4, p.1-7, 2001.

VILLACHICA, H. *El cultivo del camu-camu (Myrciaria dubia (H.B.K.) Mc Vaugh) en la Amazonia Peruana*. Tratado de Cooperación Amazonia, Lima-Peru, 95p., 1996.

VINCENTZ, Michel; SCHLÖGL, Paulo S.; CORRÊA, Luis G. G.; KÜHNE, Fabiana; LEITE, Adilson. Phylogenetic relationships between Arabidopsis and sugarcane bZIP transcriptional regulatory factors. *Genet.Mol.Biol.*, v. 24, n.1-4, pp. 55-60, 2001.

YAMAMOTO, K.; SADAHI, T. Large-scale EST sequencing in rice. *Plant Mol. Biol.*, v. 35, pp. 135-144, 1997.

Capítulo 3

IDENTIFICAÇÃO DE POTENCIAIS GENES CODIFICANTES DE CHAPERONAS EM FRUTOS DE CAMU-CAMU

RESUMO

As chaperonas moleculares são responsáveis pelo dobramento das proteínas, união, translocação e degradação nas vias de processos celulares normais. Muitas moléculas chaperonas são proteínas de estresse e são originalmente identificadas como proteínas de choque térmico. O objetivo deste capítulo foi obter o perfil da expressão das chaperonas em frutos de camu-camu com auxílio da bioinformática a partir de seqüências ESTs. A biblioteca de EST de frutos de camu-camu armazenada no Sistema Genoma foi analisada quanto à presença de chaperonas molecular, as seqüências identificadas (*singlets* e o consenso dos *contigs*) foram submetidas aos programa ORF Finder do NCBI (*National Center for Biotechnology Information*) para identificação de orfs, e estas foram submetidas ao programa Blastp contra o banco de dados *nr*. As seqüências ESTs identificadas de genes potenciais codificantes de chaperonas moleculares foram analisadas quanto à homologia no programa de alinhamento LALING e anotado o percentual de identidade das seqüências. Foram identificadas 138 seqüências (4,3%) que codificam chaperonas moleculares das seguintes famílias: HSP100/Clp (1%), HSP90 (12%), HSP70 (24%), HSP70 co-chaperonas (6%), sHSP (33%), HSTF (1%), PPIase (7%), PDIase (7%), HSP60/*Chaperonin-like* (4%), *Chaperonin-like* co-chaperona (3%) e Calnexina/calreticulina (2%). O percentual de expressão das chaperonas em frutos de camu-camu pode demonstrar a adapção da espécie ao cultivo em terra-firme, entretanto, estudos mais aprofundados devem ser feitos neste sentido.

ABSTRACT

The molecular chaperones are responsible by proteins folding, union, translocation and degradation in the roads of normal cellular process. Many chaperones molecular are stress proteins and are identified originally like heat shock proteins. The aim of this chapter was obtain the expression profile of chaperones in camu-camu fruits with bioinformática aid from ESTs sequences. The EST library of camu-camu fruits stored in the Genome System was analyzed as regards the presence molecular chaperones, the identified sequence (singlets and contig consensus) were submitted to ORF Finder program of the NCBI (National Center for Biotechnology Information) for orfs identification, these were submitted to the Blasp program against the nr database. The ESTs sequences identified of gens potential coding of molecular chaperone were analyzed as regards the ratified in the alignment LALING program and noted the identity percentage. They were 138 sequences identified (4,3%) that they encode following molecular chaperones family: HSP100/Clp (1%), HSP90 (12%), HSP70 (24%), HSP70 co-chaperonas (6%), sHSP (33%), HSTF (1%), PPIase (7%), PDIase (7%), HSP60/Chaperonin-like (4%), Chaperonin-like co-chaperone (3%) and Calnexin/calreticulin (2%). The expression percentage of chaperones in camu-camu fruits can show to specie adaptation in *terra-firme* cultivation, however, more deepened studies should be deeds in these sense.

1. INTRODUÇÃO

As chaperonas moleculares são componentes chave para a homeostase celular sob condições ótimas e adversas de crescimento. Elas também são responsáveis pelo dobramento das proteínas, união, translocação e degradação nas vias de processos celulares normais. Muitas moléculas chaperonas são proteínas de estresse e são originalmente identificadas como proteínas de choque térmico (*Heat shock protein - HSP*) (WANG et al., 2004).

As proteínas de choque térmico são sintetizadas pela célula em resposta a temperaturas elevadas, bem como por outros estresses ambientais. As Hsp/chaperonas também são expressas nas plantas em resposta a estresse hídrico, salino, osmótico, de temperatura e oxidativo. Embora seja conhecido o papel de Hsp/chaperonas em outros organismos, pouco se conhece sobre sua expressão em plantas (WATERS et al., 1996; WANG et al., 2004).

As moléculas Hsps/chaperonas são localizadas tanto no citoplasma como nos diferentes compartimentos celulares, tais como núcleo, mitocôndria, cloroplastos e retículo endoplasmático (WATERS et al., 1996; BOSTON et al., 1996; VIERLING, 1991; WANG et al., 2004). Elas não fazem ligações covalentes com seus alvos e não fazem parte do produto final (BUKAU & HORWICH, 1998). Muitas proteínas celulares, se não todas, interagem com as chaperonas durante seu tempo de vida celular. Diferentes sistemas de chaperonas são requeridos para a síntese, maturação e degradação de proteínas em todos os compartimentos celulares (BOSTON, et al., 1996).

As chaperonas moleculares formam um grupo diverso de proteínas, que compartilham as propriedades de se ligarem a substratos protéicos desnovoado. As cinco famílias mais importantes de Hsps/Chaperonas são: HSP70 (DnaK); Chaperoninas (GroEL e Hsp60); Hsp90; Hsp100 (Clp); e pequenas Hsp (sHsp). Além

destas famílias, existem outras proteínas com funções de chaperonas, tal como Proteína disulfeto isomerase (PDI), Peptidil prolil *cis-trans* isomerase (PPIase), Fator de transcrição de choque térmico (*Heat shock transcription factor - HSTF*) e calnexina/calreticulina, entre outras (VIERLING, 1991; WATERS et al., 1996; BOSTOR, et al., 1996).

As chaperonas Hsp100 são similares à seqüência ClpA de *Escherichia coli*, as quais tem clara função reguladora da protease ATP-dependente de Caseinolítica protease (Clp). As proteínas Hsp100/Clp são uma família com uma grande variedade de funções, tais como aumentam a tolerância a elevadas temperaturas, promover proteólise de substratos celulares específicos e regulam a transcrição. Alguns dados sugerem uma capacidade comum de desenovelar grandes estruturas proteicas e agregados, unificando as funções moleculares desta família (SCHIRMER et al., 1996).

Em plantas, vários genes de homólogos a Clp foram clonados, como por exemplo, em soja (LEE et al., 1994), *Arabidopsis* (SCHIRMER et al., 1994), milho (NIETO-SOTELO et al., 1999), arroz (KATIYAR-ARGAWAL et al., 2003) e tabaco (ARGAWAL et al., 2001). Em *Arabidopsis*, arroz, milho, soja e tabaco as Hsp100 são induzidas pelo estresse térmico (LEE et al., 1994; SCHIRMER et al., 1994; PAREEK et al., 1995; NIETO-SOTELO et al., 1999).

As proteínas de choque térmico com 90KDa (Hsp90) são chaperonas moleculares abundantes e altamente conservadas que são essenciais para a viabilidade celular em eucariotos. As Hsp90 tem papel chave no dobramento, ativação de proteínas cliente, possivelmente envolvido em sinal de transdução tal como receptores de esteróides e uma variedade de proteínas quinases e no controle do ciclo celular (TOFT 1998; PEARL E PRODRMOU 2000).

As Hsp90 funcionam num complexo com várias co-chaperonas moduladas pela atividade ATPase e pela interação com outras proteínas, por exemplo, a co-chaperona

HOP (proteína organizada HSP70/HSP90) é envolvida no dobramento de proteínas clientes imaturas e inibindo a atividade da ATPase de HSP90. Quando a HSP90 é ligada ao ATP, HOP é trocada pela p23 e pelas imunofilinas/prolil-isomerases (YOUNG et al., 2001; PICARD et al., 2002; WEGELE et al., 2004; SANGSTER & QUEITSCH, 2005).

Sete isoformas de Hsp90 existem em *Arabidopsis*: quatro são relatadas como isoformas citoplasmáticas (Hsp90,1-Hsp90.4); uma é cloroplastídica (Hsp90.5); uma é mitocondrial (Hsp90.6); e uma está localizada no retículo endoplasmático (Hsp90.7) (KRISHNA & GLOOR, 2001). Das isoformas citoplasmáticas, uma é induzida pelo calor, enquanto as outras são constitutivamente expressas (SANGSTER & QUEITSCH, 2005).

Quanto a família Hsp70 de proteínas de choque térmico, estas incluem proteínas citoplasmáticas induzidas pelo calor (HSP70), proteínas expressas constitutivamente (HSC70) e proteínas de 78 kDa reguladas pela glicose (GRP78) (EFREMOVA et al., 2002).

As Hsp70 tem dois domínios funcionais, um de ligação ao ATP e o outro de ligação ao substrato. O domínio ligado ao substrato contém dois sub-domínios: o N-terminal que é enovelado num β -sanduíche compacto; e o sub-domínio C-terminal que é composto de cinco alfa-hélices estendidas que são provavelmente responsáveis pelas mudanças conformacionais que ocorrem durante a ligação e liberação do substrato (ZHU et al., 1996, apud ZHANG & GLASER, 2002).

Em mitocôndrias, a chaperona Hsp70 funciona como um motor molecular, empurrando o precursor que atravessa as membranas mitocôndriais. Em geral, 97% das mitocôndrias vegetais contem sítio de ligação de pré-sequências de Hsp70. Nos cloroplastos, a Hsp70 do envelope externo, do espaço intermembrana e do estroma são proteínas importadas do citoplasma; sendo 82,5% dos cloroplastos possuem um peptídeo de trânsito ligado a Hsp70 (ZHANG & GLASER, 2002).

Os estudos da especificidade da proteína homologa a Hsp70 (DnaK bacteriana, Hsc70 citosólica e a Bip do retículo endoplasmático) mostraram que o substrato se liga a proteína Hsp70 preferencialmente nos aminoácidos hidrofóbicos (Ile, Leu, Val, Ala, Phe, Trp e Tyr) e positivamente carregados (Arg e Lys) mas não a resíduos acídicos (ZHANG et al., 1999; ZHANG & GLASER, 2002).

A Hsp70 tem um papel chave no dobramento das proteínas. O mecanismo do enovelamento auxiliado pela proteína Dnak é bem entendido. A DnaK funciona em cooperação com o DnaJ e as chaperonas GrpE, que são proteínas de choque térmico 40 kDa e 20 KDa, respectivamente. No ciclo de reação da DnaK, a primeira DnaJ se une um peptídeo desenovelado e então o substrato se desloca para o sítio de ligação de ATP para formar um complexo DnaJ-substrato-DnaK-ATP (ZHANG & GLASER, 2002).

As chaperonas moleculares melhor estudadas, as chaperoninas, consistem em dois membros: GroEL ou chaperonin 60 (Cpn60) e GroES ou chaperonin 10 (Cpn10) (HIROHASHI et al., 1999, p. 512). Chaperoninas, em geral, estão localizadas principalmente no citosol em bactéria, nos cloroplastos em plantas e nas mitocôndrias em animais. Formam estruturas multiméricas abrangendo um par empilhado de anéis contendo sete subunidades cada (WU et al., 2000).

As Cnp60 pertencem ao grupo ubíquo de chaperonas que compreende as proteínas de choque térmico, nucleoplasminas e chaperoninas; são proteínas altamente conservadas presentes em diferentes organismos, de bactéria à mamíferos (WU et al., 2000, p. 156) A análise estrutural e bioquímica revelou que a Cpn60 é formada por dois anéis simétricos de sete dobras cada uma (HIROHASHI et al., 1999).

A Cpn10 também é formada por um anel de estrutura homo-oligomérica de sete-subunidades associadas com o aligômero da Cnp60 ATP-dependente durante uma reação de dobramento mediada pela Cpn60-Cpn10 (HIROHASHI et al., 1999)

Os cloroplastos de plantas superiores contêm uma proteína de 21kDa, chaperonina 21 (Cpn 21), que é uma homóloga funcional da chaperonina 10 (Cpn 10), e que consiste de dois domínios Cpn10 fundidos em bloco (HIROHASHI et al., 1999).

As pequenas HSP (*small HSP - sHSP*) são definidas pelo domínio C-terminal conservado de aproximadamente 90 aminoácidos, chamado domínio α -cristalino, que é flanqueado por uma extensão C-terminal curta e um não conservado braço N-terminal de comprimento variável. As sHSP variam entre 15 a 40kDa de massa molecular e formam grandes complexos de 9 a 32 subunidades oligoméricas (BASHA et al., 2004).

As sHSP são muito eficientes na ligação de proteínas desnaturadas, e modelos atuais propõem que elas funcionam para prevenir irrevesivelmente a agregação e a insolubilização protéica, aumentando assim a resistência celular ao estresse (BASHA et al., 2004).

Em plantas, as sHSP são expressas no citoplasma e isoformas específicas ligadas a organelas intracelulares. Há dois tipos de sHSP no citoplasma, referidas como proteínas da classe I e II, que compartilham aproximadamente 50% de identidade do domínio α -cristalino. Cinco subfamílias de genes codificam sHSP localizadas na mitocôndria, plastídio, perossomos, núcleo e no retículo endoplasmático, cada uma com seu sinal para a organela alvo (WATERS & VIERLING, 1999; BASHA et al., 2004).

O seqüenciamento parcial de ESTs de frutos de camu-camu possibilita, com o auxílio da bioinformática, obter o perfil genético deste organismo quanto a presença de chaperonas.

2. MATERIAIS E MÉTODOS

A biblioteca de EST de frutos de camu-camu armazenada no Sistema Genoma foi analisada quanto à presença de chaperonas molecular seguindo a seguinte estratégia de anotação (Figura 01): as seqüênciaa identificadas como chaperones (*singlets* e o consenso dos *contigs*) foram submetidas ao programa ORF Finder do NCBI (*National Center for Biotechnology Information*) para identificação orfs, estas foram submetidas ao programa Blastp contra o banco de dados nr.

As seqüências ESTs identificadas de genes potenciais codificantes de chaperonas moleculares foram analisadas quanto à homologia no programa de alinhamento LALING (http://www.ch.embnet.org/software/lalign_form.html) e anotado o percentual de identidade das seqüências.

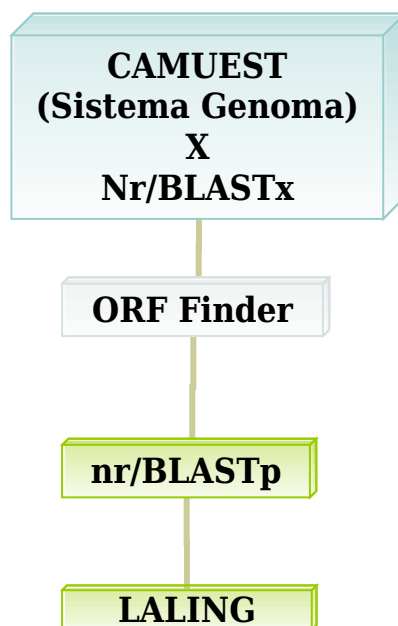


Figura 01 – Esquema da estratégia de anotação de chaperonas molecular da biblioteca de ESTs de camu-camu armazenada no programa Sistema Genoma (BIOFOCO).

3. RESULTADOS E DISCUSSÃO

O banco de seqüências ESTs de camu-camu contém, até o presente momento, 3192 seqüências ESTs e destas foram identificadas 138 seqüências (4,3%) que codificam chaperonas moleculares das seguintes famílias: HSP100/Clp (1%), HSP90 (12%), HSP70 (24%), HSP70 co-chaperonas (6%), sHSP (33%), HSTF (1%), PPIase (7%), PDiase (7%), HSP60/*Chaperonin-like* (4%), *Chaperonin-like* co-chaperona (3%) e Calnexina/calreticulina (2%) (Figura 2).

As chaperonas são relativamente abundantes em frutos de camu-camu se comparado ao percentual de chaperonas encontrados na cana-de-açúcar (SUCEST) com 1,8% das 4.164 seqüências ESTs analisadas (BORGES et al., 2001).

Foi identificada uma seqüência da família HSP100/Clp com similaridade a HSP101 (Classe I – tipo B) com 74,0% de identidade (Tabela 1). A família HSP100/Clp chaperones têm a capacidade de promover dissociação de proteínas agregadas ATP-dependente (PARSELL et al., 1994), sua expressão é regulada pelo desenvolvimento em plantas (QUEISTCH et al., 2000). Segundo Zhang e Guy (2005), a Hsp101 está presente no citoplasma de células vegetais formando um complexo com Hsp70. Em *Arabidopsis* foram identificadas Hsp101 citoplasmáticas e plastidiais (AGARWAL et al., 2001).

O acúmulo da Hsp101 em sementes madura de trigo, milho, *Brassica juncea* e arroz sugere que esta proteína tenha um papel importante no seu desenvolvimento (SINGLA et al., 1998). Não foi observada deficiência de Hsp101 em plantas com crescimento e desenvolvimento em temperaturas normais. A expressão da Hsp100 é regulada pelo desenvolvimento em plantas (QUEISTCH et al., 2000).

Da classe HSP90 foi identificada uma seqüência EST similar a HSP90 com 83,2% de identidade e uma seqüência similar a *p23 co-chaperone* com 69,4% de identidade. A

classe HSP83 apresentou nove seqüências similares a HSP83 com 70,7% de identidade (tabela 01).

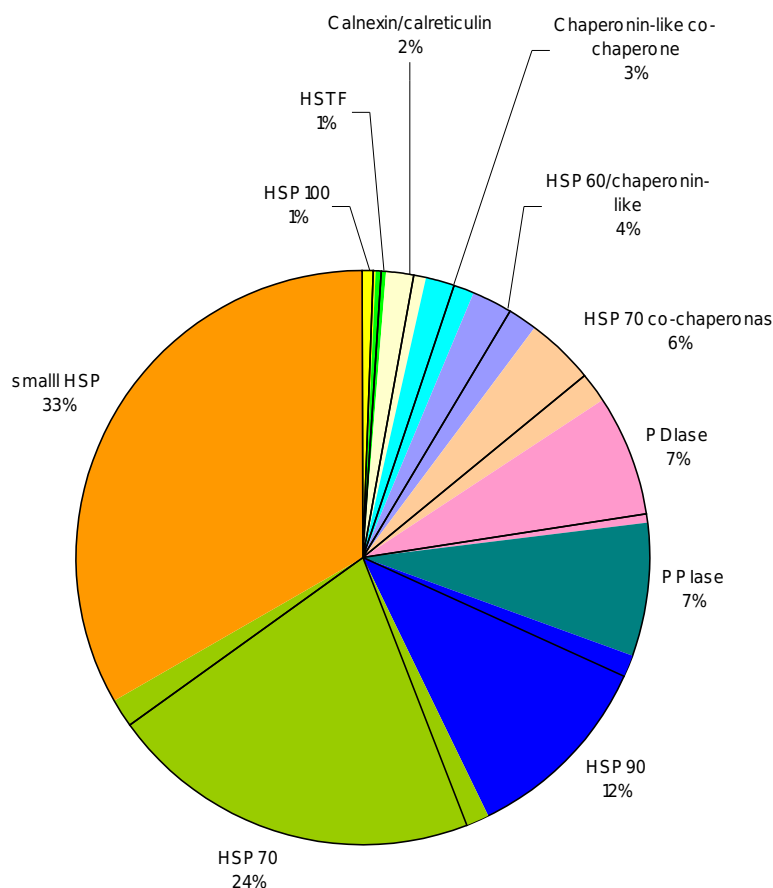


Figura 02 – Percentual de genes potenciais codificantes de famílias de chaperones, anotadas de seqüências ESTs de frutos de camu-camu. *HSP* = *heat shock protein*; *sHSP* = *small heat shock protein*; *HSPTF* = *heat shock transcription factor*; *PDIase* = *Peptidyl-disulfide isomerase*; *PPIase* = *Peptidyl-prolyl cis, trans isomerase*.

A HSP82 foi similar com duas seqüências ESTs com 79,3%, Na classe HSP81 três seqüências ESTs apresentaram similaridade a HSP81-1 ligadao ao ATP/ ligada a proteína não-enoveladas e uma provável HSP81-2, com 72,5% e 83,0% de identidade, respectivamente. Uma proteína regulada por glicose Grp94 apresentou 76% de similaridade a um ESTs de camu-camu (Tabela 01).

Genes de Hsp90 tem sido isolados de várias espécies de planta (CONNER et al 1990; FELSHEIM & DAS 1992; KONING et al 1992; TAKAHASHI et al 1992; MARRS et al 1993; YABE et al 1994; KRISHNA et al 1995). O sistema Hsp90 é de interesse particular em plantas, primeiro por causa de um compartimento subcelular adicional, o plastídio, e segundo porque plantas são submetidas a mudanças repentinas em seu ambiente, que invocam respostas moleculares rápidas. Uma proteína 94 KDa regulada por glicose, Grp94, que é 50% idêntica a Hsp90 citoplasmática, reside no retículo endoplasmático (ER). O complexo chaperona HSP90 é implicado nos mecanismos de plasticidade fenotípica a condições adversas ambientais (SANGSTER & QUEITSCH, 2005).

A segunda maior familiar, HSP70, apresentou as classes HSP70 (7 ESTs) e HSC70 (24 ESTs) (Tabela 01). A classe HSP70 apresentou três ESTs similares a HSP70 com 77,5% de identidade, um EST similar uma proteína de interação a HSP70/tioredoxina quimérica, com 74,4% de identidade, um EST com similaridade uma provável HSP70, com 82,0% de similaridade, um EST similar a uma HSP70 T-2 de ligação ao ATP com 73,9% de identidades, e um EST similar a Hsp70 do tipo DnaK, com 85,2% de identidade (Tabela 01).

As proteínas de choque térmico 70 KDa (Hsp70) funcionam como chaperonas moleculares em uma variedade de processos celulares, prevenindo agregação protéica e o redobramento de proteínas não-nativas sob condições normais e de estresse ambiental (WANG et al., 2004). Ajudam a célula a enfrentar uma variedade de condições ambientais desfavoráveis, tais como, calor, frio, seca, estresse químico e outros.

CHAPERONAS MOLECULARES		ANOTAÇÃO DE GENES POTENCIAIS	
Classe	Membro	Nº. de ESTs	Identidade (%)
Classe I - Tipo			
B	HSP101	1	74,0
HSP90	HSP90	1	83,2
	Provável p23 co-chaperone	1	69,4
HSP83	HSP83	9	70,7*
HSP82	HSP82	2	79,3*
HSP81	HSP81-1; ligada ao ATP/proteína de desdobraimento	2	72,5*
	Provável HSP81-2	1	83,0
Grp94	Grp94	1	76,0
HSP70	HSP70	3	77,5*
	Proteína HSP70 de interação/tioredoxina químera	1	74,4
	Provável HSP70	1	82,0
	HSP70 T-2 ligado ao ATP	1	73,9
	HSP70 tipo DnaK	1	85,2
HSC70	HSC70	6	76,5*
	HSC70-1	15	80,7*
	HSC70-3	3	83,4*
HSP40/DnaJ	Provável proteína DnaJ de choque térmico contendo domínio N-terminal	2	73,0*
	DnaJ proteína de choque térmico	1	57,4
	Proteína tipo DnaJ	3	66,4*
	DnaJ homóloga	1	80,4
Cpn60	Chaperonina 60 subunidade alfa	3	79,0*
	Provável Chaperonina 60 precursora beta	1	79,2
	Chaperonina subunidade citosólica delta	1	80,3
Cpn21	Chaperonina 21 cloroplastídica	3	77,2*
sHSP CI	HSP21	1	60,9
	sHSP17.7 CI	9	73,0*
	sHSP18 CI	1	71,3*
	sHSP18.5 CI	14	68,3*

	sHSP17,3	1	69,9
	SHSP citoplasmática classe I	1	60,5
sHSP CII	sHSP17.8 CII	1	65,3
	sHSP17.6 CII	3	69,2*
sHSP CIII	HSP20.1 CIII	1	79,3
	HSP26 CIII	2	64,9*
sHSP CIV	HSP22.3 CIV	1	55,9
	sHSP22 CIV	6	62,5*
HSF	Fator de transcrição de choque térmico 29	1	62,7
FK506-BP	Ligada a FK506/ peptidil-prolil <i>cis-trans</i> isomerase/ TWD1 (<i>TWISTED DWARF 1</i>)	1	79,4
	FKBP-tipo peptidil-prolil <i>cis-trans</i> isomerase	1	63,6
	Provável peptidil-prolil <i>cis-trans</i> isomerase	1	77,4
	Provável peptidilprolil isomerase 70 kDa	1	74,0
Cyclophilin	Ciclofilina	5	76,1*
	Provável ciclofilina (70.8 KD) (Cyp-15)	1	74,4
PDI	PDI precursora (PDI)	1	80,0
	Proteína tipo PDI	8	66,3*
	Provável proteína relatada como disulfeto isomerase	1	70,7
Calnexin	Proteína tipo calnexina	2	70,0*
	Proteína de interação com calreticulina	1	77,2
TOTAL		138	

Cnp = chaperonin; HSP = heat shock protein; HSC = heat shock cognate; sHSP = small heat shock protein; HSPTF = heat shock transcription factor; PDIase = Peptidyl-disulfide isomerase; PPIase = Peptidyl-prolyl *cis, trans* isomerase.* valores médios.

Tabela 01 – Anotação de genes potenciais codificantes de chaperonas molecular em frutos de camu-camu.

As Hsp70 facilitam a translocação de correntes nascentes através da membrana e mediante a reunião e separação de complexos de proteínas multiméricas e apontam a degradação de proteínas pelos lisossomos ou proteossomos. As Hsp70 pertencem a uma família de proteínas que são altamente conservadas entre os organismos e entre diferentes organelas celulares no mesmo organismo (ZHANG & GLASER, 2002)

Um total de 24 seqüências se mostraram similares às proteínas de choque térmico cognata: seis com a HSC70 com 76,5%; 15 ESTs com a HSC70-1 com 80,7%; e três ESTs com HSC70-3 com 83,4% (Tabela 01). Segundo Wang et al. (2004), as HSC70 estão freqüentemente envolvidas como auxiliares no dobramento de novos polipeptídios sintetizados e a importação/translocação de precursores de proteínas.

As HSP70 co-chaperonas anotadas foram similares a classe HSP40/DnaJ com dois ESTs relacionados a uma provável proteína com domínio DnaJ HS, com 73,0% de identidade, uma seqüências similar a uma proteína de choque térmico DnaJ, com 57,4% de identidade, três ESTs relacionados a uma proteína tipo DnaJ, com 66,4% de identidade e uma seqüência similar a DnaJ homóloga com identidade de 80,4% (Tabela 01).

Segundo Borges et al. (2001) as mais freqüentes chaperonas encontradas no banco de genes de cana-de-açúcar foram das famílias da HSP70 co-chaperones (22%) e HSP70 (21%). Duas proteínas tipo DnaJ, uma de *Nicotiana tabacum* (tabaco) e uma de *Arabidopsis thaliana* tem sido identificadas como proteínas de interação (BARGEN et al., 2001).

Em relação as Hsp60/Chaperoninas, cinco ESTs foram similares a classe Cnp60: três com a subunidade alfa da Cnp60, com 79,0% de identidade; uma seqüência similar a uma provável precursora da subunidade beta da Cnp60, com 79,2% de identidade e uma seqüência similar a subunidade delta de uma chaperonina citosólica, com 80,3% de identidade (Tabela 01).

Em cana-de-açúcar, de 4.164 seqüências ESTs analisados 3% eram relacionados a Hsp60/Chaperonina (BORGES et al., 2001), percentual inferior ao encontrado na biblioteca do camu-camu, 4% (Figura 02). Bonk et al. (1996) relataram o aumento na abundância de chaperonina 60 na diferenciação do cloroplasto em cromoplastos durante a formação da flor de *Narcissus pseudonarcissus*.

Da classe Cpn21 foram identificadas duas proteínas: uma chaperonina 21 cloroplastídica similar a três ESTs de camu-camu com 77,2% de identidade e uma Hsp21 com 60,9% de identidade (Tabela 01). Hirorashi et al. (1999) isolaram o cDNA de Cpn21 precursor de *Arabidopsis thaliana* e demonstraram sua elevada expressão em cloroplastos. Em cana-de-açúcar 29 ESTs foram anotados como codificantes de Cpn21 em cloroplasto (BORGES et al., 2001).

As sHSP representaram o maior grupo de chaperonas identificadas (33%) (Figura 2). As proteínas de choque térmico de baixo peso molecular anotadas pertencem às classes: I, II, III e IV. As seqüências ESTs relacionadas a classe I foram similares as chaperonas sHSP17,3 (69,9% identidade) sHSP17,7 (73,0% identidade), sHSP18 (71,3% identidade), sHSP18,5 (68,3% identidade) e a sHSP citoplasmática classe I (60,5% identidade) (Tabela 1).

Em relação classe II, quatro ESTs foram anotados com similaridade a sHSP17,6 (69,2% identidade) e sHSP17,8 (65,3% identidade); três ESTs identificadas foram similares sHSP20,1, sHSP26 da classe III, com 79,3%; 64,9% de identidade, respectivamente; quanto a classe IV, sete seqüências foram similares as chaperonas sHSP22 (62,5% identidade) e sHSP22,3 (55,9% identidade) (Tabela 01).

Waters et al. (1996) relataram que as sHSP são induzidas pelo calor e que se apresentam diversas e abundantes em plantas. Entretanto, Borges et al. (2001) anotaram apenas 4,7% de seqüências sHSP em cana-de-açúcar indicando que não são muito abundantes nesta espécie. Em frutos de tomate, também foram identificados

cDNA que codificam sHSP21 e sHSP17,4 (LAWRENECE et al., 1997; MEDINA-ESCOBAR et al., 1998)

A HSP20/K-cristalina é uma das seis maiores famílias de chaperonas molecular, também referidas como pequenas proteínas de choque térmico (*small heat shock protein* - sHsps). Os membros desta família compartilham uma região conhecida como centro cristalino K, que está localizado no domínio C-terminal. Muitas destas proteínas de choque térmico funcionam como chaperonas moleculares, suprimindo a agregação ou ajudando no redobramento proteínas parcialmente desnaturadas (GARCIA-RANEA et al., 2002).

Da família HSTF uma seqüência foi similar ao fator transcrição de choque térmico 29 com 62,7% de identidade (Tabela 01). Os fatores de transcrição de choque térmico (HSTF) apresentam um papel central na expressão de HSP estresse-dependente em plantas (BORGES et al., 2001; PANCHUK et al., 2002). Em *Arabidopsis*, mais de 20 genes diferentes de HSTF já foram identificados em seqüências de DNA (NOVER et al., 2001)

Em plantas, múltiplas HSFs tem sido clonadas de tomate (SCHARF et al., 1990), *Arabidopsis* (HÜBEL & SCHÖFFL, 1994), grão de soja (CZARNECKA-VERNER et al., 1995) e milho (GAGLIARDI et al., 1995). Baseado em análise filogenética de seus domínios ligados ao DNA e uma comparação de domínios de oligomerização eles tem sido designados especializar-se nas classes A e B (CZARNECKA-VERNER et al., 2000; NOVER & SCHARF, 1997).

Outras diferenças entre as classes A e B de HSFs inclui a posição relativa da seqüência de localização nuclear (*nuclear localization sequence* - NLS), e o presença ou ausência da região hidrofóbica C-terminal (*C-terminal hydrophobic region* - HR-C), que em vertebrados foi mostrado estar envolvida em regulação do estado de trimerização da HSF. Além do mais, a classe B de HSFs tem domínios C-terminais curto

e com a exceção LpHSFB1 do tomate (LpHSF24), parece faltar o motivo funcional da ativação transcricional (AHA ou TRP repete) que estão presentes em classe A de HSFs (CZARNECKA-VERNER et al., 2000)

A distribuição comum destas duas classes entre plantas superiores e suas características distintas sugerem que esta dicotomia nos tipos de HSF possa refletir uma especialização funcional que pode ser única em planta (CZARNECKA-VERNER et al., 2000)

Da família PPIase foram anotadas 10 seqüências, seis similares a classe ciclofilina: cinco de Ciclofilinas – 76,1% identidade; e uma provável ciclofilina (70,8k kD) (Cyp-15) (74,4% identidade); e quatro similares a classe FK506-BP – uma provável peptidil-prolil *cis-trans* isomerase (77,4% identidade), uma provável peptidil-prolil isomerase 70 kDa (74,0% identidade), uma FK506 ligada/peptidil-prolil *cis-trans* isomerase/TWD1 (TWISTED DWARF 1) (79,4% identidade) e uma calmodulina com 70% de identidade e uma FKBP-tipo peptidil-prolil *cis-trans* isomerase (63,6% identidade) (Tabela 01).

As PPIases catalisam a isomerização do peptídeo ligado entre uma prolina e o resíduo do peptídeo. A inter-conversão *cis-trans* acelerada pela PPIases é significativa para a estrutura final da proteína porque prolina *cis* introduz curvas dentro da proteína. FK506. Embora muita das funções celulares de diferentes FKBP's sejam desconhecidas, resultados recentes indicam que estas proteínas são reguladores magistrais no controle do desenvolvimento em plantas e animais (HARRAR et al., 2001; BREIMAN & CAMUS, 2002).

Nas plantas, as *cis-trans* prolil isomerases (PPIases) pertencentes as ciclofilinas, FKBP's e parvulinas tem sido clonadas (BOSTON et al., 1996; LANDRIEU et al., 2000). O ciclofilinas foram achadas no citosol, cloroplastos e mitocôndria (GASSER et al., 1990; BREIMAN et al., 1992; LIPPUNER et al., 1994; FULGOSI et al., 1998).

As proteínas ligadas a FK506 de planta tem sido isoladas de várias espécies tais como a KBP73 de trigo (BLECHER et al., 1996), FKBP15 de *Arabidopsis* e *Vicia faba* (LUAN et al., 1996), FKBP66 de milho (HUEROS et al., 1998), e o FKBP77 de trigo induzida pelo calor (KUREK et al., 1999). As FKBP de planta são dirigidos de diferentes organelas celulares.

As PDiases anotadas foram similares a uma precursora de PDI (80% identidade), uma proteína tipo PDI com 66,3% de identidade e oito prováveis proteínas relatadas como disulfeto isomerase (70,7% de identidade). Da família Calnexin/Calreticulín foram identificadas três seqüências, duas com similaridade a uma proteína tipo calnexina com 70,0% de identidade e uma proteína que interage com calreticulina com 77,2% de identidade (Tabela 01).

Estas duas famílias (PPIase e Calnexina/Calreticulina) estão envolvidas no dobramento de proteínas do retículo endoplasmático e são muito importante em cana-de-açúcar pois as três juntas representam 23,8% de chaperonas identificadas (BORGES et al., 2001). Entretanto, em camu-camu sua expressão é baixa (7%).

4. CONCLUSÃO

- Os dados aqui apresentados mostram que genes de chaperonas e proteínas relacionadas ao estresse são frequentes nas ESTs seqüenciadas, refletindo possivelmente no nível de mRNA expressos e tem ampla diversidade em frutos de camu-camu.
- As chaperonas mais abundantes e diversas foram sHSP, seguida pelas HSP70 e HSP90. Esses resultados também indicam a presença de outras famílias de chaperonas nos frutos de camu-camu: HSP100, HSP70 co-chaperones, PPIase, PDIase, HSTF e Calnexina/Calreticulina, porém com baixos percentual de expressão.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- AGARWAL, Manu; KATIYAR-AGARWAL, Surekha; SAHI, Chandan; GALLIE, Daniel R.; GROVER, Anil. *Arabidopsis thaliana* Hsp100 proteins: kith and kin. *Cell Stress & Chaperones*, v. 6, n.3, pp. 219-224, 2001.
- BARGEN, Susanne von; SALCHERT, Klaus; PAAPE, Martina; PIECHULLA, Birgit; KELLMAN, Jan-Wolfhard. Interactions between the tomato spotted wilt virus movement protein and plant proteins showing homologies to myosin, kinesin and DnaJ-like chaperones. *Plant Physiol. Biochem.* v. 39, pp. 1083–1093, 2001.
- BASHA, Emam; LEE, Garrett J.; DEMELER, Borries; VIERLING, Elizabeth. Chaperone activity of cytosolic small heat shock proteins from wheat. *Eur. J. Biochem.*, v. 271, p. 1426-1436, 2004.
- BLECHER, Oshra Blecher; EREL, Noa; CALLEBAUT, Isabelle; AVIEZER, Keren Aviezer; BREIMAN, Adina. A novel plant peptidyl-prolyl-*cis-trans*-isomerase (PPIase): cDNA cloning, structural analysis, enzymatic activity and expression. *Plant Mol. Biol.*, v. 32, n. 3, pp. 493-504, 1996. (Abstract)
- BONK, Michael; TADROS, Monier; VANDEKERCKHOVE, Joel; AL-BABILI, Salim; BEYER, Peter. Purification and characterization of chaperonin 60 and heat-shock protein 70 from chromoplasts of *Narcissus pseudonarcissus*. *Plant Physiol.*, v. 111, pp. 931-939, 1996.
- BORGES, Júlio C.; PEROTO, Maria C.; RAMOS, Carlos H. I. Molecular chaperone genes in the sugarcane expressed sequence database (SUCEST). *Genetics and Molecular Biology*, v. 24, n. 1-4, pp. 85-92, 2001.
- BOSTON, R. S.; VIITANEN, P. V.; VIERLING, E. Molecular chaperones and protein folding in plants *Plant Mol. Biol.* v. 32, n. 1-2, p. 191 - 222 , 1996.
- BREIMAN, Adina; CAMUS, Isabelle. The involvement of mammalian and plant FK506-binding proteins (FKBPs) in development. *Transgenic Research*, v. 11, pp. 321–335, 2002.
- BREIMAN, Adina; FAWCETT, Timothy W.; GHIRARDILL, Maria L.; MATTOOL, Autar K. Plant organelles contain distinct peptidylprolyl *cis*, *trans*-isomerases. *J. Biol. Chem.*, v. 267, n. 30, pp. 21293-21296, 1992.
- BUKAU, Bernd; HORWICH, Arthur L. the Hsp70 and Hsp60 chaperone machines. *Cell*, v. 92, p. 351-366, 1998.
- CONNER, Timothy W.; LAFAYETTE, Peter R.; NAGAO, Ronald T.; KEY, Joe L. Sequence and expression of a HSP83 from *Arabidopsis thaliana*. *Plant Physiol*, v. 94, pp. 1689–1695, 1990.
- CZARNECKA-VERNER, Eva; YUAN, Chao-Xing; SCHARF, Klaus-Dieter; ENGLICH, Gisela; GURLEY, William B. Plants contain a novel multi-member class of heat shock factors without transcriptional activator potential. *Plant Mol. Biol.*, v.43, pp. 459-471, 2000.
- CZARNECKA-VERNER, Eva; YUAN, Chao-Xing; FOX, P. C.; GURLEY, William B. Isolation and characterization of six heat shock transcription factor cDNA clones from soybean. *Plant Mol Biol.* v. 29, n. 1, p. 37, 1995 (Abstract).
- EFREMOVA, Sofia M.; et al. Heat shock protein Hsp70 expression and DNA damage in Baikalian sponges exposed to model pollutants and waste water from Baikalsk pulp and paper plant. *Aquatic Toxicology*, v. 57, pp. 267-280, 2002.
- FELSHEIM, Roderick F.; DAS, Anath. Structure and expression of a heat-shock protein 83 gene of *Pharbitis nil*. *Plant Physiol.*, v. 100, pp. 1764-1771, 1992.
- FULGOSI, Hrvoje; VENER, Alexander V.; ALTSCHMIED, Lothar; HERRMANN, Reinhold G.; ANDERSSON, Bertil. A novel multi-functional chloroplast protein: identification of a 40 kDa immunophilin-like protein located in the thylakoid lumen. *EMBO Journal*, v.17, n.6, pp.1577–1587, 1998.

GARCIA-RANEA, J. A.; MIREY, Gladys; CAMONIS, Jacques; VALENCIA, Aalfonso. p23 and HSP20/ α -crystallin proteins define a conserved sequence domain present in other eukaryotic protein families. *FEBS Letters*, v. 529, pp.162-167, 2002.

GAGLIARDI, Dominique; BRETON, Christian; CHABOUD, Annie Chaboud; VERGNE, Philippe; DUMAS, Christian. Expression of heat shock factor and heat shock protein 70 genes during maize pollen development. *Plant Mol. Biol.*, v. 29, n. 4, p. 841, 1995. (Abstract)

GASSER, CHARLES S.; GUNNING, DORIAN A.; BUDELIER, KIM A.; BROWN, SHERRI M. Structure and expression of cytosolic cyclophilin/peptidyl-prolyl cis-trans isomerase of higher plants and production of active tomato cyclophilin in *Escherichia coli*. *PNAS*.v. 87, pp. 9519-9523, 1990.

GOLOUBINOFF, Pierre; MOGK, Axel; ZVI, Anat P. B.; TOMOYASU, Toshifumi; BUKAU, Bernad. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *PNAS*, v.96, n. 24, pp. 13732-13737, 1999.

HARTL, F.U. Molecular chaperones in cellular protein folding. *Nature*, v. 381, p. 571-579, 1996

HARRAR, Yaël; BELLINI, Catherine; FAURE, Jean-Denis. FKBP: at the crossroads of folding and transduction. *Trends in Plant Science*, v.6, n.9, pp. 426-431, 2001.

HÜBEL, Anja; SCHÖFF, Fritz. Arabidopsis heat shock factor: isolation and characterization of the gene and the recombinant protein. *Plant Mol. Biol.*, v. 26, n. 1, p. 353-62, 1994. (Abstract)

HUEROS, Gregorio; RAHFELD, Jens Rahfeld; SALAMINI, Francesco; THOMPSON, Richard. A maize FK506-sensitive immunophilin, mzFKBP-66, is a peptidylproline cis-trans-isomerase that interacts with calmodulin and a 36-kDa cytoplasmic protein. *Planta*, v. 205, pp. 121-131, 1998.

HIROHASHI, Toshiya; NISHIO, Kazyaki; NAKAI, Masato. cDNA sequence and overexpression of chloroplast chaperonin 21 from *Arabidopsis thaliana*. *Biochim. Biophys. Acta*, v.1429, pp. 512-515, 1999

KARLIN, Samuel; BROCCIERI, Luciano. Heat shock protein 70 family: multiple sequence comparisons, function, and evolution. *J. Mol. Evol.* v. 47, pp. 565-577, 1998.

KATIYAR-AGARWAL, Surekha; AGARWAL, Manu; GROVER, Anil. Heat-tolerant basmati rice engineered by over-expression of hsp101. *Plant Mol. Biol.*, v.51, pp. 677-686, 2003.

KRISHNA, Priti; GLOOR, Greg. The Hsp90 family of proteins in *Arabidopsis thaliana*. *Cell Stress Chaperones*, v. 6, n. 3., pp. 238-246, 2001.

KRISHNA, Priti; SACCO, Melanie; CHERUTTI, James F; HILL, Sylvia. Cold-Induced accumulation of hsp90 transcripts in *Brassica napus*. *Plant Physiol.*, v. 107, n. 3, pp. 915-923, 1995.

KONING, Ann J.; ROSE, Ronald; CAMAI, Luca. Developmental expression of tomato heat-shock cognate protein 80. *Plant Physiol.*, v. 100, n.2, pp. 801-811, 1992.

KUREK, Isaac; AVIEZER, Keren; EREL, Noa; HERMAN, Eliot; BREIMAN, Adina. The Wheat Peptidyl Prolyl cis- trans-Isomerase FKBP77 Is Heat Induced and Developmentally Regulated. *Plant Physiol.*, v. 119, pp. 693-703, 1999.

LANDRIEU, Isabelle; VEYLDER, Lieven De; FRUCHART, Jean-Se ´bastien; ODAERT, Benoît; CASTEELSI, Peter; PORTETELLE, Daniel; MONTAGUI, Marc Van; INZÉI, Dirk Inze; LIPPENS, Guy. The *Arabidopsis thaliana* PIN1At Gene Encodes a Single-domain Phosphorylation-dependent Peptidyl Prolyl cis/trans Isomerase. *J. Biol. Chem.*, v. 275, n. 14, pp. 10577-10581, 2000.

LAWRENCE, Susan D.; CLINE, Kenneth; MOORE, Gloria A. Chromoplast development in ripening tomato fruit: identification of cDNAs for chromoplast-targeted proteins and characterization of a cDNA encoding a plastid-localized low-molecular-weight heat shock protein. *Plant Mol. Biol.*, v. 33, pp. 483-492, 1997.

LEE, Yuh-Ru J.; NAGAO, Ronald T.; KEY, Joe L. A Soybean 101-kD Heat Shock Protein Complements a Yeast HSP704 Deletion Mutant in Acquiring Thermotolerance. *Plant Cell*, v. 6, pp. 1889-1897, 1994.

- LIPPUNERT, Veronica; CHOU, Irene T.; SCOTT, Sidney V.; ETTINGER, William F.; THEGH, Steven M.; GASSER, Charles S. Cloning and Characterization of Chloroplast and Cytosolic Forms of Cyclophilin from *Arabidopsis thaliana*. *J. Biol. Chem.*, v. 269, n. 11, pp. 7863-7868, 1994.
- LUAN, Sheng; LANE, William S.; SHREIBER, Stuart L. pCyP B: A Cyclophilin Chloroplast-Localized, Heat Shock- Responsive from Fava Bean. *Plant Cell*, v. 6, pp. 885-892, 1994.
- MARRS, K. A.; CASEY, E. S.; CAPITANT, S. A.; BOUCHARD, R. A. DIETRICH, P. S. METTLER, I. J. SINIBALDI, R. M. Characterization of two maize HSP90 heat shock protein genes: expression during heat shock, embryogenesis, and pollen development. *Dev. Genet.*, v.14, n. 1, p. 27-41, 1993. (Abstract)
- MEDINA-ESCOBAR, N.; CÁRDENS, J.; MUNÓZ-BLANCO, J.; CABALLERO, J.L. cloning and molecular characterization of a strawberry fruit ripening-related cDNA corresponding a mRNA for a low-molecular-weight heat-shock protein. *Plant Mol. Biol.*, v. 36, pp. 33-42, 1998.
- NIETO-SOTELO, J.; KANNAN, K.B.; MARTÍNEZ, L.M.; SEGAL, C. Characterization of a maize heat-shock protein 101 gene, *HSP101*, encoding a ClpB/Hsp100 protein homologue. *Gene*, v. 230, pp. 187-195, 1999.
- NOVER L., BHARTI K., DÖRING P, MISHRA S.K., GANGULI A., SCHARF K.D. *Arabidopsis* and the heat stress transcription factor world: How many heat stress transcription factors do we need? *Cell Stress Chaperones* v. 6, pp. 177-189, 2001.
- NOVER, L.; SCHARF, K. D. Heat stress proteins and transcription factors. *Cell. Mol. Life Sci.*, v. 53, pp. 80-103, 1997.
- PANCHUK, I. I.; VOLKOV, R. A.; SCHOFFL, F. Heat Stress- and Heat Shock Transcription Factor-Dependent Expression and Activity of Ascorbate Peroxidase in *Arabidopsis*, *Plant Physiology*, v. 129, pp. 838-853, 2002.
- PAREEK, Ashwani; SINGLA, Sneha Lata; GROVER, Anil. Immunological evidence for accumulation of two high-molecular-weight (104 and 90 kDa) HSPs in response to different stresses in rice and in response to high temperature stress in diverse plant genera. *Abstract, Plant Molecular Biology*, v. 29, n.2, pp. 293, 1995.
- PARSELL, Dawn A.; KOWAL, Anthony S.; SINGER, Mike A.; LINDQUIST, Susan. Protein disaggregation mediated by heat-shock protein hsp104. *Nature*, v. 372, pp. 475-478, 2002. (Abstract)
- PEARL, Laurence H.; PRODRIMOU, Chrisostomos. Structure and *in vivo* function of hsp90. *Cur. Opin. Struct. Biol.*, v. 10, pp. 46-51, 2000.
- PICARD, D. Heat-shock protein 90, a chaperone for folding and regulation. *Cell Mol Life Sci*, v. 59, n. 10, pp. 1640-1648, 2002.
- QUEITSCH, Christine; HONG, Suk-Whan; VIERLING, Elizabeth; LINDQUIST, Susan. Heat shock protein 101 plays a crucial role in thermotolerance in *Arabidopsis*. *Plant Cell*, v. 12, pp. 479-492, 2000.
- SANGSTER, Todd A.; QUEITSCH, Christine. The HSP90 chaperone complex, an emerging force in plant development and phenotypic plasticity. *Current Opinion in Plant Biology*, v. 8, pp.86-92, 2005.
- SCHARF, K. D.; ROSE, S.; ZOTT, W.; SCHÖFF, F.; NOVER, L.; SCHÖFF, F. Tree tomato genes code for heat stress transcription factors with a region of remarkable homology to the DNA-binding domain of the yeast HSF. *EMBO J.*, v. 9, n. 13, pp. 4495-4501, 1990.
- SCHIRMER, Eric C.; GLOVER, John R.; SINGER, Mike A. LINDQUIST, Susan. HSP 100/Clp proteins: a common mechanism explains diverse functions. *TIBS*, v. 21, pp. 289-296, 1996.
- SCHIRMER, Eric C.; LINDQUIST, Susan; VIERLING, Elizabeth. An *Arabidopsis* Heat Shock Protein Complements a Thermotolerance Defect in Yeast. *Plant Cell*, v. 6, pp.1899-1909, 1994.
- SINGLA, Sneha L.; PAREEK, Ashwani; KUSH, Anil K.; GROVER, Anil. Distribution patterns of 104 kDa stress-associated protein in rice. *Plant Mol. Biol.*, v.37, pp. 911-919, 1998.

- TAKAHASHI, Taku; NAITO, Satoshi; KOMEDA, Yoshibumi. Isolation and analysis of the expression of two genes for the 81 kilodalton heat-shock proteins from *Arabidopsis*. *Plant Physiol.*, v. 99, n. 2, pp. 383-390, 1992.
- TOFT, David O. Recent Advances in the study of hsp90 structure and mechanism of action. *TEM*, v. 9, n.6, pp. 238-243, 1998.
- VIERLING, Elizabeth. The Roles of Heat Shock Proteins in Plants. *Ann. Rev. of Plant Physiol. and Plant Mol. Biol.*, v. 42, pp. 579-620, 1991
- WANG, Wangxia; VINOCCUR, Basia; SHOSEYOV, Oded; ALTMAN, Arie. Role of plant heat shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, v. 9, n. 5, pp. 244-252, 2004.
- WATERS, Elizabeth R.; LEE, Garrett J.; VIERLING, Elizabeth. Evolution, structure and function of the small heat shock proteins in plants. *J. Exp. Bot.* v. 47, n. 3, pp. 325-338, 1996.
- WATERS, Elizabeth R.; VIERLING, Elizabeth. Chloroplast small heat shock proteins: Evidence for atypical evolution of an organelle-localized protein. *PNAS.*, v. 96, n. 25, pp. 14394-14399, 1999.
- WEGELE, H.; MULLER, L.; BUCHNER, J. Hsp70 and Hsp90—a relay team for protein folding. *Rev Physiol Biochem Pharmacol*, v. 151, pp. 1-44, 2004.
- WU, Yang; EGERTON, Gillian; BALL, Amanda; TANGUAY, Robert M.; BIANCO, Albert E. Characterization of the heat-shock protein 60 chaperonin from *Onchocerca volvulus*. *Mol. Biochem. Parasit.*, v.107, pp. 155-168, 2000.
- YABE, Naoto; TAKAHASHI, Taku; KOMEDA, Yoshibumi. Analysis of tissue specific expression of *Arabidopsis thaliana* Hsp90 – family gene Hsp81. *Plant and Cell Physiol.*, v. 35, n.8, pp. 1207-1219, 1994.
- YOUNG, Jason C.; MOAREFI, Ismail; HARTL, F. Ulrich. Hsp90: a specialized but essential protein-folding tool. *J Cell Biol*, v. 154, pp. 267-273, 2001.
- ZHANG, Chun; GUY, Charles L. Co-immunoprecipitation of Hsp101 with cytosolic Hsc70. *Plant Physiol. Biochem.*, v. 43, pp. 13-18, 2005.
- ZHANG, Xiao-Ping; GLASER, Elzbieta. Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. *TRENDS in Plant Science*, v.7, n.1, pp. 14-21, 2002
- ZHANG, Xiao-Ping; ELOFSSON, Arne; ANDREU, David; GLASER, Elzbieta. Interaction of mitochondrial presequences with DnaK and mitochondrial Hsp70. *J. Mol. Biol.* v. 288, n. 1, pp. 177-190, 1999.
- ZHU, X.; ZHAO, X.; BURKHOLDER, W.F.; GRAGEROV, A.; OGATA, C.M.; GOTTESMAN, M.E.; HENDRICKSON, W.A. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science*, v. 272, n. 5268, pp. 1606-1614, 1996.

Capitulo 4

POTENCIAIS GENES RELACIONADOS AO METABOLISMO DO ÁCIDO ASCÓRBICO EM FRUTOS DO CAMU-CAMU

RESUMO

A vitamina C, termo genérico para todos os compostos que exibem atividade biológica do L-ácido ascórbico, exerce funções antioxidantes e metabólicas em plantas e animais. Pretende-se identificar seqüências ESTs de frutos de camu-camu relacionadas ao metabolismo do ácido ascórbico. A partir das análise com auxílio da bioinformática foram identificadas nove ESTs relacionados a síntese do ácido ascórbico: uma provável manose-6-fosfato isomerase (ORF de 236pb e 71,0% de identidade), fosfomanose isomerase (ORF de 438pb e 67,9% de identidade), GDP-manose pirofosforilase (ORF de 504pb e 72,5% de identidade), GDP-manose-3",5"-epimerase GME-1 (81,6% de identidade), cinco ESTs de GDP-manose-3,5-epimerase/catalítica (um EST com 80,1% de identidade e quatro formando o *contig* CL69 com ORF de 678pb e 78,8% de identidade). Da conversão e reciclagem do ácido áscórbico foram identificadas: ascorbato peroxidase de estroma cloroplastídico (ORF de 339pb e 67,0% de identidade), ascorbato peroxidase citosólica (ORF de 552pb e 81,5% de identidade - *contig* com ORF de 520pb e 81,3% de identidade), a precursora de L-ascorbato oxidase (Ascorbase) (ORF de 375pb e 58,0% de identidade), monodeidroascorbato redutase (ORF de 568pb e 76,5% de identidade). Outros genes potencialmente relacionados ao metabolismo do ácido ascórbico foram identificados como *Vitamin C defective 2* - VTC2.

ABSTRACT

The vitamin C, generic term for everybody the composed that exhibit biological activity of the L-ascorbic acid, exercises antioxidantes and metabolic functions in plants and animal. It intends to identify ESTs sequences of camu-camu fruits related to the ascorbic acid metabolism. From the analysis with bioinformatic aid were identified nine ESTs related the synthesis of the ascorbic acid: one putative mannose-6-phosphato isomerase (ORF of 236pb and identity 71,0%), phosphomannose isomerase (ORF of 438pb and identity 67,9%), GDP-mannose pyrophosphorylase (ORF of 504pb and identity 72,5%), GDP-mannose-3",5"-epimerase GME-1 (identity 81,6%), five ESTs of GDP-mannose-3,5-epimerase/catalitic (an EST with identity 80,1% and four compouse one contig CL69 with 678pb and identity 78,8%). Ascorbic acid conversion and recycling were identified: ascorbate peroxidase of cloroplast estrom (ORF of 339pb and identity 67,0%), ascorbate peroxidase citosolic (ORF of 552pb and identity 81,5% - contig with ORF of 520pb and identity 81,3%), the L-ascorbate oxidase precursor (Ascorbase) (ORF of 375pb and identity 58,0%), monodehydroascorbate reductase (ORF of 568pb and identity 76,5%). Others genes potentially related to the ascorbic acid metabolism were identified like Vitamin C defective 2 – VTC2.

1. INTRODUÇÃO

A vitamina C, termo genérico para todos os compostos que exibem atividade biológica do L-ácido ascórbico, exerce funções antioxidante e metabólica em plantas e animais. É muito importante na nutrição humana e animal (VANDAMME, 1992), atua como cofator enzimático e como precursor da síntese de oxalato e tartrato. Participa em uma variedade de processos incluindo a fotossíntese, fotoproteção, crescimento da parede celular e expansão celular, resistência a estresse ambiental e oxidativo, na síntese de etileno, giberilinas, antocianinas e na hidroxiprolina (SMIRNOFF & WHEELER, 2000; CONKLIN, 2001).

Aproximadamente 50% do ácido ascórbico sintético são usados em suplementos vitamínico e preparações farmacêuticas. O uso do ácido ascórbico como um aditivo para cosméticos é um mercado em rápida expansão, em virtude de suas propriedades de antioxidante e seu potencial em estimular produção de colágeno (LUPO, 2001).

As propriedades antioxidantes do ácido ascórbico também são exploradas no processamento de alimento (25% da produção total) e na fabricação de bebidas (15%) para prevenir a descoloração de pigmento e o marrom enzimático, para projetar sabor e aroma, além de proteger ou aumentar o valor nutricional (HANCOCK & VIOLA, 2002).

A vitamina C é exigida na função cardiovascular, no desenvolvimento da imunidade celular, na conexão entre os tecidos e na utilização de ferro. Embora plantas e a maioria de animais possam sintetizar ácido ascórbico, em humano falta a enzima *L-gulono-1,4-lactone oxidoreductase*, que é exigida na etapa final da síntese do ácido ascórbico (CHEN et al., 2003).

Duas vias biosintéticas foram inicialmente propostas para a síntese da vitamina C, uma sugeriu D-galactose como substrato inicial via D-galacturonato, ácido L-

galactônico e L-galactonolactona, oxidado pela enzima L-galactono- γ -lactona desidrogenase para L-ácido ascórbico (LOEWUS, 1999); a outra rota, proposta por Loewus (1990), sugere a síntese da vitamina C a partir da D-Glicose, via D-glicosona e L-sorbose.

A hipótese de que o ascorbato é sintetizado a partir da glicose é bastante aceita. Saito et al. (1990) também, sugeriram a conversão da glicose em ascorbato via D-glicosona e L-sorbose, baseados em estudo usando carbono-14 na glicose (6^{14}C). Entretanto, Conklin et al. (1997) informaram que nenhuma D-glicosona e nem L-sorbose tiveram efeitos positivos na acumulação de ascorbato no mutante *vtc1* de *Arabidopsis* e nenhuma enzima que converta D-glicose a D-glicosona ou de D-glicosona à L-sorbose foi encontrada.

O isolamento do cDNA da enzima L-galactono- γ -lactona desidrogenase a partir de couve-flor (*Brassica oleraceae* var. *botrytis*), expresso em levedura, tiveram como substrato extrato de couve-flor. Os resultados sugeriram que o L-galactono- γ -lactone, precursor do ácido ascórbico, é um constituinte natural de tecido vegetal (OSTERGAARD et al., 1997).

Em 1998, foi proposta uma rota metabólica via GDP-mannose, GDP-L-galactose e L-galactono-1,4-lactona precursor da vitamina C em plantas superiores (WHEELER et al., 1998) na qual cerca de nove enzimas participam. Seis destas enzimas convertem a D-glicose a GDP-manose e GDP-L-galactose que fornecem precursores GDP-açúcares que estão envolvidos na síntese de polissacarídeos. Através de hidrólise do GDP-L-galactose e da oxidação da L-galactose a L-galactono-1,4-lactona é formada, esta passa por oxidação sob atuação da L-galactono-1,4-deidrogenase formando o ácido ascórbico (SIENDONES et al., 1999).

O cDNA que codifica a enzima GDP-manose pirofosforilase (GMP) foi isolado por Keller et al., em 1999, de *Solanum tuberosum* e utilizado em plantas transgênicas para

avaliar o efeito da redução da atividade da GMP nos níveis do AsA, os resultados mostraram uma correlação entre a atividade desta enzima (GMP) e os níveis de vitamina C.

Wolucka et al. (2001) purificaram a GDP-manose-3,5-epimerase de *Arabidopsis thaliana* - enzima dimérica composta de subunidades aparentemente idênticas, e identificaram o gene clonando e expresso em *E. coli*. Esta enzima é chave na biossíntese de ácido ascórbico pois converte GDP-manose em GDP-L-galactose.

Recentemente, a L-galactose-1-fosfato fosfatase foi purificada de frutos de kiwi (*Actinidia deliciosa*). A enzima tem 65K Da dependente de Mg^{2+} . O cDNA desta enzima foi similar a proteína *myo-inositol-1-phosphato* de *Arabidopsis thaliana* que, quando purificada, também apresentou propriedades similares. A enzima L-galactose-1-fosfato converte a L-galactose-1-P em L-galactose na biossíntese de ácido ascórbico (LAING et al., 2004).

O ácido ascórbico produzido nas plantas pode ser oxidado pela Ascorbato peroxidase (APX) ao radical Monodeidroascorbato (MDHA), este por sua vez é não enzimaticamente convertido a Deidroascorbato (DHA). A DHA pode sofrer uma hidrólise e formar irreversivelmente o ácido 2,3-diketogulônico (CHEN et al., 2003), ou ser catabolizado formando produtos o oxalato e tartrato (NOCTOR & FOYER, 1998)

Quatro isoenzimas de APX tem sido demonstradas em termos de localização celular: APX das membranas tilacóides (tAPX), APX do estroma de cloroplastos (sAPX), APX citosólica (cAPX) e a APX de microcorpos (mAPX) (MITTLER et al., 1998; YOSHIMURA et al., 1998).

O cDNA que codifica a ascorbato peroxidase já foi isolado de diferentes tecidos vegetais, como tubérculo de batata (KAWAKAMI et al., 2002), de raízes de *Japanese radish* (OHYA et al., 1997), folhas de soja (CALDWELL et al., 1998), tomate (GADEA et al., 1999) e calo de *Pimpinella brachycarpa* (SOHN et al., 2002), .

A reciclagem do ácido ascórbico nas plantas ocorre pela conversão do MDHA utilizando a MDHA redutase (MDHAR) ou pela conversão do DHA pela Deidroascorbato redutase (DHAR), que usa glutathiona (GSH) como agente redutante. Assim, DHAR permite a planta reciclar DHA recapturando o ácido ascórbico antes que seja perdido (CHEN et al., 2003).

O cDNA da DHAR de trigo foi clonado e introduzido em tabaco e milho. A expressão da enzima aumentou 32X em tabaco e 100X em milho, e isso ocasionou o aumento da expressão do ácido ascórbico – 2X em folhas de tabaco e 4X em milho, também em função desta enzima ser responsável pela reciclagem do ascorbato (CHEN et al., 2003).

A MDAR desempenha um importante papel no sistema antioxidante em vegetais por ser capaz de regenerar o ascorbato e manter o *pool* de ascorbato na célula. Recentemente, MDAR foi mostrado ser também capaz de reduzir radicais fenoxil a fenol (SAKIHAMA, 2000; LETERRIER et al., 2005).

Isoformas de monodehidroascorbato redutase existe na mitocôndria, cloroplastos, citosol e micro-corpos (OBARA et al., 2002) e o cDNA de MDAR foi clonado de pepino (HOSSAIN & ASADA, 1985), de nódulos de raízes de soja (DALTON et al., 1992), ervilha cultivada (*Pisum sativum* L) (MURTHY & ZILINSKAS, 1994), tomate (GRANTZ et al., 1995), arroz (Número de acesso D85764) e *Arabidopsis* (Número de acesso D84417).

Na Amazônia, como potencial fonte de genes para manipulação do teor de ácido ascórbico em plantas, temos o camu-camu (*Myrciaria dúbia* (H.B.K.) McVaugh), espécie nativa pertencente à família Myrtaceae, presente nas margens de rios e lagos de toda região. Seus frutos apresentam altíssimo teor de vitamina C, com ampla variabilidade, chegando a atingir 6.1g de ácido ascórbico em 100 g de polpa (CALZADA BENZA, 1980; ANDRADE et al., 1991; YUYAMA et al., 2002).

Justi et al. (2000) estudaram a estabilidade da vitamina C de camu-camu, produzido no Estado do Paraná, a baixa temperatura, e verificaram que a polpa de camu-camu armazenada a -18º C diminui 23% em sua concentração de ácido ascórbico até 28 dias (de 1.57 até 1.21 g/100 g). Após 335 dias o conteúdo de vitamina C era de aproximadamente 1.16 g/100 g de polpa, totalizando 26% de perda de ácido ascórbico, todavia este conteúdo é ainda mais elevado comparado a outras fruteiras.

Outra característica observada no camu-camu é que grande parte de sua vitamina C se mantêm em produtos industrializados, o sorvete produzido a partir da polpa contém cerca de 560mg de vitamina C e a geléia apresenta concentrações superiores a 710mg de vitamina C por 100g de produto, demonstrando que a vitamina C de camu-camu apresenta grande estabilidade, mesmo submetida a diferentes tipos de processamentos e temperaturas (YUYAMA et al., 2001).

Conhecer os genes expressos no fruto é um passo importante para tentar compreender seu metabolismo e a partir daí tentar inferir os possíveis fatores determinantes no acúmulo de vitamina C em concentrações tão elevadas e variadas em frutos de diferentes populações e além de sua estabilidade em produtos processados (CALZADA BENZA, 1980; ANDRADE et al., 1991; YUYAMA et al., 2001 e 2002).

Diante do exposto, pretende-se identificar na biblioteca de ESTs de frutos de camu-camu possíveis genes relacionados ao metabolismo do ácido ascórbico.

2. MATERIAIS E MÉTODOS.

A biblioteca de EST de frutos de camu-camu armazenada no Sistema Genoma foi analisada para a identificação de possíveis genes relacionados ao metabolismo do ácido ascórbico, seguindo a estratégia de anotação descrita na Figura 01. As seqüências ESTs identificadas de genes potenciais codificantes relacionado ao metabolismo do ácido ascórbico foram analisadas quanto à homologia no programa de alinhamento LALING (http://www.ch.embnet.org/software/lalign_form.html) e anotado o percentual de identidade das seqüências.

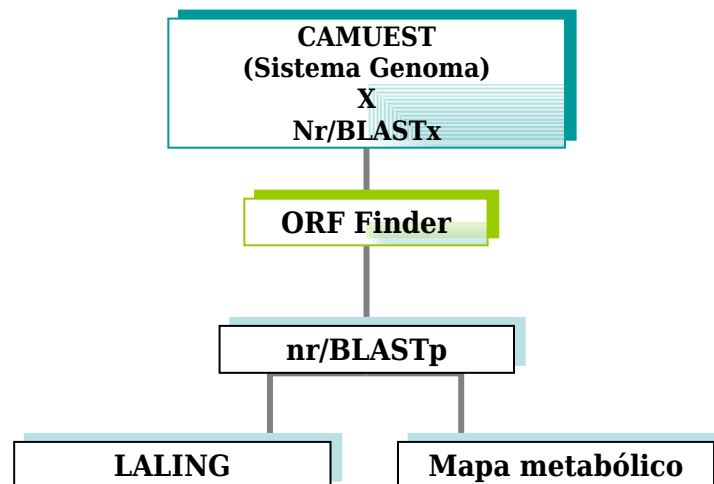


Figura 01- Esquema da estratégia de anotação de possíveis genes relacionados ao metabolismo do ácido ascórbico da biblioteca de ESTs de camu-camu armazenada no programa Sistema Genoma (BIOFOCO).

Baseado em literatura especializada (SMIRNOFF, 1996; WHEELER et al., 1998; PALLANCA E SMIRNOFF, 1999; SMIRNOFF E WHEELER, 2000; HANCOCK E VIOLA, 2002; CHEN et al., 2003; ZOU et al., 2006) organizou-se um mapa metabólico da biossíntese, oxidação e regeneração do ácido ascórbico, com indicação das enzimas cujas seqüências ESTs foram identificados em camu-camu.

3. RESULTADOS E DISCUSSÃO

3.1. Biossíntese de Ácido Ascórbico

Baseado na via proposta por Wheeler et al. (1998) as 2586 ESTs seqüências com similaridade com seqüências do banco de genes foram analisadas e nove ESTs relacionados a síntese do ácido ascórbico foram anotados (Tabela 1, Figura 2).

Uma seqüência com ORF de 236pb apresentou 71,0% de identidade com a provável manose-6-fosfato isomerase de *Oryza sativa* (*e-value* 4e-33), e outra com ORF de 438pb foi similar a *Phosphomanose isomerase* (PMI) de *Cyamopsis tetragonoloba* (*e-value* 4e-49), com 67,9% de identidade (Tabela 1).

As enzimas catalisam a inter-conversão da manose-6-fosfato a frutose-6-fosfato (Figura 2), sendo, portanto, uma enzima chave no metabolismo da manose, expressa em nível muito baixas em muitos vegetais, (BARB et al., 2003) e participante da síntese de ácido ascórbico (WHEELER et al., 1998).

Um EST, com ORF de 504pb, foi similar a provável GDP-manose pirofosforilase de *Oryza sativa* (*e-value* 5e-61) com 72,5% de identidade (Tabela 1), esta enzima participa na inter-conversão da Manose-1-P em GDP-D-manose usado na biossíntese de ácido ascórbico (Figura 2), biossíntese de carboidratos da parede celular e na glicosilação protéica (CONKLIN et al., 1999).

A GDP-manose pirofosforilase teve seu cDNA caracterizado a partir de *Solanum tuberosum* (KELLER et al., 1999) e de folhas de tabaco (TABATA et al., 2004), também foi demonstrado que o locus VTC1 do mutante de *Arabidopsis thaliana* codifica está enzima (CONKLIN et al., 1999).

A GDP-manose-3,5-epimerase (GME), outra enzima que participa da biossíntese de ácido ascórbico (Wheeler et al., 1998), também foi identificada na biblioteca de

ESTs de frutos de camu-camu. Um ESTs, que não apresentou ORF, foi similar GDP-manose-3",5"-epimerase GME-1 de *Oryza sativa*, com 81,6% de identidade. Cinco ESTs foram similares a GDP-manose-3,5-epimerase/catalítica de *Arabidopsis thaliana*, um com *e-value* 4e-18 e 80,1% de identidade não apresentou ORF; e quatro formando o *contig* CL69, com ORF de 678pb, *e-value* de 4e-97 e 78,8% de identidade (Tabela 1).

A GDP-D-manose-3,5-epimerase participa na inter-convesão da GDP-D-manose em GDP-L-galactose (Figura 2) (WHEELER et al., 1998; CONKLIN et al., 1999), considerada uma enzima chave na biossíntese da vitamina C em plantas (WOLUCKA et al., 2001). Esta enzima foi identificada em ervilha e *Arabidopsis thaliana* (WHEELER et al., 1998).

Apesar da elevada produção de vitamina C em frutos de camu-camu pouco genes foram identificados relacionados à sua síntese, a expressão da GDP-D-manose-3,5-epimerase nos leva acreditar que a via proposta por Wheeler et al. (1998) ocorre em camu-camu, contudo o baixo número de ESTs analisados e sendo estes referentes a cDNA de 500 a 1000pb, possivelmente dificultou a identificação de um maior número de genes da via.

3.2. Conversão e reciclagem do Ácido Ascórbico

O ácido ascórbico depois de sintetizado é oxidado a monodehidroascorbato (MDHA) pela ascorbato peroxidase por meio da transferência de elétrons e conseqüente conversão de H₂O₂ em água (LISENBEE et al., 2005) e pela ascorbato oxidase na parede celular vegetal (NOSTOR & FOYER, 1998) (Figura 2).

Código	Definição	Organismo	GI	e-value	No. ESTs	ORF (pb)	Identidad e %
MDI27C03	Provável manose-6-fosfate isomerase	<i>Oryza sativa</i>	34895502	4E-33	1	236	71,0
MDI05A05	Fosfomanose isomerase	<i>Cyamopsis tetragonoloba</i>	60617303	4E-49	1	438	67,9
MDI35B05	Provável GDP-manose pirofosforilase	<i>Oryza sativa</i>	50920467	5e-61	1	504	72,5
MDI25C09	GDP-manose-3",5"-epimerase GME-1	<i>Oryza sativa</i>	54650505	3E-45	1	-	81,6
MDI15F06	GDP-manose-3,5-epimerase/ catalítica	<i>Arabidopsis thaliana</i>	30692011	4E-18	1	-	80,1
CL69	GDP-mannose-3,5-epimerase/ catalítica Ascorbato peroxidase do estroma	<i>Arabidopsis thaliana</i>	15241945	4E-97	4	678	78,8
MDI29A05	cloroplastídico	<i>Vigna unguiculata</i>	45268439	9E-17	1	339	67,0
CL210	Ascorbato peroxidase citosólica	<i>Fragaria x ananassa</i>	2754860	1E-84	2	552	81,5
CL118	Ascorbato peroxidase citosólica	<i>Fragaria x ananassa</i>	2754860	2E-82	3	520	81,3
MDI21G04	Precursora da L-ascorbato oxidase (Ascorbase)	<i>Nicotiana tabacum</i> <i>Brassica rapa subsp.</i>	2493321	1e-20	1	375	58,0
CL86	Monodeidroascorbato redutase	<i>Pekinensis</i>	14764532	2E-86	4	568	76,5
MDI15B02	Vitamina C defectiva 2 - VTC2	<i>Arabidopsis thaliana</i>	18416877	2E-28	1	296	68,6
CL136	Vitamina C defectiva 2 - VTC2	<i>Arabidopsis thaliana</i>	18416877	5E-81	3	705	70,0
CL293	Vitamina C defectiva 2 -VTC2	<i>Arabidopsis thaliana</i>	18416877	4E-12	2	252	73,0

*GI = número de identificação do gene; ORF = *Oper Read Frame*

Tabela 01 – Relação de possíveis genes identificados na biblioteca de ESTs de camu-camu com similaridade a proteínas relacionadas ao metabolismo do ácido ascórbico.

Um EST, com ORF de 339pb, foi similar a ascorbato peroxidase de estroma cloroplastídico de *Vigna unguiculata* (*e-value* 9e-17), com 67,0% de identidade. Dois *contigs* formados por cinco ESTs foram similares a ascorbato peroxidase citosólica (cAPX) de *Fragaria x ananassa*: um com ORF de 552pb (*e-value* 1e-84) e 81,5% de identidade e o outro *contig* com ORF de 520pb (*e-value* 2e-82) com 81,3% de identidade (Tabela 01).

A APX tem um papel chave na recuperação de peróxido de hidrogênio e as isoenzimas ascorbato peroxidase estão localizadas no cloroplasto, citosol, peroxissomos e mitocôndria (MITTLER & ZILINSKAS 1992; ASADA 1992).

As isoenzimas cloroplastídicas e citosólica da ascorbato peroxidase são distintas uma da outra nas seguintes propriedades: a) a isoenzima cloroplastídica, tem um curto período de vida em meio sem ascorbato; b) A sensibilidade da isoenzima cloroplastídica ao reagente tiol e a inibidores de suicídio é mais elevada que na isoenzima citosólica; c) A isoenzima cloroplastídica é mais específica ao ascorbato como doador de elétrons que a isoenzima citosólica que pode, por exemplo, oxidar o pirogalol num índice apreciável (ASADA, 1992).

Um cDNA de 1379 pb da ascorbato peroxidase do estroma de cloroplastos foi isolado de cotilédones de espinafre, a seqüência completa revela uma ORF de 1095pb que codifica 365 aminoácidos (ISHIKAWA et al., 1996). O cDNA que codifica a Ascorbato peroxidase já foi isolado de tubérculo de batata . O clone completo contém 1039 bp e apresenta uma ORF de 750 bp que codifica 250 aminoácidos. A seqüência de aminoácido deduzida de APX teve alta homologia com a cAPX de outras planta (KAWAKAMI et al., 2002).

A APX citosólica difere da APX cloroplastídica em composição, estrutura, especificidade do substrato, e estabilidade num meio sem ascorbato (MITTLER et al., 1998; YOSHIMURA et al., 1998).

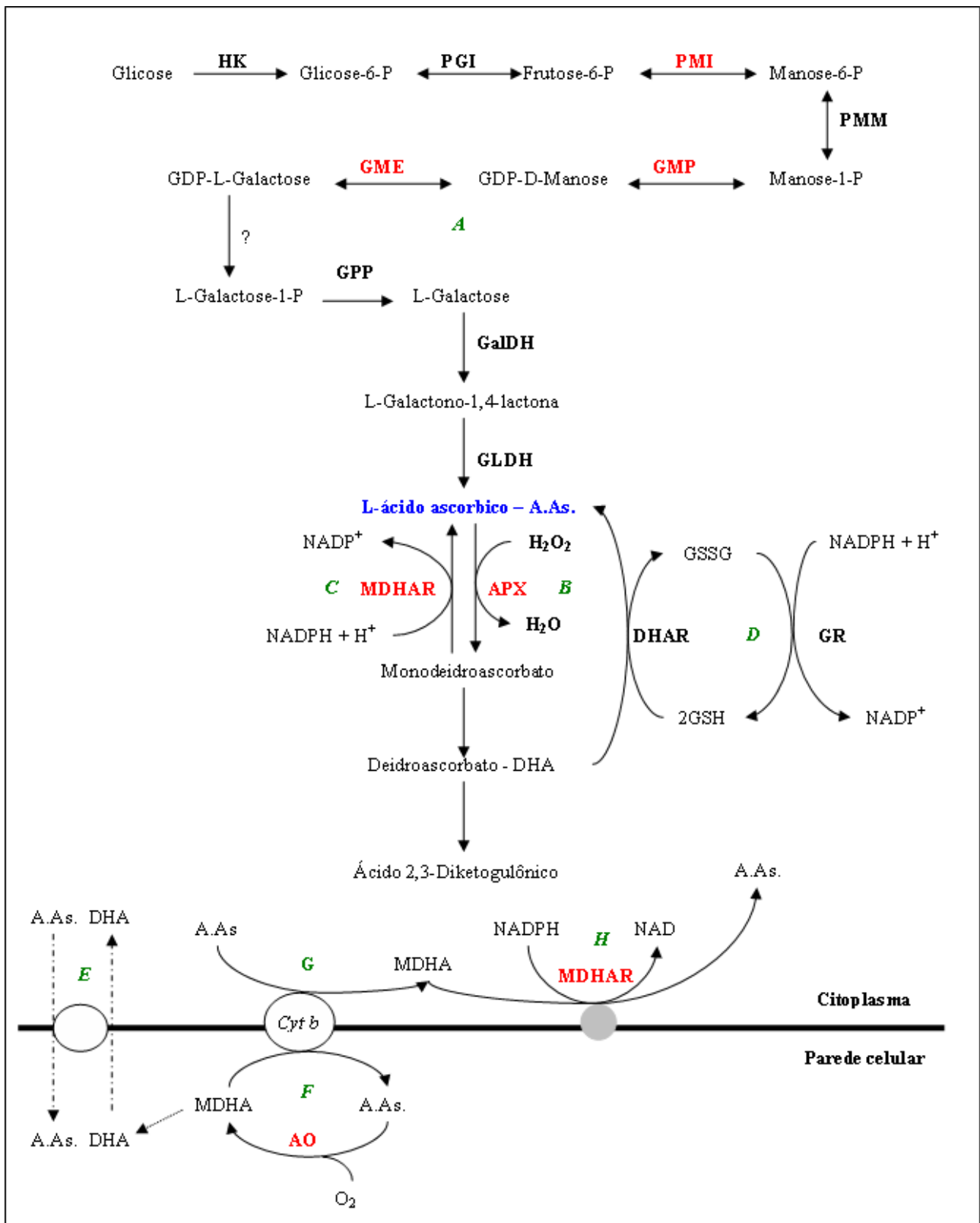


Figura 2 – Biossíntese, degradação e reciclagem do ácido ascórbico (A.As.) em plantas. A) Biossíntese do A.As. a partir da D-glicose, é uma complexa via com nove etapas com a participação das seguintes enzimas: HK – hexoquinase; PGI – Fosfoglicose isomerase; PMI – Fosfomanose isomerase; PMM – Fosfomanose mutase; GMP – GDP-manose pirofosforilase; GME – GDP-manose-3,5-epimerase; GPP – L-galactose-1-fosfato fosfatase; L-GalDH – L-

galactose deidrogenase; L-GLDH – L-galactono-1,4-lactona deidrogenase; B) A redução do peróxido de hidrogênio pelo ascorbato é catalisado pela ascorbato peroxidase citosólica (cAPX); C) O produto da oxidação, o radical monodehidroascorbato (MDHA) é reduzido à ascorbato pela MDHA redutase dependente de NAD(P); D) DHA é reduzida a ascorbato pela Dehidroascorbato redutase (DHAR) dependente de Glutathione (GSH), que é regenerada de sua forma oxidada GSSG pela Glutathione redutase (GR) a GSH; E) O A.As. é transportado através da parede celular pela difusão via um transportador de membrana plasmática na troca por DHA; F) A ascorbato oxidase (AO) é uma glicoproteína secretada que catalisa a oxidação do ascorbato na parede celular vegetal; G) O MDHA resultante é provavelmente reduzido pelo sistema citocromo *b* da membrana plasmática; H) Uma MDHAR ligada a membrana plasmática regenera o ascorbato no lado citosólico. As enzimas encontradas no banco de ESTs de camu-camu estão em vermelho.

Uma seqüência com ORF de 375pb apresentou 58,0% de identidade um precursor da L-ascorbato oxidase (Ascorbase) de *Nicotina tabacum* (*e-value* 1e-20) (Tabela 01). A ascorbato oxidase é uma glicoproteína que catalisa a oxidação do ácido ascórbico em MDHA na parede celular e pode funcionar apoplasticamente como um oxidante do ascorbato durante o processo de alojamento celular (KATO & ESAKA, 1999).

O cDNA de ascorbato oxidase foi isolado de células de tabaco (KATO & ESAKA, 1996), de tomate (ZOU et al., 2006) e de *Capsicum annuum* (GARCIA-PINEDA et al., 2004).

O *contig* CL86, com ORF de 568pb e formado por quatro ESTs, foi similar a Monodehidroascorbato redutase de *Brassica rapa subsp. pekinensis* (*e-value* 2e-86) com 76,5% de identidade (Tabela 1). A enzima Monodehidroascorbate converte a MDHA em ácido ascórbico via transferência de elétrons do NADH (Figura 2) (LISENBEE et al., 2005).

A MDHAR foi purificada de frutos de pepino (HOSSAIN & ASADA, 1985). O cDNA que codifica a enzima foi isolado de ervilha (MURTHY & ZILINSKAS, 1994), de

tomate (*Lycopersicon esculentum* Mill.), com ORF de 1299pb (GRANTZ et al., 1995), de *Brassica campestris*, com ORF de 1305pb (YOON et al., 2004).

Outros genes potencialmente relacionados ao metabolismo do ácido ascórbico foram identificados como Vitamina C defetiva 2 – VTC2 de *Arabidopsis thaliana*, um EST apresentou ORF de 296pb (*e-value* 2e-28) e 68,6% de identidade, o *contig* CL136, com ORF de 705pb e formado por três ESTs apresentou similaridade de 5e-81 (70,0% de identidade) e o *contig* CL293 apresentou ORF com 252pb e 73% de identidade (*e-value* 4e-12) (Tabela 1).

O mutante VTC2 determina como a disponibilidade do ascorbato afeta o NPQ *in vivo*. O excesso de energia luminosa absorvida pela planta pode levar à formação de perigosos radicais de oxigênio, um problema que minimizado pela dissipação do excesso na forma de calor. Este processo de dissipação térmica é chamado de NPQ (*non-photochemical quenching*) da fluorescência da clorofila. (MÜLLER-MOULÉ et al., 2002).

Smirnoff (2000) observou que o VTC2 e outros mutantes de VTC mostraram um diminuição no nível de NPQ zeaxantina- dependente, confirmando que o ascorbato é o cofator da violaxantina *de-epoxidase* e que indisponibilidade do ascorbato no lúmen da tilacóide pode limitar esta reação.

4. CONCLUSÕES

- Foram identificados diferentes seqüências ESTs relacionadas ao metabolismo de síntese do ácido ascórbico: provável manose-6-fosfate isomerase, fosfomanose isomerase, provável GDP-manose pirofosforilase, GDP-manose-3",5"-epimerase GME-1 e GDP-manose-3,5-epimerase/catalítica.
- As seqüências ESTs relacionadas a degradação e reciclagem de ácido ascórbico foram similares a ascorbato peroxidase do estroma cloroplastídico, ascorbato peroxidase citosólica, precursora da L-ascorbato oxidase (Ascorbase) e a Monodeidroascorbato redutase.

5. REFERÊNCIAS BIBLIOGRÁFICAS.

- ANDRADE, Jerusa S.; GALEZI, Maria A. M.; ARAGÃO, Carlos G.; CHÁVEZ-FLORES, Wanders B. Valor nutricional do camu-camu (*Myrciaria dubia* (H.B.K.) McVaugh) cultivado em terra firme da Amazônia Central. *Rev. Bras. Frutic.* v. 13, n. 3, pp. 307 – 311, 1991
- ASADA, Kosi. Ascorbate peroxidase – a hydrogen peroxide-scavenging enzyme in plants. *Physiol. Planta.*, v. 85, pp. 235-241, 1992.
- BARB, Adam W.; PHARR, D. Mason; WILLIAMSON, John D. A *Nicotiana tabacum* cell culture selected for accelerated growth on mannose has increased expression of phosphomannose isomerase. *Plant Science*, v. 165, pp. 639- 648, 2003.
- CALZADA BENZA, J.C. *143 Frutales nativos*. Librería El Estudiante. La Molina, 314p., 1980.
- Charles R. Caldwell, Frank J. Turano, Michael B. McMahon. Identification of two cytosolic ascorbate peroxidase cDNAs from soybean leaves and characterization of their products by functional expression in *E. coli*. *Planta* (1998) 204: 120±126
- CHEN, Zhong; YOUNG, Todd E.; LING, Jun; CHANG, Su-chih; GALLIE, Daniel R. Increasing vitamin C content of plants through enhanced ascorbate recycling. *PNAS*, v. 100, n. 6, pp. 3525-3530, 2003.
- CONKLIN, Patrícia L. Recent advances in the role and biosynthesis of ascorbic acid in plants. *Plant, Cell and Environment*, v. 24, pp. 383–394, 2001.
- CONKLIN, Patrícia L.; NORRIS, Susan R.; WHEELER, Glen L.; WILLIAMS, E.H.; SMIRNOFF, Nicholas; LAST, Robert L. Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. *PNAS*, v. 96, pp. 4198-4203, 1999.
- DALTON, D. A.; LANGE BERG, L.; ROBBINS, M. Purification and characterization of monodehydroascorbate reductase from soybean root nodules. *Arch. Biochem. Biophys.*, v. 292, n. 1, pp. 281-286, 1992. (Abstract)
- GADEA, J.; CONEJERO, V.; VERA, P. Developmental regulation of a cytosolic ascorbate peroxidase gene from tomato plants. *Mol Gen Genet.*, v. 262, pp. 212-219, 1999.
- GÁRCIA-PINEDA, Ernesto; CASTRO-MERCADO, Elda; LOZOYA-GLORIA, Edmundo. Gene expression and enzyme activity of pepper (*Capsicum annuum* L.) ascorbate oxidase during elicitor and wounding stress. *Plant Science*, v. 166, pp. 237-243, 2004.
- GRANTZ, Alexander A.; BRUMMELL, David A.; BENNETT, Alan B. Ascorbate Free Radical Reductase mRNA Levels Are Induced by Wounding. *Plant Physiol.*, v. 108, pp. 411-418, 1995.
- HANCOCK, R.D.; VIOLA, R. Biotechnological approaches for L-ascorbic acid production. *Trends in Biotechnology*, v. 20, n.7, pp. 299-305, 2002.
- HOSSAIN, Manwar A.; ASADA, Kosi. Monodehydroascorbate Reductase from Cucumber Is a Flavin Adenine Dinucleotide Enzyme. *J. Biol. Chem.*, v. 260, n. 24, pp. 12920-12926, 1985.
- ISHIKAWA, Takahiro; SAKAI, Kosuke; YOSHIMURA, Kazuya; TAKEDA, Toru; SHIGERU, Shigeru. cDNAs encoding spinach stromal and thylakoid-bound ascorbate peroxidase, differing in the presence or absence of their 3'-coding regions. *FEBS Letters*, v. 384, pp.289-293, 1996.
- JUSTI, Karin C. ; VISENTAINER, Jesuí V. ; SOUZA, Nilson E.; MATSUSHITA, M. Nutritional composition and vitamin C stability in stored camu-camu (*Myrciaria dubia*) pulp. *ALAN*, v. 50, n.4, pp. 405-408, 2000.
- KATO Naohiro; ESAKA Muneharu. Changes in ascorbate oxidase gene expression and ascorbate levels in cell division and cell elongation in tobacco cells. *Physiol. Plant*, v. 105, pp. 321–329, 1999.

- KAWAKAMI, Sachito; MATSUMOTO, Yoshio; MATSUNAGA, Arkiko; MAYAMA, Shigeyuki; MIZUNO, Masashi. Molecular cloning of ascorbate peroxidase in potato tubers and its response during storage at low temperature. *Plant Science*, v. 163, pp. 829-836, 2002.
- KELLER, Ruth; SPRINGER, Franziska; RENZ, Andreas; KOSSMANN, Jens. Antisense inhibition of the GDP-mannose pyrophosphorylase reduces the ascorbate content in transgenic plants leading to developmental changes during senescence. *Plant Journal*, v. 19, n. 2, pp. 131-141, 1999.
- LAING, William A.; BULLEY, Slan; WRIGHT, Michele; COONEY, Janine; JENSEN, Dwayne; BARRACLOUGH, Di; MACRAE, Eespeth. A highly specific L-galactose-1-phosphate phosphatase on the path to ascorbate biosynthesis. *PNAS*, v. 101, n.48, pp. 16976-16981, 2004.
- LETERRIER, Marina; CORPAS, Francisco J.; BARROSO, Juan B.; SANDALIO, Luisa M.; RÍO, Luis A. del. Peroxisomal Monodehydroascorbate Reductase. Genomic Clone Characterization and Functional Analysis under Environmental Stress Conditions. *Plant Physiology*, v. 138, pp. 2111-2123, 2005.
- LISENBEE, Cayle S.; LINGARD, Matthew J.; TRELEASE, Richard N. Arabidopsis peroxisomes possess functionally redundant membrane and matrix isoforms of monodehydroascorbate reductase. *Plant Journal*, v. 43, pp. 900-914, 2005.
- LOEWUS, Frank A. Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi. *Phytochemistry*, v. 52, pp. 193-210, 1999.
- LOEWUS, Mary W.; BEDGAR, Diana L.; SAITO, Kazumi; LOEWUS, Frank A. Conversion of L-Sorbosone to L-Ascorbic Acid by a NADP-Dependent Dehydrogenase in Bean and Spinach Leaf. *Plant Physiol.* v. 94, pp. 1492-1495, 1990.
- LUPO, Mary. P. Antioxidants and vitamins in cosmetics. *Clin. Dermatol.*, v. 19, n. 4, pp. 467-473, 2001.
- MITTLER, Ron; FENG, Xuqiao; COHEN, Mira. Post-Transcriptional Suppression of Cytosolic Ascorbate Peroxidase Expression during Pathogen-Induced Programmed Cell Death in Tobacco. *Plant Cell*, v. 10, pp. 461-473, 1998.
- MITTLER, Ron; ZILINSKAS, Barbara A. Molecular Cloning and Characterization of a Gene Encoding Pea Cytosolic Ascorbate Peroxidase. *J. Biol. Chem.*, v. 267, n. 30, pp. 21802-21807, 1992.
- MÜLLER-MOULÉ, P.; CONKLIN, Patrícia L.; NIYOGI, K.K. Ascorbate Deficiency Can Limit Violaxanthin De-Epoxidase Activity in Vivo. *Plant Physiol.*, v. 128, n.3, pp. 970-977, 2002.
- MURTHY, Siva S.; ZILINSKAS, Barbara A. Molecular Cloning and Characterization of a cDNA Encoding Pea Monodehydro-ascorbate Reductase. *J. Biol. Chem.*, v. 269, n. 49, pp. 31129-31133, 1994.
- NOCTOR, Graham; FOYER, Christine H. Ascorbate and Glutathione: Keeping Active Oxygen Under Control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, v. 49, pp. 249-79, 1998.
- OBARA, Keisuke; SUMI, Kazuyoshi; FUKUDA, Hiroo. The use of multiple transcription starts causes the dual targeting of *Arabidopsis* putative monodehydroascorbate reductase to both mitochondria and chloroplast. *Plant Cell Physiol*, v. 43, n. 7, pp. 697-705, 2002.
- OHYA, Toshihide; MORIMURA, Yoko; SAJI, Hikaru; MIHARA, Toshikazu; IKAWA, Tomoyoshi. Purification and characterization of ascorbate peroxidase in roots of *Japanese radish*. *Plant Science*, v. 125, pp.137-14, 1997.
- OSTERGAARD, J.; PERSIAN, G.; DAVEY, M.W.; BAUW, G.; MONTAGU, M.V. Isolation of a cDNA coding for L-galactono- γ -lactone dehydrogenase, an enzyme involved in the biosynthesis of ascorbic acid in plants. *J. Biol. Chem.*, v. 272, n.48, pp. 30009-30016, 1997.
- PALLANCA, Jave E.; SMIRNOFF, Nicholas. Ascorbic acid metabolism in pea seedlings. A comparison of D-Glucosone, L-sorbosone and L-galactono-1,4-lactone as ascorbate precursors. *Plant Physiol.*, v. 120, pp. 453-461, 1999.

- SAITO, Kazumi; NICK, J.A.; LOEWUS, Frank A. D-Glucosone and L-Sorbosone, Putative Intermediates of L-Ascorbic Acid Biosynthesis in Detached Bean and Spinach Leaves. *Plant Physiol.* v. 94, pp. 1496-1500, 1990.
- SAKIHAMA, Yasuko; MANO, Jun'ichi; SANO, Satoshi; ASADA, Cosí; YAMASAKI, Hideo. Reduction of Phenoxyl Radicals Mediated by Monodehydroascorbate Reductase. *Bioch. Bioph. Res. Comm.*, v. 279, pp. 949-954, 2000.
- SIENDONES, E.; GONZÁLEZ-REYES, J.A.; SANTOS-OCAÑA, C.; NAVAS, P.; CÓRDOBA, F. Biosíntesis of ascorbic acid in Kidney bean. L-galactono- γ -lactone dehydrogenase is an intrinsic protein located at mitochondrial inner membrane. *Plant Physiol.*, v. 120, pp. 907-912, 1999.
- SMIRNOFF, Nicholas. The function and metabolism of ascorbic acid in plants. *Annals of Botany*, v. 78, pp. 661-669, 1996.
- SMIRNOFF, Nicholas. Ascorbic acid: metabolism and functions of a multi-facetted molecule. *Cur. Opin. Plant Biol.*, v. 3, pp. 229-235, 2000.
- SMIRNOFF, Nicholas; WHEELER, Glen L. Ascorbic acid in plants: biosynthesis and function. *Crit. Rev. Biochem. Mol. Biol.*, v. 35, n. 4, pp. 291-314, 2000.
- SOHN, Soo-In; KIM, Joon-Chul; LEE, Kwang-Woong; RHEE, Hae-Ik; WANG, Myeong-Hyeon. Molecular cloning and expression of a cDNA encoding cytosolic ascorbate peroxidase from *Pimpinella brachycarpa*. *J. Plant Physiol.*, v. 159, pp. 1029-1035, 2002.
- TABATA, Kazufumi; TAKAOKA, Takahiko; ESAKA, Muneharu. Gene expression of ascorbic acid-related enzymes in tobacco. *Phytochemistry*, v. 61, pp. 631-635, 2002.
- VANDAMME, E.J. Production of vitamins, coenzymes and related biochemicals by biotechnological processes. *J. Chem. Technol. Biotechnol.*, v. 53, n. 3, pp. 313-327, 1992.
- WHEELER, Glen L.; JONES, Mark A.; SMIRNOFF, Nicholas. The biosynthetic pathways of vitamin C in higher plants. *Nature*, v. 393, pp. 365-369, 1998.
- WOLUCKA, B.A.; DAVEY, M.W.; BOERJAN, W. A High-Performance Liquid Chromatography Radio Method for Determination of L-Ascorbic Acid and Guanosine 59-Diphosphate-L-Galactose, Key Metabolites of the Plant Vitamin C Pathway. *Analytical Biochemistry*, v. 294, pp. 161-168, 2001.
- YOON, Ho-Sung; LEE, Hyoshin; LEE, In-Ae; KIM, Ki-Yong; JO, Jinki. Molecular cloning of the monodehydroascorbate reductase gene from *Brassica campestris* and analysis of its mRNA level in response to oxidative stress. *Bioch. Bioph. Acta*, v. 1658, pp. 181-186, 2004.
- YOSHIMURA, Kazuya; ISHIKAWA, Takahiro; NAKAMURA, Yoshihiro; TAMOI, Masahiro; TAKEDA, Toru; TADA, Toshiji; NISHIMURA, Keüchiro; SHIGEOKA, Shigeru. Comparative Study on Recombinant Chloroplastic and Cytosolic Ascorbate Peroxidase Isozymes of Spinach. *Arch. Biochem. Biophys.*, v. 353, n. 1, pp. 55-63, 1998.
- YUYAMA, Kaoru; AGUIAR, Jaime P. L.; YUYAMA, Lúcia K. O. Camu-camu: um fruto fantástico como de vitamina C. *Acta Amazônica*, v. 32, n. 1, pp. 169-174, 2002.
- YUYAMA, Kaoru; FALCÃO, Nilton S.; YUYAMA, Lúcia K. O.; AGUIAR, Jaime P. L.; MIRANDA, Ires P. A.; OLIVEIRA, Luiz A.; CLEMENT, Charles R.; SILVA FILHO, Danilo F.; ALMEIDA, Elionor R. P.; CAETANO, Aldenir C.; ALVES, Jaime C.; COSTA, Suely S.; KERR, Warwick E.; CARVALHO, Gislene; RIBEIRO, Joana D.; BARBOSA, Edelcílio M.; CALIRI, Guilherme J. A.; SOUZA, Claudia M.; SILVA, Marcicleide L.; SANTANA, Samuel C.; et al. Domesticação de germoplasma de camu-camu - *Myrciaria dubia* (H.B.K.) McVaugh - para uso em agroindústria na Amazônia. *In.*: Programa Piloto para a Proteção das Florestas Tropicais do Brasil - PPG, pp. 149-153, 2001.
- ZOU, Liping Zou; LI, Hanxia; OUYANG, Bo; ZHANG, Junhong; YE, Zhibiao. Cloning and mapping of genes involved in tomato ascorbic acid biosynthesis and metabolism. *Plant Science*, v. 170, pp. 120-127, 2006.

OUTROS GENES INTERESSE BIOTECNOLÓGICO E PERSPECTIVAS FUTURAS

Os estágios de desenvolvimento do fruto, bem como, os fatores bióticos e abióticos levam as células vegetais a expressarem uma gama de genes com finalidades específicas. No fruto do tomate, por exemplo, uma série de eventos fisiológicos ocorrem durante o amadurecimento como a produção de etileno, modificações na parede celular, degradação da clorofila, biossíntese de carotenóides e amolecimento dos tecidos, refletindo as significativas mudanças que ocorrem na expressão de genes durante o desenvolvimento dos frutos (MOCTEZUMA et al., 2003).

Em camu-camu foram identificadas seqüências ESTs relacionadas ao desenvolvimento e amadurecimento do fruto, como os responsáveis pela síntese de etileno: 1-aminociclopropano-1-carboxilato oxidase e 1-aminociclopropano-1-carboxilate síntase (Tabela 01).

Etileno é um hormônio vegetal que regula o amadurecimento de frutos por coordenar a expressão de genes que são responsáveis por uma variedade de processo, incluindo o aumento em respiração, a mudanças na cor, textura, aroma e no sabor (OETIKER & YANG, 1995).

O etileno é biossintetizado a partir da S-adenosilmetionina via ácido 1-aminociclopropano-1-carboxílico (ACC), catalizada pelas enzimas ACC sintase e ACC oxidase. Ambas enzimas são limitadas em frutos pré-climatericos, porém são grandemente induzidas durante o amadurecimento (OETIKER & YANG, 1995).

Em frutos climatérico, tais como a banana, o aumento na produção de etileno esta associado com o aumento da atividade da enzima ACC oxidase seguida da ACC síntase. Quando etileno exógeno foi usado para promover o amadurecimento da banana, a ACC oxidase foi induzida após 4 horas e a ACC síntase foi induzida 8 horas depois (LOPEZ-GOMES, 1997).

Definição	Organismo	GI	E-value
Amadurecimento do Frutos	Biossíntese de antocianina		
Fenilalanina amonoa-liase	<i>Camellia sinensis</i>	1171998	3E-154
Chalcona síntase	<i>Dictamnus albus</i>	54311699	2E-100
Chalcona síntase 2 (Naringenina-chalcona síntase 2)	<i>Camellia sinensis</i>	1345787	2E-85
Flavanona 3-hidroxilase	<i>Camellia sinensis</i>	49861114	1E-89
Antocianidina síntase	<i>Malus x domestica</i>	21038958	4E-64
Leucoantocianidina dioxigenase	<i>Vitis labrusca x Vitis vinifera</i>	22266677	2E-82
2'-hidroxi isoflavona/diidroflavonol redutase homologa	<i>Glycine max</i>	6573167	4E-47
2-hidroxi isoflavanona desidratase	<i>Glycine max</i>	56692180	2E-36
Provável diidroflavonol redutase	<i>Ipomoea trifida</i>	45935133	9E-63
Flavonol síntase/ligada ao íon ferro / isopenicilina-N síntase	<i>Arabidopsis thaliana</i>	30685157	2E-70
	Biossíntese de etileno		
1-aminociclopropano-1-carboxilato oxidase	<i>Hevea brasiliensis</i>	37719660	3E-52
1-aminociclopropano-1-carboxilate síntase	<i>Petunia x hybrida</i>	20508	2E-71
Mecanismo de defesa celular	Estresse oxidativo		
Catalase	<i>Prunus persica</i>	32526568	0.0
Ascorbato peroxidase estroma cloroplastídico	<i>Vigna unguiculata</i>	45268439	8E-14
Ascorbato peroxidase citosólica	<i>Fragaria x ananassa</i>	2754860	1E-82
Glutaciona peroxidase	<i>Hevea brasiliensis</i>	18026892	5E-56
superoxide dismutase [Fe]	<i>Lycopersicon esculentum</i>	33413303	3E-51
Fe-SOD	<i>Cinnamomum camphora</i>	3599469	3E-20
iron superoxide dismutase 3	<i>Arabidopsis thaliana</i>	21554867	10E-22
Monodeidroascorbato redutase	<i>Brassica rapa subsp. pekinensis</i>	14764532	2E-86
	Contra Patógenos		
Quitinase	<i>Vitis vinifera</i>	10880381	8E-88
Quitinase 1 precursor (quitinase 1 de bulbo de Tulipa) (TBC-1)	<i>Tulipa bakeri</i>	47605559	9E-42
Proteína como quitinase	<i>Gossypium hirsutum</i>	34016877	2E-109
thaumatin-like protein	<i>Actinidia deliciosa</i>	71057064	3E-96
SAR1/GTP-binding secretory factor	<i>Arabidopsis thaliana</i>	2104532	4E-24
SAR DNA binding protein	<i>Oryza sativa</i>	3288883	7E-32
Proteína 10 relacionada a patogêneses	<i>Vitis vinifera</i>	11182126	1E-56

Tabela 01 - Relação de genes de interesse biotecnologia em frutos.

A coloração do camu-camu é oriunda da antocianina, o que justifica a identificação de onze diferentes seqüências ESTs relacionadas a biossíntese deste importante corante natural, entre elas: Leucoantocianidina dioxigenase; Chalcona síntase; Antocianidina síntase; Flavanona 3-hidroxilase; provável Diidroflavonol redutase; entre outras (Tabela 01).

O processo de maturação do fruto e conseqüente acumulação de antocianina, por exemplo, é influenciada por estímulo ambiental, tais como a luz, temperatura e nutrição, bem como fatores genéticos. O mecanismo regulatório da biossíntese de antocianina tem atrativo interesse porque a coloração vermelha da casca é um importante fator na aceitação do mercado para muitas espécies vegetais, como de maçã e uva, por exemplo (HONDA et al., 2002).

Em uva (*Vitis vinifera* L.) os genes da via biossintética da antocianina tem sido investigado cuidadosamente, e a UDP glicose: flavonóide 3-O-glucosiltransferase (UFGLuT) foi identificada como a enzima chave no controle da cor vermelha em casca de uvas. Em maçã foi descrito o isolamento de cinco genes de enzimas da via biossintética da antocianina: chalcone sintase (CHS), flavanona 3-hidroxilase (F3H), dihidroflavonol 4-redutase (DFR) e antocianina sintase (ANS) e a UFGLuT. Estes cinco genes são, coordenadamente, expressos durante o desenvolvimento do frutos de maçã e os nível de expressão está positivamente correlacionado com o grau de concentrações de antocianinas em frutos maduros das cultivares Orin, Fuji e Jonathan (HONDA et al., 2002) .

Em frutos de camu-camu o mecanismo de defesa encontra-se ativo observado pela expressão de genes relacionados a proteção celular contra estresse oxidativo e contra patógenos. Enzimas como: Catalase, ascorbato peroxidase do estroma de cloroplasto, glutathiona peroxidase, glutathiona S-transferase, glutathiona síntase e

monodeidroascorbato redutase foram identificadas em seqüências ESTs de camu-camu, relacionadas a defesa contra estresse oxidativo (Tabela 1).

As células desenvolveram um mecanismo de defesa contra os produtos tóxicos produzidos por elas própria. O peróxido de hidrogênio (H_2O_2), superóxido (O_2^-) e o radical hidroxil (OH^\cdot) que são coletivamente chamados de espécies reativas de oxigênio (*Reactive Oxygen Species* – ROS) e participam em múltiplos processos nas plantas, sendo formados como produtos tóxicos da respiração e fotossíntese. A produção em excesso destas ROS e nitrogênio provocam nas plantas o chamado estresse oxidativo (INZE & MONTAGU, 1995; BOLWELL, 1999).

As fontes para a produção de espécies reativas de oxigênio é a transferência de elétrons da cadeia respiratória, na mitocôndria, e pela transferência de elétrons da cadeia de transporte e processos dependentes de luz, nos cloroplastos. Os radicais superóxidos podem ser convertidos a peróxido de hidrogênio espontaneamente ou pela dismutação catalítica (NETTO, 2001).

Os principais modelos enzimáticos para a limpeza de ROS é pela superóxido dismutase (SOD), catalase (CAT), o ciclo ascorbato-glutationa e o ciclo glutaciona peroxidase (GPX). A SOD converte superóxido de hidrogênio em peróxido de hidrogênio. A CAT converte superóxido de hidrogênio em água (APEL & HIRT, 2004).

O peróxido de hidrogênio é convertido em água pelo ciclo ascorbato-glutationa. O agente redutor na primeira reação catalizada pela ascorbato peroxidase (APX) é o ascorbato, o qual é oxidado a monodeidroascorbato (MDA). A MDA redutase (MDAR) reduz o MDA em ascorbato com a ajuda do NAD(P)H. A deidroascorbato (DHA) é produzida espontaneamente pela MDA redutase com a ajuda da glutaciona (GSH) que é oxidado a glutaciona disulfeto (GSSG). O ciclo fecha com a glutaciona redutase (GR) convertendo GSSG em GSH, com agente redutor NAD(P)H. O ciclo de glutaciona peroxidase (GPX) converte peróxido de hidrogênio (água oxigenada) em água, usando

equivalentes redutores de GSH. A glutathiona disulfeto (GSSG) oxidada é novamente convertida em GSH por glutathiona redutase o agente redutor é o NAD(P)H (APEL & HIRT, 2004).

Os genes de glutathionas s-transferase são, também, induzidos pelo ataque de patógenos em vegetais, como trigo, batata e *Arabidopsis* (APEL & HIRT, 2004).

Em camu-camu entre os ESTs relacionados a defesa contra patógeno destacam-se as quitinases identificadas: quitinase, precursor de quitinase 1 (Bulbo de tulipa quitinase 1) (TBC-1) e uma proteína tipo quitinase (Tabela 01). As plantas reagem ao ataque por microrganismos, insetos e animais pela indução de genes que codificam diversas proteínas, muitas das quais desempenham papel de defesa. As quitinases, por exemplo, catalizam a hidrólise da quitina, um homopolímero linear ligado a posição β -1,4 de resíduos de acetilglucosamina. A quitina é um componente comum da parede celular de fungos e de exoesqueletos de artrópodes. As quitinases são induzidas nas plantas pelo ataque de patógenos e são importantes nas reações de defesa vegetal (SHINSHI et al 1987; HOU et al., 1998).

Estes são alguns dos outros genes de interesse biotecnológico e agrônômico encontrados na biblioteca de ESTs de frutos de camu-camu e que poderão ser explorados na domesticação e no melhoramento genética da espécie.

As perspectivas no uso da biblioteca de ESTs desenvolvida são grandes. Elas poderão ser utilizadas em programas de melhoramento genético, por meio da tecnologia do DNA recombinante, ou engenharia genética, que permite ampliar as mais diversas características vegetais pela transferência de um gene de um organismo para qualquer outro, de forma que o indivíduo receptor poderá apresentar a característica conferida pelo gene transferido ou até super-expressar um determinado gene de interesse com o uso de promotores de expressão fortes.

Estas seqüências ESTs também poderão ser utilizadas, via engenharia genética, como uma ferramenta para melhorar nutricionalmente fruteiras que já tem mercado garantido, como é o caso da banana, pela biofortificação com vitamina C, por exemplo, trabalho que já vem sendo desenvolvido pela EMBRAPA/CENARGEN.

Como o camu-camu é uma espécie não domesticada e apresenta dificuldades na seleção de plantas com caracteres agrônômicos desejáveis, em função de sua elevada variabilidade genética das plantas, outra estratégia é o uso das ESTs como marcadores funcionais.

Os marcadores funcionais são originados a partir de seqüências de genes expressos (cDNA ou ESTs), tais como os marcadores EST-PCR e EST-SSR. Os marcadores EST-PCR são co-dominantes e tem algumas vantagens, como por exemplo: se o marcador EST for geneticamente associado com uma característica de interesse, é possível que o gene afete diretamente esta característica, como o estresse oxidativo, resistência a pragas, a biossíntese de vitaminas, biossíntese de antocianina, citocromo b, entre outros (CATO et al., 2001).

Adicionalmente os marcadores EST são úteis para alinhar mapas de ligação de genoma, comparando lócus com características quantitativa (QTL) de espécies distantemente relacionadas (CATO et al., 2001). Os ESTs-PCR já foram utilizados para o monitoramento da variação genética intrapopulacionais de *Picea abies*, no qual detectaram sete pares de primers polimórficos, que mostraram uma maior variação intrapopulacional, incluindo heterosiguidase observada, quando comparado com marcador isoenzimático (SCHUBERT et al., 2001).

Os EST-SSR, por sua vez, são microsatelites provenientes de seqüências EST - genes que transcritos e com provável função que pode ser deduzida por homologia via busca em bancos de dados. Estes EST-SSRs ajudam a determinar a variabilidade funcional em populações naturais e em bancos de germoplasma, além disso são

altamente transferíveis a outras espécies dentro da mesma família ou em outras famílias, porque são derivados de regiões transcritas altamente conservadas (CORDEIRO et al 2001).

A obtenção de plantas melhoradas geneticamente poderá proporcionar a comunidade variedades altamente produtivas e resistentes a pragas, sendo um estímulo para o desenvolvimento de cultivos pelas comunidades amazônicas em terra-firme e na várzea.

A várzea por ser um ambiente altamente fértil é ideal para o desenvolvimento de camu-camu, visto que o mesmo consegue sobreviver submerso durante os períodos das cheias. Contudo, o período de frutificação geralmente ocorre quando as plantas estão quase submersas, sendo esta uma barreira que precisa ser superada pelos melhoristas. Os marcadores funcionais poderão auxiliar programas de melhoramento genético que buscam a precocidade na frutificação de plantas cultivadas em várzea.

A amplitude dos estudos e perspectivas na utilização de ESTs com aplicabilidade direta nos vegetais cresce a cada dia. O estudo genético do camu-camu ainda precisa ser mais explorado, entretanto, deve-se destacar a importância da valiosa contribuição do estudo de ESTs como marco inicial para futuras descobertas.

REFERÊNCIAS BIBLIOGRÁFICAS

APEL, Klaus; HIRT, Heribert. Reactive oxygen species: Metabolism, Oxidative Stress, and Signal Transduction. *Annu. Rev. Plant Biol.*, v. 55, pp. 373-399, 2004.

BOLWELL, G Paul. Role of active oxygen species and NO in plant defence responses. *Cur. Opin. Plant Biol.*, v. 2, n. 4, pp. 287-294, 1999.

CATO, S. A.; GARDNER, R. C.; KENT, J.; RICHARDSON, T. E. A rapid PCR-based method for genetically mapping ESTs. *Theor Appl Genet.*, v. 102, pp. 296-306, 2001.

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*, v. 160, pp. 1115-1123, 2001.

HONDA, Chikako; KOTOBA, Nobuhiro; WABA, Masato Wada; KONDO, Satoru; KOBAYASHI, Shozo; SOEJIMA, Junichi; ZHANG, Zilian; TSUDA, Tomomi; MORIGUCHI, Takaya Moriguchi. Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. *Plant Physiol. Biochem.*, v. 40, n.11, pp. 955-962, 2002.

HOU, Wen-Chi; CHEN, Ying-Chou; LIN, Yaw-Huei. Chitinase activity of sweet potato (*Ipomoea batatas* [L.] Lam var, Tainong 57). *Bot. Bull. Acad. Sin.*, v. 39, pp. 93-97, 1998.

INZÉ, Dirk; MONTAGU, Marc Van. Oxidative stress in plants. *Cur. Opin. Biotechnol.*, v. 6, n.2, pp. 153-158, 1995.

LOPES-GOMES, Rodolfo; CAMPBELLO, Alan; DONGB, Jiang-Guo; YANG, Shang-Fa; GOMES-LIM, Miguel Angel. Ethylene biosynthesis in banana fruit: isolation of a genomic clone to ACC oxidase and expression studies. *Plant Science*, v. 123, pp. 123-131, 1997.

MOCTEZUMA, Edgar; SMITH, David L.; CROSS, Kenneth C. Effect of ethylene on mRNA abundance of three galactosidase genes in wild type and mutant tomato fruit. *Postharvest Biology and Technology*, v. 28, pp. 207-217, 2003.

NETTO, Luis E. S. Oxidative stress response in sugarcane. *Genet.Mol. Biol.*, v. 24, n. 1-4, pp. 93-102, 2001.

OETIKER, J. H.; YANG, S. F. The role of ethylene in fruit ripening. *ISHS Acta Horticulturae Postharvest Physiology of Fruits* v. 398, pp. 167-178, 1995. (Abstract) <http://www.actahort.org/books/398/398_17.htm> acesso em 01.04.2006.

SCHUBERT, R.; MUELLER-STARCK, G.; RIEGEL, R. Development of EST-PCR markers and monitoring their intrapopulation genetic variation in *Picea abies* (L.) Karst. *Theor. Appl. Genet.*, v. 103, n. 8, pp. 1223-1231, 2001.

SHINSHI, Hideaki; MOHNEN, Debra; MEINS JR, Frederick. Regulation of a plant pathogenesis-related enzyme: Inhibition of chitinase and chitinase mRNA accumulation in cultured tobacco tissues by auxin and cytokinin. *PNAS*, v. 84, pp. 89-93, 1987.

APENDICE 1

CÓDIGO	No. ESTs	DEFINIÇÃO	ORGANISMO	GI	e- value	KO
MDI05-A09	1	(1-4)-beta-mannan endohydrolase	<i>Coffea arabica</i>	10178872	2E-22	S
CL164 contig1	3	(S)-norcochlorine synthase	<i>Thalictrum flavum subsp. glaucum</i>	38492394	3E-19	S
CL9 contig1	9	(S)-norcochlorine synthase	<i>Thalictrum flavum subsp. glaucum</i>	38492394	2E-17	S
MDI16_H09	1	101 kDa heat shock protein; HSP101	<i>Nicotiana tabacum</i>	11561806	6E-26	O
MDI08-B08	1	14-3-3 brain protein homolog	<i>Vicia faba</i>	695765	1E-33	O
MDI20_H12	1	14-3-3 family protein	<i>Malus x domestica</i>	5537759	2E-64	O
MDI15_H04	1	14-3-3 family protein	<i>Lycopersicon esculentum</i>	15637114	1E-58	O
CL63 contig1	4	14-3-3 family protein	<i>Malus x domestica</i>	55375985	7E-73	O
MDI10-A09	1	14-3-3 family protein	<i>Malus x domestica</i>	55375985	7E-94	O
CL439 contig1	2	14-3-3 protein	<i>Populus x canescens</i>	8099061	8E-56	O
MDI37_F06	1	14-3-3 protein	<i>Populus x canescens</i>	8515888	3E-50	O
CL481 contig1	2	14-3-3 protein	<i>Populus x canescens</i>	8515890	1E-82	O
MDI32-A01	1	14-3-3-like protein B	<i>Glycine max</i>	3023195	2E-43	O
MDI24_B06	1	14-3-3-like protein D (SGF14D)	<i>Glycine max</i>	3023197	3E-16	O
MDI37_C09	1	16.9 kDa oleosin	<i>Theobroma cacao</i>	21311553	2E-18	S
MDI25_G05	1	17.7 kDa heat shock protein	<i>Carica papaya</i>	37933812	2E-15	O
MDI33_B02	1	17.7 kDa heat shock protein	<i>Carica papaya</i>	37933812	4E-33	O
MDI42_D11	1	17.7 kDa heat shock protein	<i>Carica papaya</i>	37933812	2E-61	O
MDI48-H11	1	17.7 kDa heat shock protein	<i>Carica papaya</i>	37933812	2E-59	O
CL228 contig1	2	17.8 kDa class II heat shock protein, putative	<i>Oryza sativa</i>	50915316	4E-31	O
CL7 contig2	7	18.2 kDa class I heat shock protein	<i>Medicago sativa</i>	123556	2E-74	O
CL256 contig1	2	18.5 kDa class I heat shock protein (HSP 18.5)	<i>Glycine max</i>	18654	2E-42	O
CL7 contig1	5	18.5 kDa class I heat shock protein (HSP 18.5)	<i>Glycine max</i>	18654	6E-56	O
CL469 contig1	2	18S ribosomal RNA gene	<i>Heteropyxis natalensis</i>	7595447	1E- 150	J
CL312 contig1	2	1-aminocyclopropane-1-carboxylate oxidase	<i>Hevea brasiliensis</i>	37719660	5E-68	S
MDI51-E08	1	1-aminocyclopropane-1-carboxylate oxidase	<i>Hevea brasiliensis</i>	37719660	6E-52	S
MDI20_H10	1	1-aminocyclopropane-1-carboxylate synthase	<i>Petunia x hybrida</i>	20508	2E-71	GT
MDI25_C05	1	1-deoxy-D-xylulose-5-phosphate reductoisomerase	<i>Catharanthus roseus</i>	7621710	3E-95	S
CL51Contig2	3	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (Phosphoglyceromutase) (PGAM-I)	<i>Mesembryanthemum crystallinum</i>	3914394	2E-85	GT
MDI32-H11	1	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (Phosphoglyceromutase) (PGAM-I)	<i>Oryza sativa</i>	3914394	2E-85	GT
MDI10-G11	1	20S proteasome alpha 6 subunit	<i>Nicotiana benthamiana</i>	22947842	5E-19	O
MDI18_A02	1	20S proteasome beta subunit PBC2, putative	<i>Arabidopsis thaliana</i>	12744997	4E-82	O

MDI29_F09	1	21D7 antigen	<i>Daucus carota</i>	217911	2E-22	T
CL50 contig1	5	22.0 kDa class IV heat shock protein precursor	<i>Glycine max</i>	18661	9E-34	O
CL239 contig1	2	26S proteasome AAA-ATPase subunit RPT3	<i>Arabidopsis thaliana</i>	8777330	8E-47	O
MDI31_E11	1	26S proteasome regulatory particle non-ATPase subunit8	<i>Oryza sativa</i>	17297979	8E-10	O
CL205 contig1	2	26S proteasome subunit 4-like	<i>Solanum tuberosum</i>	77745479	0.0	O
CL273 contig1	2	26S proteasome subunit 8; Tat binding protein	<i>Fagus sylvatica</i>	6599051	2E-101	O
MDI47-A08	1	2-C-methyl-D-erythritol 4-phosphate Cytidyltransferase - MEPCT/ISPD; 4-diphosphocytidyl-2C-methyl-D-erythritol synthase	<i>Arabidopsis thaliana</i>	18395376	4E-17	S
MDI38_H03	1	2'-hydroxy isoflavone/dihydroflavonol reductase homolog	<i>Glycine max</i>	6573167	4E-47	S
CL308Contig1	2	2-hydroxyisoflavanone dehydratase	<i>Glycine max</i>	56692180	2E-36	S
MDI05-G09	1	2-on-2 hemoglobin	<i>Datisca glomerata</i>	31559437	2E-71	S
CL314 contig1	2	2-oxoacid-dependent dioxygenase	<i>Prunus mume</i>	82568691	8E-83	S
MDI48-C09	1	2-oxoglutarate-dependent dioxygenase	<i>Lycopersicon esculentum</i>	66730857	2E-18	E
MDI39_E02	1	2-oxoglutarate-dependent dioxygenase	<i>Lycopersicon esculentum</i>	66730858	2E-26	E
CL44 contig1	5	2S albumin seed storage protein	<i>Juglans nigra</i>	31321942	1E-14	S
MDI02-G11	1	2S albumin seed storage protein	<i>Juglans nigra</i>	31321942	1E-10	S
MDI18_A07	1	2S albumin seed storage protein	<i>Juglans nigra</i>	31321942	2E-11	S
MDI43_B07	1	34 kDa outer mitochondrial membrane protein porin-like protein	<i>Solanum tuberosum</i>	83283993	3E-20	S
MDI48-F03	1	39 kDa EF-Hand containing protein	<i>Solanum tuberosum</i>	3550661	2E-54	S
MDI22_C07	1	3-dehydroquininate dehydratase / shikimate dehydrogenase isoform 2	<i>Nicotiana tabacum</i>	57920902	5E-16	E
MDI25_E01	1	3-dehydroquininate dehydratase / shikimate dehydrogenase isoform 2	<i>Nicotiana tabacum</i>	57920902	2E-23	S
MDI08-E03	1	3-deoxy-D-arabino heptulosonate-7-phosphate synthase	<i>Oryza sativa</i>	46359608	2E-51	S
MDI09-C09	1	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase 3, putative	<i>Fagus sylvatica</i>	76782190	3E-53	S
MDI14_B04	1	3-hydroxyacyl-[acyl-carrier protein] dehydratase/ catalytic	<i>Arabidopsis thaliana</i>	18399910	6E-34	S
MDI13_A01	1	3-hydroxyisobutyryl-coenzyme A hydrolase-like protein	<i>Arabidopsis thaliana</i>	7268039	1E-84	GE
CL234 contig1	2	3-isopropylmalate dehydratase large subunit, putative	<i>Oryza sativa</i>	50904927	2E-89	E
CL243 contig1	2	40S ribosomal protein S13	<i>Glycine max</i>	44662862	2E-49	J
CL343 contig1	2	40S ribosomal protein S15	<i>Oryza sativa</i>	218131	8E-30	J
MDI24_D02	1	40S ribosomal protein S15A, putative	<i>Oryza sativa</i>	5025167	1E-25	J
MDI22_A08	1	40S ribosomal protein S17-like	<i>Arabidopsis thaliana</i>	79326995	7E-42	J
MDI49-D12	1	40S ribosomal protein S19	<i>Cicer arietinum</i>	3860315	5E-24	J
MDI20_F11	1	40S ribosomal protein S19-like	<i>Solanum tuberosum</i>	82623403	7E-65	J
MDI51-E03	1	40S ribosomal protein S2	<i>Picea abies</i>	12580855	6E-23	J
MDI47-F03	1	40S Ribosomal Protein S20 homolog	<i>Arabidopsis thaliana</i>	5541704	5E-52	J
CL264 contig1	2	40S ribosomal protein S5 [<i>Cicer arietinum</i>]	<i>Cicer arietinum</i>	3043428	2E-91	J
MDI07-C07	1	40S ribosomal protein SA (p40)	<i>Glycine max</i>	3334320	1E-46	J
MDI48-G12	1	40S ribosomal protein, putative	<i>Oryza sativa</i>	50939157	7E-76	J

MDI30_D11	1	40S ribosomal S4 protein	<i>Glycine max</i>	22138108	7E-36	J
MDI49-F03	1	4-coumarate:coenzyme A ligase	<i>Eucalyptus camaldulensis</i>	73665529	2E-84	S
MDI21_F11	1	4-coumarate-CoA ligase, putative	<i>Lycopersicon chilense</i>	56713950	2E-34	S
MDI21_A12	1	4-coumarate-CoA ligase-like protein	<i>Arabidopsis thaliana</i>	21593705	1E-22	S
MDI01-G08	1	4-coumarate-CoA ligase-like protein	<i>Arabidopsis thaliana</i>	21593705	1E-19	S
MDI07-E04	1	4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis protein, putative	<i>Arabidopsis thaliana</i>	21536528	3E-71	S
CL361 contig1	2	50S ribosomal protein L24	<i>Oryza sativa</i>	77552956	9E-28	J
CL229 contig1	2	5-enolpyruvylshikimate 3-phosphate synthase, putative	<i>Fagus sylvatica</i>	76782198	3E-67	S
MDI34_A03	1	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase (Vitamin-B12-independent methionine synthase isozyme) (Cobalamin-independent methionine synthase isozyme)	<i>Catharanthus roseus</i>	81134570	1E-40	S
MDI23_H04	1	5-oxoprolinase, similar	<i>Oryza sativa</i>	34905886	3E-117	S
MDI48-G03	1	60s acidic ribosomal protein	<i>Prunus dulcis</i>	24473796	4E-22	J
CL168 contig1	3	60S acidic ribosomal protein P3, putative	<i>Oryza sativa</i>	53792736	8E-24	J
MDI01-E07	1	60S ribosomal protein L11, putative	<i>Oryza sativa</i>	53749326	6E-89	J
CL255 contig1	2	60S ribosomal protein L12	<i>Capsicum annuum</i>	40287508	5E-48	J
CL305 contig1	2	60S ribosomal protein L13E	<i>Picea abies</i>	12580867	2E-52	J
MDI11_C03	1	60S ribosomal protein L15, probable	<i>Picea mariana</i>	2982318	4E-44	J
MDI40_F04	1	60s ribosomal protein L15, putative	<i>Oryza sativa</i>	50878349	7E-79	J
MDI44-A07	1	60s ribosomal protein L15, putative	<i>Oryza sativa</i>	50878349	2E-23	J
CL331 contig1	2	60S ribosomal protein L18a	<i>Castanea sativa</i>	13430184	1E-89	J
MDI07-B05	1	60S ribosomal protein L18a	<i>Castanea sativa</i>	13430184	1E-68	J
CL397 contig1	2	60S ribosomal protein L19	<i>Capsicum annuum</i>	40287526	2E-82	J
MDI07-H01	1	60S ribosomal protein L21	<i>Oryza sativa</i>	31432571	2E-70	J
MDI45-C06	1	60S ribosomal protein L21	<i>Oryza sativa</i>	31432571	1E-50	J
MDI35_E06	1	60S ribosomal protein L23A	<i>Arabidopsis thaliana</i>	3355475	4E-30	J
MDI46-H09	1	60S ribosomal protein L31, putative	<i>Oryza sativa</i>	54290930	3E-26	J
MDI05-A03	1	60S ribosomal protein L34	<i>Pisum sativum</i>	730557	2E-48	J
MDI19_E04	1	60S ribosomal protein L34	<i>Nicotiana tabacum</i>	730558	1E-20	J
CL214Contig1	2	60S ribosomal protein L35, putative	<i>Rheum australe</i>	71535092	4E-44	J
MDI23_H09	1	60S ribosomal protein L39, putative	<i>Oryza sativa</i>	50915614	9E-24	J
MDI43-H04	1	60S ribosomal protein L39, putative	<i>Oryza sativa</i>	50915614	6E-20	J
MDI16_E05	1	60S ribosomal protein L5	<i>Cucumis sativus</i>	34484312	2E-109	J
MDI26_E05	1	60S ribosomal protein L5	<i>Cucumis sativus</i>	34484312	3E-49	J
MDI46-C09	1	60S ribosomal protein L5	<i>Cucumis sativus</i>	34484312	4E-38	J
MDI49-F05	1	60S ribosomal protein L6 (YL16-like)	<i>Mesembryanthemum crystallinum</i>	464621	2E-48	J
MDI36_G04	1	60S ribosomal protein L7	<i>Arabidopsis thaliana</i>	20148739	1E-74	J

MDI29_E01	1	65kD microtubule associated protein	<i>Daucus carota</i>	28200790	4E-76	T
MDI15_F02	1	6-phosphofructokinase	<i>Arabidopsis thaliana</i>	15221156	4E-74	G
MDI12_B11	1	6-phosphogluconolactonase, putative	<i>Oryza sativa</i>	50725145	9E-36	G
CL3 contig1	15	70 kDa heat shock cognate protein 1	<i>Vigna radiata</i>	45331281	0.0	O
MDI23_F02	1	70 kDa peptidylprolyl isomerase, putative	<i>Oryza sativa</i>	50947791	9E-68	O
MDI02-B05	1	AAA-type ATPase, putative	<i>Oryza sativa</i>	19387258	9E-27	O
CL212 contig1	2	AAA-Type ATPase, salt-induced	<i>Mesembryanthemum crystallinum</i>	37894600	8E-27	O
MDI13_A04	1	ABA 8'-hydroxylase CYP707A2	<i>Solanum tuberosum</i>	76803521	6E-30	S
MDI32-H08	1	ABA Insensitive 2 - ABI2; protein phosphatase type 2C	<i>Arabidopsis thaliana</i>	15242022	1E-21	S
MDI18_F03	1	ABA-glucosyltransferase	<i>Vigna angularis</i>	18151384	4E-30	S
MDI46-C04	1	ABC transporter, putative	<i>Arabidopsis thaliana</i>	7267235	8E-50	Q
MDI12_B10	1	ABC transporter-like protein	<i>Brassica rapa</i>	83032241	4E-46	Q
MDI25_E04	1	ABC1 family protein-like	<i>Oryza sativa</i>	50898992	1E-49	Q
MDI51-E02	1	abscisic stress ripening-like protein	<i>Prunus persica</i>	16588758	7E-17	S
MDI19_H10	1	acetyl Co-A acetyltransferase	<i>Hevea brasiliensis</i>	16417944	2E-48	G
MDI06-F03	1	acetylglutamate kinase/ glutamate 5-kinase	<i>Arabidopsis thaliana</i>	15230338	2E-64	E
MDI05-D09	1	aci-reductone dioxygenase-like protein	<i>Oryza sativa</i>	61807497	3E-66	S
MDI02-H05	1	aconitase, putative	<i>Prunus avium</i>	34851120	8E-83	CE
MDI29_G02	1	Actin	<i>Stevia rebaudiana</i>	23955912	7E-88	S
CL119 contig1	3	Actin	<i>Ricinus communis</i>	38259662	4E-75	S
MDI19_D09	1	actin binding	<i>Arabidopsis thaliana</i>	15240762	5E-30	DZ
CL201 contig1	2	actin depolymerizing factor	<i>Gossypium hirsutum</i>	68166241	1E-63	S
MDI32-B11	1	actin depolymerizing factor, putative	<i>Sorghum bicolor</i>	48374972	1E-42	S
CL285 contig1	2	actin-depolymerizing factor 2	<i>Petunia x hybrida</i>	14906210	2E-60	S
MDI33_G01	1	acyl carrier protein	<i>Cuphea lanceolata</i>	454261	4E-23	CIQ
MDI06-E04	1	acyl carrier protein	<i>Casuarina glauca</i>	1838961	4E-32	CIQ
MDI10-H02	1	acyl-CoA binding AT4G24230 transcript variant	<i>Arabidopsis thaliana</i>	42567094	2E-78	S
MDI18_E10	1	acyl-CoA oxidase	<i>Glycine max</i>	15553480	3E-34	S
MDI33_F07	1	acyl-CoA synthetase long chain, putative	<i>Oryza sativa</i>	55168156	9E-32	I
MDI16_E11	1	acyl-CoA synthetase, putative	<i>Capsicum annuum</i>	16797908	2E-36	I
MDI07-C10	1	acyl-CoA synthetase, putative long chain	<i>Oryza sativa</i>	55168156	2E-32	I
MDI12_H11	1	acyl-CoA-binding protein	<i>Ricinus communis</i>	1938236	9E-35	I
MDI33_G03	1	acyltransferase/ dihydrolipoyllysine-residue acetyltransferase/ protein binding	<i>Arabidopsis thaliana</i>	18404837	4E-20	S
MDI19_E06	1	acyltransferase/ dihydrolipoyllysine-residue acetyltransferase/ protein binding	<i>Arabidopsis thaliana</i>	18400212	2E-36	S
MDI43-D10	1	adapitin protein, putative	<i>Ipomoea trifida</i>	45935135	5E-99	S
CL320 contig1	2	adenosine kinase-like protein	<i>Oryza sativa</i>	29367547	6E-71	S
MDI17_B02	1	adenosine-5'-phosphosulfate-kinase	<i>Catharanthus roseus</i>	2832300	1E-28	S

MDI37_F05	1	<i>Adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase)</i>	<i>Medicago sativa</i>	1710838	2E-82	E
MDI25_G11	1	<i>adenylosuccinate synthase</i>	<i>Nicotiana tabacum</i>	37983671	2E-103	S
MDI32-G06	1	<i>adenylylsulfate kinase, putative</i>	<i>Arabidopsis thaliana</i>	6006852	3E-10	S
MDI08-G07	1	<i>ADP, ATP carrier protein precursor</i>	<i>Picea abies</i>	12580863	7E-66	C
MDI10-B04	1	<i>ADP-ribosylation factor 002</i>	<i>Daucus carota</i>	77812440	4E-94	TZ
CL31 contig1	6	<i>ADP-ribosylation factor 1</i>	<i>Medicago truncatula</i>	56311394	7E-96	TZ
MDI12_A01	1	<i>ADP-ribosylation factor, putative</i>	<i>Oryza sativa</i>	50912455	6E-59	TZ
MDI33_D07	1	<i>ADP-ribosylation factor-directed GTPase activating protein, putative</i>	<i>Oryza sativa</i>	50910930	1E-48	TZ
MDI03-A03	1	<i>Agamous homolog</i>	<i>Liquidambar styciflua</i>	5031217	2E-32	S
MDI16_B10	1	<i>AKIN gamma</i>	<i>Medicago truncatula</i>	32364482	5E-54	S
MDI08-A02	1	<i>Alanine:Glyoxylate Aminotransferase 2 - AGT2; pyridoxal phosphate binding / transaminase</i>	<i>Arabidopsis thaliana</i>	18420498	1E-32	E
MDI37_E09	1	<i>Albino 3 - ALB3; protein translocase</i>	<i>Arabidopsis thaliana</i>	30684018	3E-67	S
CL263 contig1	2	<i>alcohol acyl-transferases, putative</i>	<i>Cucumis melo</i>	57471999	7E-14	S
MDI14_A05	1	<i>alcohol dehydrogenase</i>	<i>Lotus corniculatus var. japonicus</i>	51587338	4E-65	GQ
MDI21_E01	1	<i>alcohol dehydrogenase</i>	<i>Citrus x paradisi</i>	68138959	3E-70	GQ
MDI50-E06	1	<i>alcohol dehydrogenase</i>	<i>Alnus glutinosa</i>	71793966	2E-63	GQ
CL12 contig1	8	<i>alcohol dehydrogenase 2</i>	<i>Vitis vinifera</i>	18027092	0.0	GQ
MDI32-A10	1	<i>alcohol dehydrogenase PAN2, putative</i>	<i>Oryza sativa</i>	52075618	2E-10	GQ
MDI09-G07	1	<i>aldehyde decarbonylase, possible</i>	<i>Arabidopsis thaliana</i>	1199467	4E-30	S
CL107 contig1	3	<i>aldehyde dehydrogenase (NAD+)</i>	<i>Nicotiana tabacum</i>	1743354	4E-120	CG
CL334 contig1	2	<i>aldehyde dehydrogenase 1 precursor</i>	<i>Lotus corniculatus</i>	29373073	7E-109	CG
MDI10-H01	1	<i>aldo/keto reductase</i>	<i>Fragaria x ananassa</i>	53988164	1E-19	R
MDI03-F09	1	<i>Aldolase, plastidic</i>	<i>Solanum brevidens</i>	27368033	1E-53	S
CL55 contig1	3	<i>Al-induced protein</i>	<i>Gossypium hirsutum</i>	34541994	6E-85	S
MDI11_B12	1	<i>alkaline alpha galactosidase I</i>	<i>Cucumis melo</i>	29838629	8E-103	GI
CL62 contig1	4	<i>allergen-like protein BRSn20</i>	<i>Sambucus nigra</i>	6561156	3E-45	S
CL89 contig1	3	<i>allyl alcohol dehydrogenase</i>	<i>Nicotiana tabacum</i>	6692816	3E-150	S
MDI20_B03	1	<i>allyl alcohol dehydrogenase</i>	<i>Nicotiana tabacum</i>	6692816	8E-75	S
MDI26_A01	1	<i>alpha subunit of F-actin capping protein</i>	<i>Arabidopsis thaliana</i>	15230004	1E-19	Z
CL53 contig1	3	<i>alpha tubulin 1</i>	<i>Pseudotsuga menziesii var. menziesii</i>	56481497	7E-122	Z
CL64 contig1	4	<i>alpha tubulin 1</i>	<i>Pseudotsuga menziesii var. menziesii</i>	56481497	0.0	Z
MDI05-D04	1	<i>alpha-1,4-fucosyltransferase</i>	<i>Lycopersicon esculentum</i>	18477220	1E-29	O
MDI08-F03	1	<i>alpha-amylase, putative</i>	<i>Aspergillus fumigatus Af293</i>	66853317	6E-64	G

MDI26_E08	1	<i>alpha-galactosidase, putative</i>	<i>Oryza sativa</i>	56202255	1E-35	GI
MDI22_F06	1	<i>alpha-glucosidase</i>	<i>Solanum tuberosum</i>	2648032	7E-44	M
MDI20_C11	1	<i>alpha-mannosidase</i>	<i>Arabidopsis thaliana</i>	79598781	2E-80	G
MDI25_F01	1	<i>Alpha-soluble NSF attachment protein (N-ethylmaleimide-sensitive factor attachment protein, alpha)</i>	<i>Solanum tuberosum</i>	18203271	8E-34	S
MDI24_G05	1	<i>alpha-tubulin</i>	<i>Miscanthus sinensis</i>	37936220	2E-85	Z
MDI32-A11	1	<i>alpha-tubulin-like protein</i>	<i>Solanum tuberosum</i>	76573359	7E-36	Z
MDI29_C04	1	<i>Altered meristem program 1 - AMP1 ; dipeptidase</i>	<i>Arabidopsis thaliana</i>	18410120	7E-30	S
MDI29_E02	1	<i>amidohydrolase, putative</i>	<i>Oryza sativa</i>	40714660	4E-31	GFI
MDI29_G01	1	<i>Amino acid ABC transporter, permease protein 3-TM region, His/Glu/Gln/Arg/opine amino acid or GABA permease, putative</i>	<i>Pseudomonas fluorescens Pfo-1</i>	77381205	4E-58	E
MDI43-E02	1	<i>amino acid permease</i>	<i>Oryza sativa</i>	34912494	7E-14	EU
MDI17_G01	1	<i>amino acid permease</i>	<i>Arabidopsis thaliana</i>	18422139	4E-18	EU
CL375 contig1	2	<i>Amino acid permease</i>	<i>Oryza sativa</i>	77556627	1E-95	EU
MDI08-C01	1	<i>amino acid transporter-like protein 1</i>	<i>Arabidopsis thaliana</i>	5688864	5E-21	E
MDI17_E02	1	<i>amino acid/peptide transporter</i>	<i>Prunus dulcis</i>	6635838	1E-68	E
MDI18_G08	1	<i>aminopeptidase like protein</i>	<i>Arabidopsis thaliana</i>	62321473	3E-30	ORE
MDI34_C11	1	<i>aminopeptidase N, putative</i>	<i>Oryza sativa</i>	50948547	2E-41	ORE
CL209 contig1	2	<i>AMP binding / acetate-CoA ligase</i>	<i>Arabidopsis thaliana</i>	15228909	3E-70	S
MDI21_D04	1	<i>ANAC087; transcription factor</i>	<i>Arabidopsis thaliana</i>	42567931	1E-19	S
MDI30_G10	1	<i>anionic peroxidase</i>	<i>Petroselinum crispum</i>	559235	4E-46	K
MDI27_B08	1	<i>ankyrin repeat-containing protein-like</i>	<i>Oryza sativa</i>	51535518	7E-24	M
MDI06-H09	1	<i>ankyrin-kinase, putative</i>	<i>Oryza sativa</i>	57899508	3E-64	M
CL356 contig1	2	<i>ankyrin-like protein</i>	<i>Oryza sativa</i>	56784487	1E-83	M
MDI20_B11	1	<i>ankyrin-like protein</i>	<i>Oryza sativa</i>	56784487	4E-63	M
MDI14_F09	1	<i>Annexin</i>	<i>Gossypium hirsutum</i>	152144	8E-91	U
MDI21_F06	1	<i>annexin-like protein</i>	<i>Medicago sativa</i>	3881978	1E-41	U
CL420 contig1	2	<i>anthocyanidin synthase</i>	<i>Malus x domestica</i>	21038958	4E-64	T
MDI38_H11	1	<i>anthranilate phosphoribosyltransferase-like protein</i>	<i>Arabidopsis thaliana</i>	6729531	9E-16	E
MDI30_B06	1	<i>AP-1 Golgi-related complex component; clathrin coated vesicles; clathrin assembly protein</i>	<i>Camptotheca acuminata</i>	1762309	1E-75	U
MDI18_F10	1	<i>AP2-binding protein, putative</i>	<i>Jatropha curcas</i>	70905563	8E-18	S
MDI37_H11	1	<i>Aquaporin</i>	<i>Ricinus communis</i>	28198150	4E-55	G
MDI01-G05	1	<i>Aquaporin</i>	<i>Vitis vinifera</i>	56608597	3E-81	G
MDI42_C12	1	<i>Aquaporin</i>	<i>Vitis vinifera</i>	56608597	2E-71	G
MDI01-E02	1	<i>Aquaporin</i>	<i>Prunus mune</i>	82568687	1E-33	G
MDI03-F04	1	<i>aquaporin 2</i>	<i>Samanea saman</i>	3158476	1E-76	G
CL67 contig1	4	<i>aquaporin PIP1-2, putative</i>	<i>Vitis berlandieri x Vitis rupestris</i>	8071624	3E-76	G
MDI32-E12	1	<i>aquaporin TIP3, putative</i>	<i>Vitis berlandieri x Vitis rupestris</i>	8699604	5E-11	G

MDI18_B03	1	<i>aquaporin, putative</i>	<i>Phaseolus vulgaris</i>	62945921	6E-54	G
CL398 contig1	2	<i>arabinogalactan-like protein</i>	<i>Pinus taeda</i>	607774	3E-20	S
MDI15_D10	1	<i>ARG10</i>	<i>Vigna radiata</i>	2970051	3E-10	S
MDI34_C10	1	<i>arginine/serine-rich protein, putative</i>	<i>Arabidopsis thaliana</i>	15027957	3E-20	S
MDI20_H09	1	<i>arginine/serine-rich splicing factor, putative</i>	<i>Medicago sativa subsp. X varia</i>	3334756	2E-48	S
MDI25_B10	1	<i>arginyl-tRNA synthetase</i>	<i>Arabidopsis thaliana</i>	2632105	1E-62	J
MDI03-E03	1	<i>arm repeat protein, putative</i>	<i>Oryza sativa</i>	78709013	8E-26	S
CL75 contig1	3	<i>ARPN (Plantacyanin); copper ion binding / electron transporter</i>	<i>Arabidopsis thaliana</i>	15227048	1E-40	S
MDI09-B07	1	<i>ARR17; transcription regulator/ two-component response regulator</i>	<i>Arabidopsis thaliana</i>	18410485	7E-38	K
CL65 contig1	4	<i>Asparaginase</i>	<i>Hordeum vulgare</i>	11066973	4E-60	EC
MDI26_H06	1	<i>asparaginyl-tRNA synthetase, putative</i>	<i>Arabidopsis thaliana</i>	6041857	6E-17	J
MDI33_D05	1	<i>aspartate aminotransferase glyoxysomal isozyme AAT1 precursor</i>	<i>Glycine max</i>	2654094	5E-62	EC
CL368Contig1	2	<i>aspartic proteinase</i>	<i>Theobroma cacao</i>	21616051	2E-57	O
CL105 contig1	3	<i>aspartic proteinase</i>	<i>Theobroma cacao</i>	21616053	9E-79	O
MDI27_D12	1	<i>aspartic proteinase 3</i>	<i>Nepenthes alata</i>	12231176	2E-44	O
MDI29_A10	1	<i>aspartic proteinase 3</i>	<i>Nepenthes alata</i>	12231176	2E-16	O
MDI21_H10	1	<i>aspartic proteinase 3</i>	<i>Nepenthes alata</i>	12231176	2E-44	O
MDI18_G01	1	<i>aspartic proteinase nepenthesin I, putative</i>	<i>Oryza sativa</i>	51535935	3E-19	O
MDI42_B09	1	<i>aspartokinase-homoserine dehydrogenase</i>	<i>Glycine max</i>	2970447	6E-44	E
MDI19_A04	1	<i>ATARFC1; GTP binding</i>	<i>Arabidopsis thaliana</i>	15228912	2E-51	T
MDI45-E02	1	<i>ATG5; phosphoinositide 5-phosphatase</i>	<i>Arabidopsis thaliana</i>	30693541	2E-19	I
MDI15_E02	1	<i>AT-Hook DNA-binding protein, putative</i>	<i>Oryza sativa</i>	37537056	2E-28	S
MDI46-C02	1	<i>ATP binding / ATPase/ nucleoside-triphosphatase/ nucleotide binding</i>	<i>Arabidopsis thaliana</i>	15233037	1E-65	S
MDI47-F10	1	<i>ATP binding / ATPase/ nucleoside-triphosphatase/ nucleotide binding</i>	<i>Arabidopsis thaliana</i>	30678418	1E-24	A
MDI11_D01	1	<i>ATP binding / ATP-dependent helicase/ helicase/ nucleic acid binding</i>	<i>Arabidopsis thaliana</i>	15231074	3E-63	A
MDI36_B06	1	<i>ATP binding / damaged DNA binding</i>	<i>Arabidopsis thaliana</i>	18398289	6E-67	S
MDI11_E01	1	<i>ATP binding / glutamate-tRNA ligase/ glutamine-tRNA ligase/ tRNA ligase</i>	<i>Arabidopsis thaliana</i>	22329783	9E-38	S
CL42 contig1	5	<i>ATP binding / GTP binding / phosphoenolpyruvate carboxykinase (ATP)</i>	<i>Arabidopsis thaliana</i>	15235730	2E-130	S
CL450 contig1	2	<i>ATP binding / hydrolase/ nucleoside-triphosphatase/ nucleotide binding</i>	<i>Arabidopsis thaliana</i>	15231775	2E-152	S
CL428 contig1	2	<i>ATP binding / kinase/ protein kinase/ protein serine/threonine kinase</i>	<i>Arabidopsis thaliana</i>	30687278	3E-50	S
MDI30_A05	1	<i>ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase</i>	<i>Arabidopsis thaliana</i>	18394385	2E-23	S
MDI16_D07	1	<i>ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase</i>	<i>Arabidopsis thaliana</i>	22327668	6E-87	S
MDI25_C04	1	<i>ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase</i>	<i>Arabidopsis thaliana</i>	30689316	3E-28	S
MDI22_C03	1	<i>ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase/ ubiquitin-protein ligase</i>	<i>Arabidopsis thaliana</i>	18408768	5E-10	S
MDI45-C08	1	<i>ATP binding / microtubule motor</i>	<i>Arabidopsis thaliana</i>	22326629	2E-16	S

MDI46-F12	1	ATP binding / nucleotide binding	<i>Arabidopsis thaliana</i>	18423220	1E-68	S
CL112 contig1	3	ATP binding / nucleotide kinase/ phosphotransferase, phosphate group as acceptor / uridylylate kinase	<i>Arabidopsis thaliana</i>	30690243	4E-82	S
MDI32-D10	1	ATP binding / protein binding	<i>Arabidopsis thaliana</i>	79537402	3E-26	S
MDI08-F09	1	ATP binding / protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase	<i>Arabidopsis thaliana</i>	15222211	4E-87	S
MDI27_G04	1	ATP binding / protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase	<i>Arabidopsis thaliana</i>	15225805	2E-56	S
MDI26_E02	1	ATP binding / protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase/ signal transducer	<i>Arabidopsis thaliana</i>	18416060	7E-21	S
MDI25_F10	1	ATP citrate lyase b-subunit	<i>Lupinus albus</i>	15919089	2E-35	GC
MDI17_A04	1	ATP citrate lyase, putative	<i>Arabidopsis thaliana</i>	17065616	2E-99	GC
CL399 contig1	2	ATP synthase beta chain, putative	<i>Oryza sativa</i>	50932681	1E-117	C
MDI48-C02	1	ATP synthase beta chain, putative	<i>Oryza sativa</i>	50932681	3E-26	C
CL162 contig1	3	ATP(GTP)-binding protein, putative	<i>Oryza sativa</i>	50918291	1E-49	G
MDI06-H12	1	ATPase 6	<i>Anastrepha suspensa</i>	15637356	4E-38	L
MDI32-B10	1	ATP-binding ABC transporter protein, probable	<i>Bordetella bronchiseptica RB50</i>	33568028	4E-14	Q
MDI40_D11	1	ATP-dependent Clp protease ATP-binding subunit precursor	<i>Oryza sativa</i>	26518520	5E-92	O
MDI13_G09	1	ATP-dependent peptidase	<i>Arabidopsis thaliana</i>	15222235	8E-30	O
CL265Contig1	2	ATPP2-A13	<i>Arabidopsis thaliana</i>	30695337	3E-34	S
CL237 contig1	2	ATSDAT/ATTDIT; malate transporter/ sodium:dicarboxylate symporter	<i>Arabidopsis thaliana</i>	15238130	6E-83	S
MDI12_B05	1	ATTPS1 (Trehalose-6-phosphate synthase); transferase, transferring glycosyl groups	<i>Arabidopsis thaliana</i>	15218422	8E-64	S
MDI27_G01	1	ATTPS1 (Trehalose-6-phosphate synthase); transferase, transferring glycosyl groups	<i>Arabidopsis thaliana</i>	15218422	3E-28	S
MDI30_A08	1	ATTPS5; transferase, transferring glycosyl groups / trehalose-phosphatase	<i>Arabidopsis thaliana</i>	18414960	3E-43	S
CL222Contig1	2	Auxin Efflux Carrier	<i>Pseudomonas fluorescens PfO-1</i>	77457058	1E-46	S
CL261 contig1	2	auxin growth promotor protein, putative	<i>Lycopersicon esculentum</i>	15144512	1E-40	S
MDI46-B09	1	auxin influx carrier protein, putative	<i>Prunus avium</i>	56562199	2E-59	S
MDI27_C01	1	auxin response factor 4	<i>Cucumis sativus</i>	47496698	1E-41	S
MDI13_B09	1	auxin:hydrogen symporter	<i>Arabidopsis thaliana</i>	15241659	3E-58	S
MDI19_H07	1	auxin:hydrogen symporter	<i>Arabidopsis thaliana</i>	18409758	1E-59	S
MDI36_E11	1	auxin-amidohydrolase precursor, putative	<i>Populus alba</i> x <i>Populus tremula</i>	49524068	3E-35	S
MDI45-F09	1	auxin-binding protein 1	<i>Populus tomentosa</i>	56236456	3E-75	S
MDI33_B06	1	auxin-independent growth promoter - axi1-like protein	<i>Arabidopsis thaliana</i>	10176802	2E-22	S
MDI06-D05	1	auxin-independent growth promoter putative	<i>Oryza sativa</i>	5297229	4E-53	S
MDI16_E09	1	auxin-independent growth promoter, similar	<i>Arabidopsis thaliana</i>	6587842	2E-14	S
CL477 contig1	2	auxin-regulated protein	<i>Zinnia elegans</i>	20257219	8E-68	S
CL154 contig1	3	auxin-repressed protein, putative	<i>Prunus armeniaca</i>	2688824	2E-26	S
MDI26_A12	1	basic blue copper protein	<i>Cicer arietinum</i>	3860333	3E-31	S
CL147 contig1	3	Bax inhibitor-1 like	<i>Arabidopsis thaliana</i>	21593125	5E-63	S

MDI45-H03	1	<i>benzoyl coenzyme A: benzyl alcohol benzoyl transferase</i>	<i>Clarkia breweri</i>	22854876	1E-28	S
CL444 contig1	2	<i>benzoyl-CoA:benzyl alcohol/phenylethanol benzoyltransferase; BPBT</i>	<i>Petunia x hybrida</i>	1556912	2E-77	S
CL294 contig1	2	<i>Bet v I allergen-like</i>	<i>Oryza sativa</i>	50906525	6E-43	S
MDI05-C09	1	<i>beta Galactosidase-like protein</i>	<i>Arabidopsis thaliana</i>	7258347	4E-22	G
MDI05-C06	1	<i>beta transducin-like protein HET-E2C-like</i>	<i>Oryza sativa</i>	57900202	5E-68	T
MDI03-B12	1	<i>beta-1,3-glucanase</i>	<i>Fragaria x ananassa</i>	62362438	3E-67	M
MDI29_A11	1	<i>beta-1,3-glucanase, putative</i>	<i>Arabidopsis thaliana</i>	20465431	2E-44	M
CL401Contig1	2	<i>beta-1,3-glucanase, putative</i>	<i>Oryza sativa</i>	56784589	5E-18	M
MDI11_F10	1	<i>beta-1,3-glucanase, putative</i>	<i>Oryza sativa</i>	50905193	1E-39	M
MDI21_E04	1	<i>beta-6-xylosyltransferase</i>	<i>Vitis vinifera</i>	55956970	1E-118	G
MDI11_E11	1	<i>beta-alanine synthases, putative</i>	<i>Oryza sativa</i>	50937181	6E-78	S
MDI16_E07	1	<i>beta-amylase</i>	<i>Prunus ameniaca</i>	5031285	2E-79	G
CL176 contig1	2	<i>beta-carotene hydroxylase</i>	<i>Citrus unshiu</i>	11245486	3E-89	I
MDI08-H06	1	<i>beta-D-glucan exohydrolase</i>	<i>Nicotiana tabacum</i>	3582436	4E-63	G
CL231 contig1	2	<i>beta-D-glucosidase</i>	<i>Gossypium hirsutum</i>	33391721	2E-156	G
MDI15_H10	1	<i>beta-D-glucosidase</i>	<i>Gossypium hirsutum</i>	33391721	8E-47	G
CL66 contig1	3	<i>beta-galactosidase</i>	<i>Phagemid cloning vector pTZ19U</i>	244016	1E-45	G
MDI32-G09	1	<i>beta-glucosidases, similar</i>	<i>Arabidopsis thaliana</i>	6056418	2E-18	GIQ
MDI38_E08	1	<i>beta-ketoacyl-ACP reductase</i>	<i>Cuphea lanceolata</i>	18045	3E-24	I
MDI10-G04	1	<i>beta-mannosidase enzyme</i>	<i>Lycopersicon esculentum</i>	17226270	3E-68	G
MDI43-D08	1	<i>beta-N-acetylhexosaminidase, putative</i>	<i>Oryza sativa</i>	551683	1E-17	G
MDI24_G03	1	<i>beta-phosphoglucomutase, putative</i>	<i>Arabidopsis thaliana</i>	21593582	3E-56	G
CL455Contig1	2	<i>beta-tubulin</i>	<i>Medicago sativa subsp. falcata</i>	14331109	4E-29	Z
MDI05-G03	1	<i>beta-tubulin</i>	<i>Lolium perenne</i>	59709767	6E-73	Z
MDI30_F09	1	<i>beta-tubulin</i>	<i>Lolium perenne</i>	59709767	2E-65	Z
MDI13_G11	1	<i>beta-tubulin 1</i>	<i>Hordeum vulgare subsp. vulgare</i>	1743277	2E-77	Z
MDI10-H04	1	<i>bHLH transcription factor, putative</i>	<i>Arabidopsis thaliana</i>	21592325	4E-70	K
MDI05-F02	1	<i>Binding</i>	<i>Arabidopsis thaliana</i>	15223949	1E-25	S
MDI11_E07	1	<i>Binding</i>	<i>Arabidopsis thaliana</i>	18396541	3E-65	S
MDI10-G08	1	<i>BONZAI1</i>	<i>Arabidopsis thaliana</i>	15487382	1E-19	S
MDI10-H10	1	<i>brain protein 44-like</i>	<i>Oryza sativa</i>	50899866	7E-38	S
MDI50-C04	1	<i>branched-chain alpha keto-acid dehydrogenase E1 alpha subunit-like protein</i>	<i>Arabidopsis thaliana</i>	9955517	4E-37	C
MDI10-G10	1	<i>branched-chain amino acid aminotransferase -like</i>	<i>Oryza sativa</i>	56784110	4E-47	E
MDI16_B09	1	<i>BRI1-KD interacting protein, putative</i>	<i>Oryza sativa</i>	50906049	2E-24	S
MDI50-A04	1	<i>BRK1</i>	<i>Arabidopsis thaliana</i>	42569246	1E-23	S
MDI19_E11	1	<i>bZIP protein</i>	<i>Arabidopsis thaliana</i>	7576208	2E-66	K

MDI25_D02	1	<i>bZIP protein, putative; 48652-45869</i>	<i>Arabidopsis thaliana</i>	12323124	3E-10	K
MDI52_C01	1	<i>bZIP transcription factor ATB2</i>	<i>Glycine max</i>	22597162	8E-31	K
CL211 contig1	2	<i>bzip-like transcription factor-like</i>	<i>Oryza sativa</i>	34905788	3E-11	K
MDI48-B02	1	<i>C2 domain-containing protein</i>	<i>Cicer arietinum</i>	45720180	1E-47	S
CL193 contig1	2	<i>C2 domain-containing protein</i>	<i>Hordeum vulgare subsp. vulgare</i>	61889374	2E-38	S
MDI21_C06	1	<i>C2 domain-containing protein-like</i>	<i>Oryza sativa</i>	50907613	2E-40	S
MDI51-C07	1	<i>Caffeic acid 3-O-methyltransferase (S-adenosyl-L-methionine:caffeic acid 3-O-methyltransferase)</i>	<i>Eucalyptus gunnii</i>	1169009	1E-49	Q
MDI48-C08	1	<i>caffeic acid O-3-methyltransferase</i>	<i>Populus tomentosa</i>	7528266	1E-27	Q
MDI13_B07	1	<i>caffeic acid O-methyltransferase</i>	<i>Rosa chinensis var. spontanea</i>	32440933	1E-16	Q
MDI13_A06	1	<i>caffeic acid O-methyltransferase</i>	<i>Rosa chinensis var. spontanea</i>	32440935	3E-34	Q
MDI35_G05	1	<i>caffeic acid O-methyltransferase</i>	<i>Rosa chinensis var. spontanea</i>	32440935	7E-23	Q
MDI20_A12	1	<i>caffeic O-methyltransferase</i>	<i>Prunus dulcis</i>	602588	1E-30	Q
MDI11-H09	1	<i>caffeoyl-CoA O-methyltransferase</i>	<i>Eucalyptus gunnii</i>	1934859	4E-91	Q
MDI34_G11	1	<i>calcium ion binding</i>	<i>Arabidopsis thaliana</i>	15241944	8E-24	T
CL348 contig1	2	<i>calmodulin 8</i>	<i>Daucus carota</i>	34304717	5E-74	T
MDI22_C11	1	<i>calmodulin binding</i>	<i>Arabidopsis thaliana</i>	15231175	5E-16	T
MDI11_B11	1	<i>calmodulin binding</i>	<i>Arabidopsis thaliana</i>	18394794	1E-10	T
MDI44-B05	1	<i>calmodulin binding / cyclic nucleotide binding / ion channel - ATCNGC17;</i>	<i>Arabidopsis thaliana</i>	30688654	2E-11	T
MDI13_H02	1	<i>calmodulin binding / triacylglycerol lipase</i>	<i>Arabidopsis thaliana</i>	42572787	6E-20	T
MDI23_C03	1	<i>calmodulin binding / triacylglycerol lipase</i>	<i>Arabidopsis thaliana</i>	42572787	4E-36	T
MDI50-C11	1	<i>calmodulin cam-210</i>	<i>Daucus carota</i>	50299519	2E-38	T
MDI15_F03	1	<i>calmodulin-binding protein, putative</i>	<i>Oryza sativa</i>	50909549	1E-34	T
MDI08-C08	1	<i>calmodulin-like protein</i>	<i>Arabidopsis thaliana</i>	21536925	8E-18	T
MDI09-B08	1	<i>calmodulin-like-domain protein kinase PCPK1, phloem</i>	<i>Curcubita maxima</i>	20453013	3E-88	T
MDI40_D05	1	<i>calnexin-like protein</i>	<i>Lycopersicon esculentum</i>	67077850	4E-88	O
MDI05-B06	1	<i>calnexin-like protein</i>	<i>Lycopersicon esculentum</i>	67077850	2E-36	O
MDI02-G01	1	<i>calreticulin interacted protein</i>	<i>Oryza sativa</i>	29420481	5E-24	O
MDI24_E04	1	<i>carbohydrate transporter/ sugar porter</i>	<i>Arabidopsis thaliana</i>	15219226	1E-33	G
MDI15_C10	1	<i>carbohydrate transporter/ sugar porter</i>	<i>Arabidopsis thaliana</i>	15234491	1E-36	G
MDI02-F04	1	<i>carbonate dehydratase/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	18418245	8E-47	S
MDI52_A05	1	<i>carbonic anhydrase</i>	<i>Populus tremula x Populus tremuloides</i>	1354515	4E-77	C
MDI33_B08	1	<i>carbon-sulfur lyase</i>	<i>Arabidopsis thaliana</i>	15237879	1E-27	S
MDI15_E09	1	<i>carbonyl reductase -like protein</i>	<i>Oryza sativa</i>	50940121	3E-39	I
CL388Contig1	2	<i>carboxylic ester hydrolase</i>	<i>Araabidopsis thaliana</i>	79328183	1E-57	S
MDI06-B09	1	<i>carboxylic ester hydrolase</i>	<i>Arabidopsis thaliana</i>	79328183	2E-69	S
CL422 contig1	2	<i>carboxylic ester hydrolase/ hydrolase, acting on ester bonds</i>	<i>Arabidopsis thaliana</i>	15228189	2E-98	S
MDI12_B07	1	<i>carboxylic ester hydrolase/ hydrolase, acting on ester bonds</i>	<i>Arabidopsis thaliana</i>	15237350	1E-12	S

MDI37_A12	1	carboxyl-terminal proteinase like protein	<i>Arabidopsis thaliana</i>	2245109	7E-33	S
MDI22_B03	1	carboxyl-terminal proteinase, putative	<i>Oryza sativa</i>	50938543	3E-31	S
CL60 contig1	4	carboxymethylenebutenolidase, putative	<i>Arabidopsis thaliana</i>	3298540	1E-88	G
MDI13_H09	1	carboxypeptidase, putative	<i>Arabidopsis thaliana</i>	51971367	3E-43	S
MDI52_D10	1	carnitine/acylcarnitine translocase, putative	<i>Arabidopsis thaliana</i>	21593041	2E-43	I
CL254 contig1	2	casein kinase I, putative	<i>Oryza sativa</i>	34914048	7E-74	TDK
MDI19_B12	1	casein kinase I, putative	<i>Oryza sativa</i>	50915736	5E-47	TDK
MDI12_F01	1	casein kinase, putative	<i>Oryza sativa</i>	55733867	7E-88	TDK
MDI33_D08	1	Catalase	<i>Raphanus sativus</i>	8050693	2E-14	PC
CL13 contig1	9	Catalase	<i>Prunus persica</i>	32526568	0.0	PC
MDI14_E07	1	Catalytic	<i>Arabidopsis thaliana</i>	15229371	1E-67	S
MDI21_E11	1	Catalytic	<i>Arabidopsis thaliana</i>	15242253	3E-43	S
MDI30_E07	1	Catalytic	<i>Arabidopsis thaliana</i>	22329127	6E-37	S
MDI05-F08	1	Catalytic	<i>Arabidopsis thaliana</i>	30689997	5E-79	S
MDI06-B10	1	Catalytic	<i>Arabidopsis thaliana</i>	30683676	1E-56	S
CL418 contig1	2	Catalytic	<i>Arabidopsis thaliana</i>	42569479	7E-59	S
MDI13_B06	1	catalytic/ hydrolase	<i>Arabidopsis thaliana</i>	30681816	1E-48	S
MDI15_G06	1	catalytic/ hydroxyethylthiazole kinase	<i>Arabidopsis thaliana</i>	15229565	4E-25	S
MDI27_G07	1	catalytic/ long-chain-fatty-acid-CoA ligase	<i>Arabidopsis thaliana</i>	30678112	5E-60	TI
MDI34_A05	1	catalytic/ mannosyltransferase	<i>Arabidopsis thaliana</i>	79324944	6E-18	G
MDI39_C01	1	catalytic/ phosphogluconate dehydrogenase (decarboxylating)	<i>Arabidopsis thaliana</i>	15233455	8E-33	G
MDI14_C12	1	catalytic/ transferase, transferring glycosyl groups	<i>Arabidopsis thaliana</i>	42565357	1E-16	S
MDI10-C02	1	catalytic/ tryptophan synthase	<i>Arabidopsis thaliana</i>	42573494	2E-60	E
MDI14_A11	1	catechol O-methyltransferase; Omt II;THATU;1	<i>Thalictrum tuberosum</i>	4808522	1E-20	E
MDI41_B10	1	cation diffusion facilitator 10	<i>Stylosanthes hamata</i>	28628306	3E-15	P
MDI12_C05	1	cation diffusion facilitator 9	<i>Stylosanthes hamata</i>	28628304	4E-62	P
MDI24_B05	1	CBF1 interacting corepressor CIR, putative	<i>Oryza sativa</i>	53792619	2E-17	T
MDI06-A10	1	CBL-interacting protein kinase	<i>Brassica napus</i>	17224924	2E-65	T
MDI50-G06	1	CBL-interacting protein kinase 21 - CIPK21; ATP binding / protein kinase/ protein serine/threonine/tyrosine kinase	<i>Arabidopsis thaliana</i>	18424001	1E-35	T
CL140 contig1	3	CBL-Interacting protein kinase 23 - CIPK23; ATP binding /protein kinase/ protein serine/threonine/tyrosine kinase	<i>Arabidopsis thaliana</i>	18397430	2E-143	T
MDI11_C07	1	CBL-Interacting protein kinase 6 - CIPK6 ; ATP binding /protein serine/threonine kinase/ protein-tyrosine kinase	<i>Arabidopsis thaliana</i>	15235768	2E-16	T
CL381 contig1	2	CBS domain protein-like	<i>Oryza sativa</i>	55296028	1E-35	S
MDI05-D06	1	cell cycle control crn (crooked neck) protein-like	<i>Arabidopsis thaliana</i>	25083215	4E-90	S
MDI21_E10	1	cell wall apoplastic invertase	<i>Vitis vinifera</i>	47078691	9E-58	S
CL386 contig1	2	cellulose synthase	<i>Populus tremuloides</i>	39933010	3E-105	G

MDI02-H11	1	cellulose synthase 6	<i>Populus tremuloides</i>	37781495	4E-84	G
MDI07-D02	1	cellulose synthase catalytic subunit	<i>Gossypium hirsutum</i>	6446577	4E-89	G
MDI05-C08	1	cellulose synthase catalytic subunit-like protein	<i>Arabidopsis thaliana</i>	15237873	1E-75	G
MDI11_G07	1	cellulose synthase Cesa1	<i>Boehmeria nivea</i>	68566323	1E-58	G
MDI15_A10	1	cellulose synthase-like protein CslG	<i>Nicotiana tabacum</i>	73624747	1E-89	G
CL98 contig1	3	chalcone synthase	<i>Dictamnus albus</i>	54311699	2E-100	Q
MDI43_D01	1	Chalcone synthase 2 (Naringenin-chalcone synthase 2)	<i>Camellia sinensis</i>	1345787	2E-85	Q
CL156 contig1	3	chaperonin 21, chloroplast	<i>Vitis vinifera</i>	50660327	2E-98	O
MDI27_B09	1	chaperonin 60 alpha subunit	<i>Canavalia lineata</i>	3790441	3E-28	O
CL247 contig1	2	chaperonin 60 alpha subunit	<i>Canavalia lineata</i>	3790441	1E-56	O
MDI01-F04	1	chaperonin 60 beta precursor, putative	<i>Oryza sativa</i>	34897924	5E-63	O
MDI10-B07	1	chaperonin subunit, putative	<i>Arabidopsis thaliana</i>	21595439	5E-84	O
MDI09-C06	1	chemotaxis sensory transducer, Cache sensor	<i>Pseudomonas fluorescens PfO-1</i>	77460652	1E-50	S
MDI20_E03	1	Chitinase	<i>Beta vulgaris subsp. vulgaris</i>	829258	8E-10	S
CL101 contig1	3	Chitinase	<i>Vitis vinifera</i>	10880381	8E-88	S
MDI07-E06	1	Chitinase	<i>Chenopodium amaranticolor</i>	2570164	4E-39	S
MDI33_E07	1	Chitinase 1 precursor (Tulip bulb chitinase-1) (TBC-1)	<i>Tulipa bakeri</i>	47605559	9E-42	S
CL169 contig1	3	chitinase-like protein	<i>Gossypium hirsutum</i>	34016877	2E-109	S
MDI52_C09	1	CHL6/CNX/CNX1/SIR4; molybdenum ion binding	<i>Arabidopsis thaliana</i>	15242124	4E-78	S
CL393 contig1	2	chlorophyll a /b binding protein	<i>Beta vulgaris</i>	33438474	6E-90	S
CL476 contig1	2	chlorophyll a/b binding protein CP29	<i>Vigna radiata</i>	4689382	2E-73	S
MDI11_A02	1	chlorophyll a/b binding protein, putative	<i>Arabidopsis thaliana</i>	21594018	1E-74	S
MDI11_G11	1	chlorophyll a/b-binding protein	<i>Pisum sativum</i>	20671	2E-73	S
MDI08-E05	1	chlorophyll a/b-binding protein	<i>Lycopersicon esculentum</i>	19180	1E-86	S
MDI17_C08	1	chlorophyll a/b-binding protein type III precursor, putative	<i>Oryza sativa</i>	50905979	2E-48	S
CL454 contig1	2	chlorophyll a/b-binding protein; LHCP II	<i>Pinus thunbergii</i>	688315	1E-97	S
CL468 contig1	2	chlorophyll a/b-binding protein-like	<i>Arabidopsis thaliana</i>	21593520	2E-87	S
CL72 contig1	3	Chlorophyll a-b binding protein of LHCP type I, chloroplast precursor (CAB) (LHCP)	<i>Cucumis sativus</i>	115784	3E-95	S
MDI03-A10	1	chlorophyll antenna size regulatory protein	<i>Chlamydomonas reinhardtii</i>	22536150	2E-22	S
MDI33_B01	1	chloroplast ATP/ADP transporter protein	<i>Helianthus tuberosus</i>	51243142	2E-73	S
CL258 contig1	2	chloroplast FtsH protease, putative	<i>Oryza sativa</i>	54291028	2E-138	S
MDI20_D05	1	chloroplast latex aldolase-like protein	<i>Manihot esculenta</i>	56122688	4E-80	S
MDI17_A07	1	chloroplast mRNA-binding protein CSP41 precursor	<i>Spinacia oleracea</i>	1532135	2E-13	S
CL382 contig1	2	chloroplast nucleoid DNA-binding protein, putative	<i>Arabidopsis thaliana</i>	19424106	5E-73	S
MDI50-F03	1	chloroplast oxygen-evolving enhancer protein	<i>Manihot esculenta</i>	56122682	4E-29	S

MDI43_B10	1	<i>chloroplast protease</i>	<i>Capsicum annuum</i>	3808101	1E-35	S
MDI29_A05	1	<i>chloroplast stromal ascorbate peroxidase</i>	<i>Vigna unguiculata</i>	45268439	8E-14	S
MDI27_H10	1	<i>chorismate mutase/prephenate dehydratase, putative</i>	<i>Arabidopsis thaliana</i>	3860250	3E-55	E
CL396Contig1	2	<i>chromosome I BAC T6A9 genomic sequence</i>	<i>Arabidopsis thaliana</i>	7958959	5E-20	S
MDI08-G01	1	<i>CID9; RNA binding / protein binding</i>	<i>Arabidopsis thaliana</i>	15231858	1E-19	S
MDI18_B09	1	<i>CIG2</i>	<i>Nicotiana tabacum</i>	15216226	4E-23	S
MDI32-G02	1	<i>cinnamate 4-hydroxylase</i>	<i>Populus balsamifera subsp. trichocarpa</i> <i>x P. deltoids</i>	12276037	3E-27	E
MDI24_C02	1	<i>cinnamoyl CoA reductase</i>	<i>Eucalyptus globules</i>	50345916	2E-51	V
CL106 contig1	3	<i>cinnamoyl-CoA reductase, putative</i>	<i>Solanum demissum</i>	47900734	1E-128	V
MDI39_C05	1	<i>cinnamoyl-CoA reductase-like protein</i>	<i>Arabidopsis thaliana</i>	21553548	2E-65	V
CL333 contig1	2	<i>cinnamyl alcohol dehydrogenase</i>	<i>Aralia cordata</i>	474300	2E-76	Q
MDI07-B07	1	<i>cinnamyl alcohol dehydrogenase</i>	<i>Eucalyptus gunnii</i>	1143445	8E-85	Q
MDI20_F09	1	<i>cinnamyl alcohol dehydrogenase</i>	<i>Eucalyptus gunnii</i>	1143445	1E-95	Q
MDI24_C03	1	<i>Citrate synthase, glyoxysomal precursor (GCS)</i>	<i>Curcubita maxima</i>	1345933	1E-79	S
MDI52_H05	1	<i>class II small heat shock protein Le-HSP17.6</i>	<i>Lycopersicon esculentum</i>	1773291	3E-43	O
MDI08-D05	1	<i>class IV chitinase</i>	<i>Vitis vinifera</i>	33413754	2E-69	GM
MDI09-C01	1	<i>class IV chitinase</i>	<i>Arabidopsis thaliana</i>	2597826	2E-61	GM
MDI04-A04	1	<i>clathrin assembly protein AP180 short form-like</i>	<i>Oryza sativa</i>	56785152	2E-36	TU
MDI30_G09	1	<i>clathrin binding</i>	<i>Arabidopsis thaliana</i>	15221714	2E-34	TU
MDI01-E09	1	<i>Clathrin coat assembly protein AP50, putative</i>	<i>Oryza sativa</i>	50911657	2E-48	TU
MDI18_E12	1	<i>clathrin heavy chain, putative</i>	<i>Oryza sativa</i>	77552802	2E-82	TU
MDI08-F08	1	<i>clathrin-coat assembly protein-like</i>	<i>Oryza sativa</i>	50942829	1E-28	TU
MDI26_E01	1	<i>cleavage and polyadenylation specific factor, putative</i>	<i>Oryza sativa</i>	50945445	4E-11	A
CL310 contig1	2	<i>CLPR1; endopeptidase Clp</i>	<i>Arabidopsis thaliana</i>	18402863	9E-91	S
MDI12_C06	1	<i>CMP-sialic acid transporter, putative</i>	<i>Oryza sativa</i>	50938547	1E-47	S
MDI44-A08	1	<i>c-myc binding protein, putative</i>	<i>Oryza sativa</i>	31432776	4E-34	S
CL405 contig1	2	<i>coatomer alpha subunit-like protein</i>	<i>Lotus corniculatus var. japonicus</i>	33945873	7E-20	P
CL227Contig1	2	<i>coatomer epsilon subunit, putative</i>	<i>Arabidopsis thaliana</i>	3033380	5E-34	P
MDI27_F04	1	<i>cobalamine-independent methionine synthase</i>	<i>Solenostemon scutellarioides</i>	974782	2E-48	H
MDI01-F10	1	<i>COP1 homolog (Constitutive Photomorphogenic 1)</i>	<i>Lycopersicon esculentum</i>	4090943	5E-43	OU
MDI41_F09	1	<i>COP1-interacting protein 7 (CIP7)-like</i>	<i>Oryza sativa</i>	50939273	3E-13	OU
MDI49-C05	1	<i>COP9 signalosome complex subunit 2, putative</i>	<i>Oryza sativa</i>	56783671	8E-94	TOU
MDI06-F08	1	<i>COP9 signalosome subunit 6</i>	<i>Arabidopsis thaliana</i>	17940314	5E-78	TOU
CL238 contig1	2	<i>copine III-like</i>	<i>Oryza sativa</i>	56785364	8E-68	T
MDI33_D06	1	<i>copper ion binding / electron transporter</i>	<i>Arabidopsis thaliana</i>	15242279	8E-41	P
CL402 contig1	2	<i>COPT5-like protein</i>	<i>Oryza sativa</i>	51091411	4E-25	S

MDI25_B09	1	<i>coronatine-insensitive 1</i>	<i>Glycine max</i>	72154229	5E-19	S
MDI05-C04	1	<i>COX VIIa-like protein, putative</i>	<i>Pinguicula sp. Jobson 240</i>	47606405	2E-21	S
MDI34_F01	1	<i>CTD phosphatase-like 3</i>	<i>Arabidopsis thaliana</i>	22212705	4E-16	K
MDI25_C08	1	<i>C-terminal zinc-finger</i>	<i>Glycine max</i>	558543	3E-23	S
MDI17_H02	1	<i>CTV.2, putative</i>	<i>Oryza sativa</i>	56783655	3E-67	S
MDI13_E10	1	<i>CTV.22</i>	<i>Poncirus trifoliata</i>	24461867	4E-17	S
MDI30_H10	1	<i>Cu2+-transporting ATPase-like protein</i>	<i>Arabidopsis thaliana</i>	4006855	1E-23	P
MDI19_A06	1	<i>CUL1</i>	<i>Oryza sativa</i>	54290813	5E-71	S
MDI24_H01	1	<i>cullin protein, putative</i>	<i>Oryza sativa</i>	50355720	1E-68	D
MDI08-C02	1	<i>cyclase-like protein</i>	<i>Oryza sativa</i>	50898414	3E-54	I
CL295 contig1	2	<i>cyclin-dependent kinases regulatory subunit</i>	<i>Glycine max</i>	42362268	4E-31	UD
MDI20_F05	1	<i>cyclin-dependent protein kinase</i>	<i>Arabidopsis thaliana</i>	15233185	6E-45	UD
MDI21_F02	1	<i>cycloartenol synthase</i>	<i>Betula platyphylla</i>	18147592	1E-33	I
CL103 contig1	3	<i>Cyclophilin</i>	<i>Ricinus communis</i>	18076088	3E-74	O
CL217 contig1	2	<i>Cyclophilin</i>	<i>Ricinus communis</i>	18076088	1E-61	O
MDI27_E05	1	<i>cyclophilin (70.8 kD) (cyp-15), putative</i>	<i>Oryza sativa</i>	50948445	1E-55	O
MDI26_H05	1	<i>CYP716A2 heme binding / iron ion binding / monooxygenase/ oxygen binding</i>	<i>Arabidopsis thaliana</i>	15239310	2E-10	S
MDI30_E06	1	<i>cystathionine gamma-synthase isoform 2</i>	<i>Solanum tuberosum</i>	8439543	2E-52	OCE
MDI30_B01	1	<i>cysteine protease</i>	<i>Prunus armeniaca</i>	2677828	3E-62	O
MDI05-A12	1	<i>cysteine protease</i>	<i>Ipomoea batatas</i>	13491752	6E-47	O
MDI04-F03	1	<i>cysteine protease</i>	<i>Gossypium hirsutum</i>	27728675	1E-76	O
MDI25_A04	1	<i>cysteine protease</i>	<i>Gossypium hirsutum</i>	27728675	1E-35	O
CL157 contig1	3	<i>cysteine protease CP1</i>	<i>Manihot esculenta</i>	62526575	2E-70	O
CL37 contig1	5	<i>cysteine protease inhibitor</i>	<i>Populus tremula</i>	52851096	1E-36	O
MDI50-F10	1	<i>cysteine protease, putative</i>	<i>Pisum sativum</i>	6624721	1E-10	O
CL35 contig1	5	<i>cysteine proteinase precursor</i>	<i>Phaseolus vulgaris</i>	2511689	8E-63	O
MDI51-E11	1	<i>cysteine proteinase precursor</i>	<i>Phaseolus vulgaris</i>	2511689	2E-67	O
CL148Contig1	3	<i>cysteine proteinase-like protein</i>	<i>Ipomoea batatas</i>	56961686	2E-47	O
CL370Contig1	2	<i>cysteine synthase</i>	<i>Citrullus lanatus</i>	540497	2E-92	C
MDI30_H08	1	<i>cysteine-type endopeptidase/ cysteine-type peptidase</i>	<i>Arabidopsis thaliana</i>	18378947	6E-32	S
MDI31_A09	1	<i>cysteine-type endopeptidase/ ubiquitin thiolesterase</i>	<i>Arabidopsis thaliana</i>	15235599	4E-11	S
CL286Contig1	2	<i>cytochrome b5</i>	<i>Olea europaea</i>	2695711	4E-55	G
MDI36_F11	1	<i>cytochrome b</i>	<i>Vitis vinifera</i>	52082696	2E-18	G
CL207Contig1	2	<i>cytochrome c oxidase subunit 5c-like protein</i>	<i>Arabidopsis thaliana</i>	7340711	5E-24	C
MDI18_A01	1	<i>cytochrome P450</i>	<i>Nicotiana tabacum</i>	1237250	7E-33	Q
MDI06-C06	1	<i>cytochrome P450</i>	<i>Ammi majus</i>	46370700	4E-34	Q
MDI52_H08	1	<i>cytochrome P450</i>	<i>Ammi majus</i>	46370700	1E-51	Q

MDI18_A05	1	cytochrome P450	<i>Ammi majus</i>	46370702	1E-18	Q
MDI30_F01	1	Cytochrome P450 77A3	<i>Glycine max</i>	5915821	9E-29	Q
MDI13_D09	1	cytochrome P450 CYP1A	<i>Picea sitchensis</i>	59800274	6E-16	Q
MDI51-B07	1	cytochrome P450 CYP2A	<i>Picea sitchensis</i>	59800276	9E-27	Q
MDI02-F12	1	cytochrome P450 CYP2A	<i>Picea sitchensis</i>	59800276	3E-35	Q
MDI06-E02	1	cytochrome P450 CYP2A	<i>Picea sitchensis</i>	59800276	1E-48	Q
MDI42_G07	1	cytochrome P450 monooxygenase CYP716A	<i>Medicago truncatula</i>	84514135	8E-16	Q
CL39 contig1	5	cytochrome P450 monooxygenase CYP72A5	<i>Zea mays</i>	21805645	2E-61	Q
MDI27_H11	1	cytochrome p450, putative	<i>Arabidopsis thaliana</i>	20198022	5E-38	Q
MDI32-C10	1	cytochrome P450, putative	<i>Oryza sativa</i>	34912916	9E-19	Q
MDI22_F07	1	cytochrome P450-like	<i>Arabidopsis thaliana</i>	8777295	6E-24	Q
MDI04-D01	1	cytochrome P450-like	<i>Arabidopsis thaliana</i>	8777295	3E-30	Q
MDI23_E01	1	cytoplasmic Cu/ZnSOD	<i>Gossypium hirsutum</i>	74229677	4E-64	P
MDI42_A04	1	cytoplasmic ribosomal protein L18	<i>Arabidopsis thaliana</i>	606970	6E-53	J
CL391 contig1	2	cytoplasmic small heat shock protein class I	<i>Capsicum frutescens</i>	40646968	1E-42	O
CL210 contig1	2	cytosolic ascorbate peroxidase	<i>Fragaria x ananassa</i>	2754860	9E-80	G
CL118 contig1	3	cytosolic ascorbate peroxidase	<i>Fragaria x ananassa</i>	2754860	1E-77	G
MDI04-H10	1	Heat shock protein class II small Le-HSP17.6	<i>Lycopersicon esculentum</i>	1773291	1E-19	O
MDI10-F10	1	Heat shock protein 17.7 kDa (hsp17.7).	<i>Pisum sativum</i>	169099	3E-34	O
MDI11_A07	1	Heat shock protein 17.7 kDa	<i>Carica papaya</i>	37933812	2E-29	O
MDI26_A11	1	Heat shock protein 17.7 kDa	<i>Carica papaya</i>	37933812	2E-55	O
MDI27_C11	1	Heat shock protein 17.7 kDa	<i>Carica papaya</i>	37933812	4E-57	O
MDI39_H02	1	Heat shock protein Le-HSP17.6 class II small	<i>Lycopersicon esculentum</i>	1773291	8E-64	O
MDI29_H02	1	Heat shock protein 17.7 kDa	<i>Carica papaya</i>	37933812	1E-48	O
MDI42_H06	1	cytosolic factor, putative	<i>Trifolium pratense</i>	84453207	7E-19	S
MDI07-A05	1	DAG protein	<i>Antirrhinum majus</i>	1200205	1E-10	S
MDI44-B02	1	DAL1 protein	<i>Arabidopsis thaliana</i>	2440029	7E-16	S
MDI07-E09	1	D-alanyl-D-alanine endopeptidase	<i>Arabidopsis thaliana</i>	18395487	6E-26	S
MDI18_B12	1	defective in exine formation	<i>Arabidopsis thaliana</i>	11138671	1E-18	V
MDI21_A08	1	defense-related F-box protein	<i>Oryza sativa</i>	78498846	5E-59	OTV
MDI12_D10	1	dehydration responsive element-binding protein 3	<i>Glycine max</i>	69111452	7E-22	S
MDI26_B04	1	dehydration-responsive protein-like	<i>Oryza sativa</i>	55296764	5E-29	S
MDI01-F03	1	delta7 sterol C-5 desaturase	<i>Zinnia elegans</i>	71834080	5E-52	I
MDI15_H07	1	delta-pyrroline-5-carboxylate synthetase	<i>Glycine max</i>	40457265	7E-50	I
MDI03-B02	1	DEM2	<i>Lycopersicon esculentum</i>	37223344	2E-58	S
MDI22_F01	1	deoxycytidine deaminase, putative	<i>Oryza sativa</i>	50919917	4E-25	F
CL127 contig1	3	dessication-related protein	<i>Craterostigma plantagineum</i>	167479	7E-44	S

MDI02-G12	1	<i>dessication-related protein, putative</i>	<i>Arabidopsis thaliana</i>	12323093	5E-19	S
MDI48-G05	1	<i>developmental protein, putative</i>	<i>Nicotiana benthamiana</i>	28916459	1E-13	S
MDI50-H11	1	<i>diacylglycerol O-acyltransferase</i>	<i>Arabidopsis thaliana</i>	18409359	2E-62	I
MDI25_E05	1	<i>diadenosine 5',5'''-P1,P4-tetraphosphate hydrolase</i>	<i>Lupinus angustifolius</i>	1888557	2E-45	S
MDI13_D11	1	<i>dicarboxylate carrier protein mitochondrial, putative</i>	<i>Arabidopsis thaliana</i>	21554157	1E-45	S
MDI50-A10	1	<i>dicarboxylate carrier protein, mitochondrial putative</i>	<i>Arabidopsis thaliana</i>	13878155	1E-25	S
MDI04-C08	1	<i>dicarboxylate diiron protein, putative</i>	<i>Arabidopsis thaliana</i>	7542486	1E-99	S
MDI35_H03	1	<i>dihydroflavonol reductase, putative</i>	<i>Ipomoea trifida</i>	45935133	9E-63	Q
MDI22_H08	1	<i>dihydrolipoamide dehydrogenase precursor</i>	<i>Lycopersicon esculentum</i>	23311340	4E-65	G
MDI27_D02	1	<i>dihydrolipoamide dehydrogenase, putative</i>	<i>Oryza sativa</i>	15290186	4E-16	G
MDI32-B06	1	<i>dihydrolipoamide succinyltransferase, putative</i>	<i>Arabidopsis thaliana</i>	7269544	1E-11	EC
MDI01-H01	1	<i>Dihydroxyacid dehydratase/phosphogluconate dehydratase</i>	<i>Burkholderia fungorum LB400</i>	48780964	1E-116	E
MDI51-C10	1	<i>Dioxygenase</i>	<i>Marah macrocarpus</i>	1666096	2E-51	S
MDI47-E09	1	<i>dioxygenase, putative</i>	<i>Arabidopsis thaliana</i>	3201612	1E-45	S
MDI17_D02	1	<i>dipeptidyl-peptidase IV/ serine-type peptidase</i>	<i>Arabidopsis thaliana</i>	15237923	2E-66	S
MDI20_E11	1	<i>disease resistance-responsive family protein</i>	<i>Arachis hypogaea</i>	71040679	8E-52	T
MDI25_B12	1	<i>disulfide oxidoreductase/ monooxygenase/ oxidoreductase</i>	<i>Arabidopsis thaliana</i>	15221995	5E-45	S
MDI19_D03	1	<i>Disulphide bond formation protein DsbB</i>	<i>Pseudomonas fluorescens Pfo-1</i>	77460871	3E-11	S
MDI46-H06	1	<i>DNA binding</i>	<i>Arabidopsis thaliana</i>	152418	7E-16	S
CL262 contig1	2	<i>DNA binding</i>	<i>Arabidopsis thaliana</i>	15220725	2E-43	S
MDI03-F03	1	<i>DNA binding / protein binding / zinc ion binding</i>	<i>Arabidopsis thaliana</i>	18416854	2E-33	S
MDI01-B11	1	<i>DNA binding / protein binding / zinc ion binding</i>	<i>Arabidopsis thaliana</i>	30680213	2E-12	S
MDI02-H10	1	<i>DNA binding protein, putative</i>	<i>Arabidopsis thaliana</i>	7269553	5E-34	S
MDI40_F07	1	<i>DNA binding zinc finger protein (Pspzf)</i>	<i>Pisum sativum</i>	4240031	2E-27	K
MDI06-G08	1	<i>DNA helicase/ nucleoside-triphosphatase/ nucleotide binding</i>	<i>Arabidopsis thaliana</i>	15240788	2E-84	L
MDI03-E08	1	<i>DNA packaging protein of prophage CP-933X, putative</i>	<i>putative DNA protein of prophage CP-933X</i>	15801357	1E-18	S
MDI27_H01	1	<i>DNA polymerase delta p66 subunit -like</i>	<i>Oryza sativa</i>	56201487	1E-27	L
CL260 contig1	2	<i>DNA-directed RNA polymerase II 13.6K chain</i>	<i>Arabidopsis thaliana</i>	4678938	6E-49	S
CL252 contig1	2	<i>DNAJ heat shock N-terminal domain-containing protein, putative</i>	<i>Oryza sativa</i>	50905929	7E-60	O
CL185Contig1	2	<i>DnaJ homolog</i>	<i>Salix gilgiana</i>	4008159	8E-47	O
MDI14_D07	1	<i>DnaJ protein-like</i>	<i>Oryza sativa</i>	55775118	7E-14	O
MDI24_A07	1	<i>DnaJ-like protein</i>	<i>Phaseolus vulgaris</i>	1684851	9E-21	O
MDI01-C07	1	<i>DnaJ-like protein</i>	<i>Glycine max</i>	23428796	2E-50	O
CL171 contig1	3	<i>dnaK-type molecular chaperone hsp70 - rice</i>	<i>Oryza sativa</i>	62733235	2E-161	O
MDI24_G04	1	<i>dolichyl-diphosphooligosaccharide-protein glycotransferase/ oligosaccharyl transferase</i>	<i>Arabidopsis thaliana</i>	15226341	4E-48	G

MDI26_B08	1	<i>dolichyl-diphosphooligosaccharide-protein glycotransferase/ oligosaccharyl transferase</i>	<i>Arabidopsis thaliana</i>	15226341	3E-45	G
MDI49-H07	1	<i>dolichyl-phosphate beta-D-mannosyltransferase/ eukaryotic elongation factor-2 kinase</i>	<i>Arabidopsis thaliana</i>	18394881	2E-39	O
MDI16_G02	1	<i>dopamine beta-monoxygenase/ monoxygenase</i>	<i>Arabidopsis thaliana</i>	15239759	4E-27	I
MDI12_C12	1	<i>DRT111; nucleic acid binding</i>	<i>Arabidopsis thaliana</i>	15220757	8E-47	S
MDI42_F04	1	<i>dsRNA-binding protein ODB1, putative</i>	<i>Oryza sativa</i>	56785214	4E-13	YF
MDI27_C10	1	<i>dTDP-D-glucose 4,6-dehydratase-like</i>	<i>Oryza sativa</i>	50253123	2E-27	GM
MDI13_G10	1	<i>dymeclin, putative</i>	<i>Oryza sativa</i>	56785226	1E-64	S
MDI19_F11	1	<i>Dynamin-Like Protein - ADL1; GTP binding</i>	<i>Arabidopsis thaliana</i>	30693985	3E-73	S
MDI04-E08	1	<i>early light induced protein, putative</i>	<i>Arachis hypogaea</i>	28194659	5E-29	S
MDI37_F09	1	<i>Early light induced protein, putative</i>	<i>Arachis hypogaea</i>	28194659	1E-21	S
MDI40_A01	1	<i>early-responsive to dehydration stress protein, putative</i>	<i>Oryza sativa</i>	50912907	1E-24	S
MDI32-H04	1	<i>EBF2 (EIN3-Binding F Box Protein 2)</i>	<i>Arabidopsis thaliana</i>	15239399	6E-17	S
MDI41_G01	1	<i>EF-1-alpha-related GTP-binding protein</i>	<i>Nicotiana tabacum</i>	1009232	7E-35	J
MDI20_H04	1	<i>EF-hand Ca2+-binding protein CCD1</i>	<i>Triticum aestivum</i>	9255753	9E-26	JZ
MDI01-B09	1	<i>EIF3K</i>	<i>Arabidopsis thaliana</i>	15234123	5E-27	J
MDI11_A05	1	<i>EIN3-like protein</i>	<i>Cucumis melo</i>	15425733	2E-51	S
MDI13_E06	1	<i>electron transporter/ thiol-disulfide exchange intermediate</i>	<i>Arabidopsis thaliana</i>	15218686	6E-12	S
CL77 contig1	4	<i>ELIP</i>	<i>Brassica rapa subsp. Pekinensis</i>	38123375	6E-46	S
MDI14_G04	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	2E-34	S
MDI20_C03	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	1E-44	S
MDI35_F06	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	5E-45	S
MDI38_A11	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	2E-45	S
MDI47-G10	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	2E-43	S
MDI48-F11	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	9E-39	S
MDI49-G06	1	<i>ELIP</i>	<i>Brassica rapa subsp. pekinensis</i>	38123375	2E-42	S
MDI31_E08	1	<i>ELIP1 (Early Light-Inducable Protein); chlorophyll binding</i>	<i>Arabidopsis thaliana</i>	15228866	9E-12	S
MDI13_G05	1	<i>elongation factor</i>	<i>Triticum aestivum</i>	32400835	4E-27	J
MDI02-F07	1	<i>elongation factor 1-alpha 1</i>	<i>Elaeis guineensis</i>	48527431	1E-89	J
MDI38_C11	1	<i>elongation factor 2</i>	<i>Aspergillus nidulans FGSC A4</i>	67540320	8E-20	J
MDI36_E08	1	<i>Elongation factor Tu, chloroplast precursor (EF-Tu)</i>	<i>Glycine max</i>	949873	1E-22	J
MDI29_E11	1	<i>EMB1374; transcription regulator</i>	<i>Arabidopsis thaliana</i>	15236742	8E-42	K
MDI05-H08	1	<i>endo-1,3;1,4-beta-D-glucanase, putative</i>	<i>Oryza sativa</i>	51854423	1E-54	GM
MDI41_H11	1	<i>endo-1,3-1,4-beta-D-glucanase, similarity</i>	<i>Arabidopsis thaliana</i>	21536848	6E-10	GM
MDI42_G01	1	<i>endo-1,4-beta-xylanase, putative</i>	<i>Populus tremula x P. tremuloides</i>	60656567	5E-40	S
CL126 contig1	3	<i>endo-beta-1,4-glucanase</i>	<i>Fragaria x ananassa</i>	22208353	4E-149	M
MDI19_D12	1	<i>endo-beta-1,4-glucanase</i>	<i>Fragaria x ananassa</i>	22208353	1E-38	M

MDI24_C12	1	<i>endomembrane protein 70, putative</i>	<i>Arabidopsis thaliana</i>	27754304	8E-74	U
MDI03-D09	1	<i>endomembrane protein 70-like</i>	<i>Oryza sativa</i>	51534986	2E-18	U
MDI26_E11	1	<i>endopeptidase Clp</i>	<i>Arabidopsis thaliana</i>	18414804	2E-49	O
MDI39_E11	1	<i>endoplasmatic reticulum retrieval protein Rer1B [validated]</i>	<i>Arabidopsis thaliana</i>	11281669	1E-16	S
MDI18_H03	1	<i>endoplasmic reticulum oxidoreductin, putative</i>	<i>Oryza sativa</i>	50582733	3E-21	U
MDI20_E07	1	<i>endosomal protein-like</i>	<i>Arabidopsis thaliana</i>	9758717	8E-73	U
CL269 contig1	2	<i>Enolase</i>	<i>Gossypium barbadense</i>	33415263	1E-126	G
MDI25_E09	1	<i>enzyme inhibitor/ pectinesterase/ pectinesterase inhibitor</i>	<i>Arabidopsis thaliana</i>	30697666	3E-29	G
MDI27_F03	1	<i>ER6 protein, putative</i>	<i>Oryza sativa</i>	52077479	3E-32	S
CL189 contig1	2	<i>ERD4 protein (early-responsive dehydration protein)</i>	<i>Arabidopsis thaliana</i>	62321290	8E-24	S
MDI29_D03	1	<i>ERD4 protein, putative</i>	<i>Oryza sativa</i>	50934237	4E-42	S
MDI02-E03	1	<i>ethylene overproducer-like 1</i>	<i>Lycopersicon esculentum</i>	70780055	2E-58	S
MDI19_B04	1	<i>ethylene overproducer-like 1</i>	<i>Lycopersicon esculentum</i>	70780055	5E-64	S
CL159 contig1	3	<i>ethylene responsive element binding protein</i>	<i>Fagus sylvatica</i>	18496063	1E-47	S
MDI06-F02	1	<i>ethylene signaling protein</i>	<i>Lycopersicon esculentum</i>	77747170	5E-17	S
CL100 contig1	3	<i>ethylene-forming-enzyme-like dioxygenase-like protein</i>	<i>Oryza sativa</i>	50933567	2E-48	S
CL379Contig1	2	<i>ethylene-responsive element binding protein 5</i>	<i>Nicotiana tabacum</i>	55585713	1E-20	S
CL115 contig1	3	<i>ethylene-responsive element binding protein ERF2</i>	<i>Gossypium hirsutum</i>	59800341	9E-43	S
MDI35_D02	1	<i>ethylene-responsive small GTP-binding protein</i>	<i>Lycopersicon esculentum</i>	5669640	2E-79	S
MDI34_A12	1	<i>ethylene-responsive small GTP-binding protein, putative</i>	<i>Oryza sativa</i>	50935375	1E-56	S
MDI24_E07	1	<i>Ethylene-responsive transcription factor 3 (Ethylene-responsive element-binding factor 3 homolog)</i>	<i>Nicotiana tabacum</i>	57012880	3E-14	K
MDI12_F06	1	<i>eukaryotic initiation factor iso4E</i>	<i>Nicotiana tabacum</i>	51557995	2E-24	J
CL451 contig1	2	<i>eukaryotic translation initiation factor 5A isoform II</i>	<i>Hevea brasiliensis</i>	33325119	7E-80	J
MDI40_A06	1	<i>eukaryotic translation initiation factor 5A isoform VIII</i>	<i>Hevea brasiliensis</i>	33325131	3E-18	J
MDI29_E12	1	<i>Exonuclease</i>	<i>Arabidopsis thaliana</i>	30683321	1E-32	L
CL16 contig1	9	<i>extensin like protein</i>	<i>Populus nigra</i>	1199774	2E-14	S
CL426 contig1	2	<i>extensin like protein</i>	<i>Populus nigra</i>	1199774	6E-15	S
CL58 contig1	4	<i>extensin-like protein</i>	<i>Cucumis sativus</i>	17154773	9E-64	S
CL316Contig1	2	<i>F0 ATP synthase D chain, mitochondrial</i>	<i>Arabidopsis thaliana</i>	21555349	7E-75	C
MDI49-F01	1	<i>F1-ATP synthase delta subunit</i>	<i>Ipomoea batatas</i>	4774163	6E-50	C
MDI26_F01	1	<i>F1-ATPase gamma subunit mitochondrial</i>	<i>Ipomoea nil</i>	61675574	8E-33	C
MDI27_A12	1	<i>F1F0-ATPase inhibitor protein</i>	<i>Oryza sativa</i>	51077175	2E-22	C
MDI14_B08	1	<i>farnesylated protein, putative</i>	<i>Brassica rapa subsp. pekinensis</i>	32442810	8E-32	O
CL123 contig1	3	<i>farnesylated protein, putative</i>	<i>Oryza sativa</i>	50908877	5E-28	W
CL175 contig1	2	<i>fasciclin-like AGP 12</i>	<i>Populus alba x Populus tremula</i>	47717927	4E-52	I
MDI41_D11	1	<i>fasciclin-like AGP 12</i>	<i>Populus alba x Populus tremula</i>	47717927	7E-15	I

MDI01-D06	1	<i>Fatty acid cis-trans isomerase</i>	<i>Pseudomonas fluorescens PfO-1</i>	77459328	4E-11	I
MDI42_D09	1	<i>fatty acid desaturase 6, chloroplast</i>	<i>Olea europaea subsp. europaea</i>	54873456	1E-14	I
MDI05-E05	1	<i>fatty acid elongase, putative</i>	<i>Arabidopsis thaliana</i>	21536949	1E-29	I
MDI26_E03	1	<i>fatty acid elongase-like protein</i>	<i>Arabidopsis thaliana</i>	9294653	4E-32	I
MDI37_B04	1	<i>fatty acid hydroperoxide lyase</i>	<i>Psidium guayava</i>	13183137	1E-20	I
MDI20_F12	1	<i>Fatty acid hydroxylase 1 - FAH1; catalytic</i>	<i>Arabidopsis thaliana</i>	15226828	1E-68	I
MDI47-E04	1	<i>fatty acid hydroxylase-like protein</i>	<i>Arabidopsis thaliana</i>	7268883	3E-21	I
CL121Contig1	3	<i>Ferredoxin</i>	<i>Medicago sativa</i>	120008	6E-15	GPC
MDI16_F12	1	<i>ferric reductase</i>	<i>Oryza sativa</i>	47169677	3E-35	P
MDI09-F04	1	<i>Fe-SOD</i>	<i>Cinnamomum camphora</i>	3599469	3E-20	P
MDI48-E04	1	<i>fiber dTDP-glucose 4-6-dehydratase</i>	<i>Gossypium barbadense</i>	38016529	1E-84	S
MDI48-G01	1	<i>fiber lipid transfer protein</i>	<i>Gossypium barbadense</i>	25992534	9E-24	S
MDI38_G10	1	<i>fiber protein Fb19</i>	<i>Gossypium barbadense</i>	38016525	2E-18	S
MDI25_B06	1	<i>fiber protein Fb34</i>	<i>Gossypium barbadense</i>	38016521	2E-10	S
MDI16_C09	1	<i>fiber protein Fb35</i>	<i>Gossypium barbadense</i>	38016523	4E-10	S
MDI27_G10	1	<i>fiber protein Fb38</i>	<i>Gossypium barbadense</i>	38016533	4E-13	S
CL296Contig1	2	<i>fiber protein Fb9</i>	<i>Gossypium barbadense</i>	25992539	3E-16	S
MDI01-F08	1	<i>filamentous hemagglutinin</i>	<i>Pseudomonas syringae pv. phaseolicola 1448A</i>	71736182	5E-39	S
MDI43_A03	1	<i>FK506 binding / peptidyl-prolyl cis-trans isomerase</i>	<i>Arabidopsis thaliana</i>	15235203	4E-36	O
MDI10-C10	1	<i>flavanone 3-hydroxylase</i>	<i>Camellia sinensis</i>	49861114	1E-89	Q
MDI24_G01	1	<i>flavonoid 3'-hydroxylase</i>	<i>Vitis vinifera</i>	78183424	7E-46	Q
MDI05-A10	1	<i>flavonol synthase, putative</i>	<i>Oryza sativa</i>	50946913	9E-26	Q
CL365 contig1	2	<i>flavonol synthase/ iron ion binding / isopenicillin-N synthase</i>	<i>Arabidopsis thaliana</i>	30685157	2E-70	Q
MDI13_C02	1	<i>formamidase/ hydrolase, acting on carbon-nitrogen (but not peptide) bonds, in linear amides</i>	<i>Arabidopsis thaliana</i>	18420071	4E-36	GIC
MDI06-A08	1	<i>formate dehydrogenase</i>	<i>Lycopersicon esculentum</i>	56562181	3E-87	C
MDI45-D01	1	<i>fructokinase 2; S2 self-incompatibility locus-linked 3.16 protein, putative</i>	<i>Petunia integrifolia subsp. Inflata</i>	33329200	5E-78	G
MDI18_E09	1	<i>fructose-6-phosphate 2-kinase/fructose-2,6-bisphosphatase</i>	<i>Solanum tuberosum</i>	3309583	5E-40	G
CL161 contig1	3	<i>fructose-bisphosphate aldolase</i>	<i>Persea americana</i>	10800924	3E-125	GC
MDI02-H04	1	<i>fructose-bisphosphate aldolase</i>	<i>Arabidopsis thaliana</i>	15226185	4E-55	GC
MDI09-E07	1	<i>fructose-bisphosphate aldolase-like</i>	<i>Solanum tuberosum</i>	78191410	9E-76	GC
MDI13_A07	1	<i>FtsH protease, putative</i>	<i>Lycopersicon esculentum</i>	30526336	8E-21	O
MDI01-A01	1	<i>G protein beta subunit 2</i>	<i>Solanum tuberosum</i>	15778632	5E-37	TD
CL114 contig1	3	<i>G10 protein, putative</i>	<i>Oryza sativa</i>	52353717	5E-69	K
MDI20_D07	1	<i>G10 protein, putative</i>	<i>Oryza sativa</i>	52353717	8E-76	K
CL442 contig1	2	<i>GA 20-oxidase 2</i>	<i>Nerium oleander</i>	51011362	2E-40	Q
MDI42_H08	1	<i>GA 2-oxidase 2</i>	<i>Nerium oleander</i>	51011366	2E-43	Q

MDI43-H05	1	GA 2-oxidase 2	<i>Nerium oleander</i>	51011366	3E-38	Q
MDI18_A06	1	GAI protein	<i>Oryza sativa</i>	39841619	4E-28	S
MDI48-H02	1	galactose transporter	<i>Arabidopsis thaliana</i>	42565652	6E-16	G
MDI38_A09	1	Galactoside 2-alpha-L-fucosyltransferase (Xyloglucan alpha-(1,2)-fucosyltransferase) (PsFT1)	<i>Pisum sativum</i>	20138107	5E-35	E
MDI26_F04	1	galactosyltransferase, 3'-partial, putative	<i>Oryza sativa</i>	41469412	1E-75	G
MDI38_H02	1	gamma-aminobutyrate transaminase subunit precursor isozyme 3	<i>Lycopersicon esculentum</i>	29837286	3E-40	GE
CL277 contig1	2	gamma-glutamyltransferase-like protein	<i>Arabidopsis thaliana</i>	15233493	2E-75	H
MDI35_C10	1	gamma-lyase, putative	<i>Oryza sativa</i>	37536090	7E-25	GCE
MDI50-H02	1	gamma-TIP-like protein	<i>Trifolium repens</i>	453980	4E-36	S
MDI52_H07	1	gamma-TIP-like protein	<i>Trifolium repens</i>	453980	2E-21	S
MDI30_G01	1	gamma-tocopherol methyltransferase	<i>Glycine max</i>	62126057	4E-29	S
CL68 contig1	4	GDP dissociation inhibitor	<i>Cicer arietinum</i>	3175990	9E-132	S
MDI35_C03	1	GDP-L-fucose synthetase, putative	<i>Arabidopsis thaliana</i>	20465283	2E-37	OG
MDI25_C09	1	GDP-mannose 3',5'-epimerase	<i>Oryza sativa</i>	54650506	9E-22	G
CL69 contig1	4	GDP-mannose 3,5-epimerase/ catalytic	<i>Arabidopsis thaliana</i>	15241945	4E-97	G
MDI15_F06	1	GDP-mannose 3,5-epimerase/ catalytic	<i>Arabidopsis thaliana</i>	15241945	1E-09	G
MDI35_B05	1	GDP-mannose pyrophosphorylase, putative	<i>Oryza sativa</i>	50920467	5E-56	GOM
MDI24_A11	1	GDSL-motif lipase/hydrolase, putative	<i>Arabidopsis thaliana</i>	51968790	4E-34	S
CL56 contig1	4	Germin-Like Protein 2A - GLP2A; nutrient reservoir	<i>Arabidopsis thaliana</i>	15241662	7E-48	S
CL6 contig1	12	Gibberellin 2 oxidase	<i>Nerium oleander</i>	50293061	3E-106	Q
MDI01-E08	1	Gibberellin 2-beta-dioxygenase	<i>Arabidopsis thaliana</i>	79318890	4E-24	Q
MDI46-E10	1	gibberellin 2-oxidase 2	<i>Nicotiana tabacum</i>	46849531	2E-11	Q
MDI20_D09	1	Gigantea	<i>Arabidopsis thaliana</i>	50660455	4E-86	S
MDI19_B03	1	glucose 6 phosphate/phosphate translocator-like protein	<i>Arabidopsis thaliana</i>	9755788	1E-39	EG
CL219 contig1	2	glutamate decarboxylase 1	<i>Brassica juncea</i>	45935274	4E-44	E
MDI17_E05	1	glutamate decarboxylase, putative	<i>Arabidopsis thaliana</i>	19698825	7E-103	E
MDI39_F01	1	Glutaredoxin	<i>Tilia platyphyllos</i>	15637350	3E-32	O
MDI33_B09	1	glutathione peroxidase	<i>Hevea brasiliensis</i>	18026892	5E-56	O
CL5 contig1	13	glutathione S-transferase	<i>Carica papaya</i>	2853219	5E-92	O
MDI24_C04	1	glutathione S-transferase	<i>Euphorbia esula</i>	8099671	1E-54	O
MDI02-H07	1	glutathione S-transferase	<i>Cucurbita maxima</i>	23978432	2E-63	O
MDI08-G04	1	glutathione S-transferase	<i>Curcubita maxima</i>	23978432	2E-47	O
MDI09-A08	1	glutathione S-transferase	<i>Curcubita maxima</i>	23978432	4E-51	O
MDI11_F12	1	glutathione S-transferase	<i>Curcubita maxima</i>	23978432	6E-26	O
MDI23_E07	1	glutathione S-transferase	<i>Curcubita maxima</i>	23978432	9E-36	O

MDI27_D01	1	<i>glutathione S-transferase</i>	<i>Cucurbita maxima</i>	23978432	1E-33	O
MDI38_G02	1	<i>glutathione S-transferase</i>	<i>Cucurbita maxima</i>	23978432	7E-19	O
MDI40_F09	1	<i>glutathione S-transferase</i>	<i>Cucurbita maxima</i>	23978432	3E-58	O
MDI44-H07	1	<i>glutathione S-transferase</i>	<i>Cucurbita maxima</i>	23978432	3E-37	O
MDI52_F09	1	<i>glutathione S-transferase</i>	<i>Cucurbita maxima</i>	23978432	1E-24	O
MDI29_G04	1	<i>glutathione S-transferase</i>	<i>Arabidopsis thaliana</i>	15375408	9E-27	O
MDI46-D04	1	<i>glutathione S-transferase</i>	<i>Cucurbita maxima</i>	23978432	3E-53	O
MDI17_G05	1	<i>glutathione S-transferase GST 18</i>	<i>Glycine max</i>	11385451	1E-13	O
MDI43_B04	1	<i>glutathione S-transferase GST 18</i>	<i>Glycine max</i>	11385451	3E-16	O
MDI39_E04	1	<i>glutathione S-transferase GST 18</i>	<i>Populus alba</i> x <i>P. tremula</i> var. <i>glandulosa</i>	61287178	1E-22	O
MDI37_A01	1	<i>glutathione S-transferase GST 19</i>	<i>Glycine max</i>	11385453	4E-35	O
MDI18_F05	1	<i>glutathione S-transferase T3, putative</i>	<i>Lycopersicon esculentum</i>	10567808	2E-38	O
MDI06-E07	1	<i>glutathione S-transferase T4, putative</i>	<i>Lycopersicon esculentum</i>	10567810	5E-49	O
MDI29_A03	1	<i>glutathione synthetase</i>	<i>Arabidopsis thaliana</i>	5531231	2E-27	O
MDI16_G11	1	<i>glyceraldehyde 3-phosphate dehydrogenase</i>	<i>Magnolia quinquepeta</i>	19566	4E-53	G
CL172 contig1	3	<i>glyceraldehyde 3-phosphate dehydrogenase</i>	<i>Magnolia quinquepeta</i>	19566	1E-98	G
CL47 contig1	5	<i>glyceraldehyde-3-phosphate dehydrogenase</i>	<i>Nicotiana tabacum</i>	4539543	7E-117	G
MDI17_E09	1	<i>glycerophosphodiester phosphodiesterase/ kinase</i>	<i>Arabidopsis thaliana</i>	18416801	8E-42	C
MDI41_B02	1	<i>glycerophosphoryl diester phosphodiesterase family protein, putative</i>	<i>Nicotiana tabacum</i>	83281191	2E-11	C
MDI09-A01	1	<i>glycine decarboxylase P subunit</i>	x <i>Tritordeum</i> sp.	2565305	3E-43	S
MDI20_H01	1	<i>glycine hydroxymethyltransferase</i>	<i>Solanum tuberosum</i>	438247	7E-95	E
MDI42_F07	1	<i>glycine hydroxymethyltransferase (EC 2.1.2.1)-like protein</i>	<i>Arabidopsis thaliana</i>	7270156	6E-95	E
CL20 contig1	6	<i>glycine-rich RNA-binding protein, putative</i>	<i>Prunus avium</i>	34851124	3E-32	S
CL41 contig1	5	<i>glycine-rich RNA-binding protein, putative</i>	<i>Prunus avium</i>	34851124	5E-34	S
MDI24_C05	1	<i>glycine-rich RNA-binding protein-like</i>	<i>Arabidopsis thaliana</i>	21553602	4E-26	S
MDI22_E07	1	<i>glycolate oxidase subunit D-like; D-lactate dehydrogenase-like</i>	<i>Arabidopsis thaliana</i>	10178114	2E-38	C
MDI06-E01	1	<i>glycolate oxidase/ oxidoreductase</i>	<i>Arabidopsis thaliana</i>	15229497	3E-42	C
CL249 contig1	2	<i>glycolytic glyceraldehyde 3-phosphate dehydrogenase</i>	<i>Antirrhinum majus</i>	1345501	3E-50	G
MDI24_G06	1	<i>glycosylphosphatidylinositol-anchored protein, putative</i>	<i>Arabidopsis thaliana</i>	14193657	1E-51	O
MDI50-C05	1	<i>Glycosyltransferase</i>	<i>Gossypium raimondii</i>	63087742	1E-56	GM
MDI22_G10	1	<i>Glycosyltransferase</i>	<i>Dictyostelium discoideum AX4</i>	66811414	9E-38	GM
CL434 contig1	2	<i>glycosyltransferase NTGT5b</i>	<i>Nicotiana tabacum</i>	62241067	2E-44	GM
MDI14_A06	1	<i>glycosyltransferase, putative</i>	<i>Lotus corniculatus</i> var. <i>japonicus</i>	55956984	4E-23	GM
MDI27_B11	1	<i>glycyl-tRNA synthetase, putative</i>	<i>Oryza sativa</i>	50948039	2E-33	J
MDI35_D08	1	<i>Glyoxalase 2-2 - GLX2-2; hydroxyacylglutathione hydrolase</i>	<i>Arabidopsis thaliana</i>	15228389	4E-32	S
CL304 contig1	2	<i>glyoxalase II, putative</i>	<i>Arabidopsis thaliana</i>	12324040	4E-82	G

MDI35_E03	1	<i>granule-bound starch synthase Ib precursor</i>	<i>Phaseolus vulgaris</i>	30698530	1E-89	G
MDI22_A10	1	<i>Grp94</i>	<i>Xerophyta viscosa</i>	23477636	3E-53	O
CL177 contig1	2	<i>GSK1 (GSK3/Shaggy-like protein kinase 1); kinase</i>	<i>Arabidopsis thaliana</i>	42571361	1E-158	S
MDI12_H10	1	<i>GSK1 (GSK3/Shaggy-like protein kinase 1); kinase</i>	<i>Arabidopsis thaliana</i>	42571361	2E-20	S
MDI30_A09	1	<i>GTP binding / translation factor; nucleic acid binding / translation release factor</i>	<i>Arabidopsis thaliana</i>	79318119	1E-26	TJ
CL362 contig1	2	<i>GTP binding protein</i>	<i>Cichorium intybus x C. endivia</i>	12311680	3E-54	TJ
MDI46-E08	1	<i>GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase</i>	<i>Malus x domestica</i>	33308405	7E-24	T
MDI12_D09	1	<i>GTPase activating protein-like</i>	<i>Oryza sativa</i>	55773891	3E-20	T
MDI29_G11	1	<i>GTP-binding protein</i>	<i>Oryza sativa</i>	13195452	1E-43	T
MDI22_B07	1	<i>GTP-binding protein, putative</i>	<i>Arabidopsis thaliana</i>	12324651	3E-56	T
MDI07-E02	1	<i>GTP-binding protein, putative</i>	<i>Arabidopsis thaliana</i>	15231462	2E-44	T
MDI20_E06	1	<i>GTP-binding protein, putative</i>	<i>Arabidopsis thaliana</i>	62320005	2E-62	T
MDI01-C11	1	<i>Guanine nucleotide-binding protein beta subunit-like protein</i>	<i>Glycine max</i>	3023858	3E-48	D
MDI17_G08	1	<i>Guanine nucleotide-binding protein beta subunit-like protein</i>	<i>Glycine max</i>	3023858	5E-91	D
MDI48-B10	1	<i>Guanine nucleotide-binding protein beta subunit-like protein</i>	<i>Glycine max</i>	3023858	4E-58	D
MDI03-E07	1	<i>H⁺/Ca²⁺ exchanger 2</i>	<i>Ipomoea nil</i>	4512263	4E-14	P
CL18 contig1	8	<i>H⁺-pyrophosphatase</i>	<i>Vitis vinifera</i>	7739779	0.0	P
MDI02-G02	1	<i>harpin inducing protein</i>	<i>Nicotiana tabacum</i>	7417006	8E-32	S
CL339Contig1	2	<i>harpin-induced protein, putative</i>	<i>Arabidopsis thaliana</i>	21617965	4E-41	S
MDI05-C01	1	<i>hAT dimerisation domain-containing protein -like</i>	<i>Oryza sativa</i>	56784061	7E-25	S
MDI08-A03	1	<i>hAT dimerisation domain-containing protein, putative</i>	<i>Oryza sativa</i>	54291760	1E-57	S
MDI11_A06	1	<i>hAT family dimerisation domain, putative</i>	<i>Oryza sativa</i>	77552015	3E-25	S
MDI25_D04	1	<i>Heat shock 22 kd protein-like, mitochondrial</i>	<i>Arabidopsis thaliana</i>	21554988	8E-28	O
CL170 contig1	3	<i>Heat shock cognate 70 kDa protein</i>	<i>Petunia x hybrida</i>	123650	2E-102	O
MDI51-A05	1	<i>heat shock factor binding protein 2</i>	<i>Zea mays</i>	38349095	2E-20	O
MDI04-B04	1	<i>heat shock protein 17.3 kDa class I</i>	<i>Arachis hypogaea</i>	83701135	9E-19	O
CL363 contig1	2	<i>heat shock protein 26</i>	<i>Nicotiana glauca</i>	3256378	2E-55	O
MDI09-E03	1	<i>heat shock protein 70</i>	<i>Nicotiana tabacum</i>	19878	6E-105	O
MDI41_D10	1	<i>heat shock protein 70</i>	<i>Nicotiana tabacum</i>	19878	2E-20	O
MDI16_G05	1	<i>heat shock protein 70</i>	<i>Cucumis sativus</i>	6911549	1E-21	O
CL134 contig1	2	<i>heat shock protein 70-3</i>	<i>Nicotiana tabacum</i>	38325815	2E-77	O
MDI20_C09	1	<i>heat shock protein 70-3</i>	<i>Nicotiana tabacum</i>	38325815	2E-59	O
MDI15_H06	1	<i>heat shock protein 82</i>	<i>Nicotiana tabacum</i>	19880	2E-42	O
CL36 contig1	5	<i>Heat shock protein 83</i>	<i>Ipomoea nil</i>	1708314	0.0	O
CL70 contig1	4	<i>Heat shock protein 83</i>	<i>Ipomoea nil</i>	1708314	0.0	O

MDI02-E05	1	<i>heat shock protein 90</i>	<i>Nicotiana tabacum</i>	46093890	4E-62	O
MDI32-B07	1	<i>heat shock protein binding / unfolded protein binding</i>	<i>Arabidopsis thaliana</i>	15228802	2E-21	O
MDI17_D01	1	<i>heat shock protein binding / unfolded protein binding</i>	<i>Arabidopsis thaliana</i>	30689058	8E-11	O
MDI50-F05	1	<i>heat shock protein Gmhsp17.9-D (class VI)</i>	<i>Glycine max</i>	18651	5E-57	O
MDI14_D03	1	<i>Heat shock protein Hsp20.1 protein</i>	<i>Lycopersicon peruvianum</i>	3341464	2E-15	O
MDI14_E08	1	<i>Heat shock protein HSP21 (Heat Shock Protein 21)</i>	<i>Arabidopsis thaliana</i>	15234240	3E-15	O
MDI05-H11	1	<i>Heat shock protein HSP82</i>	<i>Zea mays</i>	7546186	1E-51	O
MDI45-H06	1	<i>heat shock transcription factor 29</i>	<i>Glycine max</i>	662925	5E-15	O
MDI12_F09	1	<i>heat-shock protein 80</i>	<i>Euphorbia esula</i>	6934298	9E-63	O
MDI24_A10	1	<i>heme binding / iron ion binding / monooxygenase/ oxygen binding</i>	<i>Arabidopsis thaliana</i>	18398119	6E-76	S
MDI12_G01	1	<i>hexose transporter 6</i>	<i>Vitis vinifera</i>	61613105	5E-34	R
MDI12_G03	1	<i>hexose transporter HT2</i>	<i>Vitis vinifera</i>	56759682	7E-40	R
MDI50-H01	1	<i>HHOA; serine-type endopeptidase/ serine-type peptidase/ trypsin</i>	<i>Arabidopsis thaliana</i>	42566926	2E-33	O
MDI30_E08	1	<i>high mobility group protein</i>	<i>Cucumis sativus</i>	63020536	2E-31	S
MDI14_C07	1	<i>histone deacetylase HD2, putative</i>	<i>Oryza sativa</i>	7804960	4E-23	B
MDI34_B09	1	<i>histone H2A, putative</i>	<i>Oryza sativa</i>	50931693	2E-22	B
CL155 contig1	3	<i>histone H3.2</i>	<i>Medicago sativa</i>	488577	5E-65	B
CL447Contig1	2	<i>histone H3.2</i>	<i>Medicago sativa</i>	15238916	6E-65	BO
CL192 contig1	2	<i>HMGB1 (High Mobility Group B 1); transcription factor</i>	<i>Arabidopsis thaliana</i>	15231066	5E-36	K
CL354 contig1	2	<i>HMG-CoA synthase</i>	<i>Brassica juncea</i>	7799986	7E-82	K
MDI39_G05	1	<i>HMT3; homocysteine S-methyltransferase</i>	<i>Arabidopsis thaliana</i>	18403505	3E-22	E
MDI24_C10	1	<i>homeobox-leucine zipper protein</i>	<i>Glycine max</i>	1234900	8E-15	K
MDI33_C06	1	<i>homeodomain leucine zipper protein HDZ2</i>	<i>Phaseolus vulgaris</i>	15148918	2E-42	K
MDI15_D12	1	<i>homo-phytochelatin synthase</i>	<i>Glycine max</i>	18699092	4E-28	P
CL120 contig1	2	<i>Hsc70</i>	<i>Lycopersicon esculentum</i>	762844	2E-55	O
MDI14_F02	1	<i>Hsc70</i>	<i>Lycopersicon esculentum</i>	762844	2E-35	O
CL88 contig1	4	<i>Hsp18p</i>	<i>Nicotiana tabacum</i>	558606	3E-50	O
MDI06-C08	1	<i>Hsp22.3</i>	<i>Glycine max</i>	710434	3E-31	O
MDI41_H05	1	<i>Hsp70 interacting protein/thioredoxin chimera</i>	<i>Vitis labrusca</i>	14581677	2E-50	OT
MDI09-D06	1	<i>HSP70, putative</i>	<i>Oryza sativa</i>	34908140	5E-48	O
MDI15_C12	1	<i>HSP70T-2; ATP binding</i>	<i>Arabidopsis thaliana</i>	15225187	2E-89	O
MDI48-G07	1	<i>HSP81-1 (Heat Shock Protein 81-1); ATP binding / unfolded protein binding</i>	<i>Arabidopsis thaliana</i>	152372	8E-29	O
MDI07-C09	1	<i>hydrolase</i>	<i>Arabidopsis thaliana</i>	15223919	8E-13	S
MDI37_G05	1	<i>hydrolase, acting on acid anhydrides, catalyzing transmembrane movement of substances</i>	<i>Arabidopsis thaliana</i>	15236722	7E-24	S
CL190 contig1	2	<i>hydrolase, acting on glycosyl bonds</i>	<i>Arabidopsis thaliana</i>	15232958	9E-87	S
CL336 contig1	2	<i>hydrolase, hydrolyzing O-glycosyl compounds</i>	<i>Arabidopsis thaliana</i>	15230330	2E-96	S
MDI39_E09	1	<i>hydrolase, hydrolyzing O-glycosyl compounds</i>	<i>Arabidopsis thaliana</i>	15239973	1E-11	S

MDI13_B05	1	hydroxyproline-rich glycoprotein homolog	<i>Arabidopsis thaliana</i>	2244847	3E-50	S
MDI37_A03	1	hypersensitive-induced reaction protein	<i>Capsicum annuum</i>	46452120	1E-50	S
CL319 contig1	2	hypersensitive-induced response protein	<i>Cucumis sativus</i>	34484310	1E-82	S
MDI34_E10	1	IAA16 protein	<i>Gossypium hirsutum</i>	22531416	3E-26	S
MDI03-E12	1	IAA19; transcription factor	<i>Arabidopsis thaliana</i>	15232612	2E-28	K
MDI07-D03	1	IAA33; transcription factor	<i>Arabidopsis thaliana</i>	15242127	3E-28	K
CL142 contig1	3	IAA-amino acid hydrolase	<i>Arabidopsis suecica</i>	18652314	3E-52	S
MDI39_F05	1	import receptor subunit mitochondrial TOM7-1 (Translocase of outer membrane 7 kDa subunit 1)	<i>Solanum tuberosum</i>	26397273	2E-14	S
MDI23_E04	1	Importin alpha 2	<i>Capsicum annuum</i>	13752562	9E-36	Y
MDI49-D08	1	Inhibitor of apoptosis-like protein	<i>Arabidopsis thaliana</i>	21593126	5E-17	S
MDI37_A08	1	initiation factor eIF4A-15	<i>Helianthus annuus</i>	38564733	6E-84	J
CL244 contig1	2	inner membrane protein, chloroplast	<i>Pisum sativum</i>	599956	3E-11	O
MDI22_D07	1	inner membrane translocase TM17-2, mitochondrial	<i>Arabidopsis thaliana</i>	38678772	7E-35	OU
MDI30_A07	1	inositol or phosphatidylinositol kinase/ phosphotransferase, alcohol group as acceptor	<i>Arabidopsis thaliana</i>	15227924	2E-38	GI
MDI43-G05	1	inositol or phosphatidylinositol kinase/ phosphotransferase, alcohol group as acceptor	<i>Arabidopsis thaliana</i>	18395629	7E-49	GI
MDI49-B06	1	inositol or phosphatidylinositol kinase/ phosphotransferase, alcohol group as acceptor	<i>Arabidopsis thaliana</i>	18395629	5E-22	GI
MDI09-D03	1	inositol or phosphatidylinositol kinase/ phosphotransferase, alcohol group as acceptor	<i>Arabidopsis thaliana</i>	42571253	7E-54	GI
MDI14_E04	1	inositol or phosphatidylinositol kinase/ phosphotransferase, alcohol group as acceptor	<i>Arabidopsis thaliana</i>	18422029	1E-70	GI
MDI02-E02	1	Inositol-3-phosphate synthase (Myo-inositol-1-phosphate synthase) (MI-1-P synthase) (IPS)	<i>Nicotiana tabacum</i>	14548096	2E-83	GI
MDI33_E11	1	integral membrane protein, putative	<i>Oryza sativa</i>	50942125	2E-15	S
CL43 contig1	5	invertase inhibitor precursor, putative	<i>Platanus x acerifolia</i>	26190140	3E-24	S
MDI14_G06	1	iron inhibited ABC transporter 2, putative	<i>Oryza sativa</i>	50948579	1E-22	P
MDI03-C04	1	iron superoxide dismutase 3	<i>Arabidopsis thaliana</i>	21554867	1E-21	P
MDI41_F06	1	isochorismatase hydrolase-like protein	<i>Oryza sativa</i>	50910291	1E-55	R
CL220 contig1	2	isocitrate dehydrogenase (NAD+)/ oxidoreductase; IDH1 (Isocitrate Dehydrogenase 1)	<i>Arabidopsis thaliana</i>	15236932	5E-77	G
MDI19_B11	1	isocitrate lyase	<i>Aspergillus fumigatus Af293</i>	66845417	2E-68	GC
MDI05-H06	1	KAK (KAKTUS)	<i>Arabidopsis thaliana</i>	42570183	8E-79	S
MDI18_B02	1	karyopherin-beta 3 variant, putative	<i>Oryza sativa</i>	50938581	1E-33	Y
MDI02-C04	1	ketol-acid reductoisomerase	<i>Arabidopsis thaliana</i>	15231092	5E-101	E
MDI06-A06	1	Ketopantoate reductase ApbA/PanE	<i>Pseudomonas fluorescens PfO-1</i>	77460643	1E-91	S
MDI20_F01	1	KH domain-containing protein NOVA-like	<i>Oryza sativa</i>	50915982	3E-23	J
MDI04-H02	1	Kinase	<i>Arabidopsis thaliana</i>	15235548	7E-51	S
MDI22_H10	1	kinesin heavy chain, putative	<i>Arabidopsis thaliana</i>	4567265	2E-53	Z
MDI12_E08	1	KUP11; potassium ion transporter	<i>Arabidopsis thaliana</i>	15226892	1E-53	L

MDI18_C05	1	<i>KUP3; potassium ion transporter</i>	<i>Arabidopsis thaliana</i>	15232841	1E-27	L
MDI10-C01	1	<i>L24 ribosomal protein, putative</i>	<i>Ipomoea batatas</i>	64420807	7E-27	J
MDI17_H04	1	<i>L24 ribosomal protein, putative</i>	<i>Ipomoea batatas</i>	64420807	5E-48	J
MDI30_G06	1	<i>L3 Ribosomal protein</i>	<i>Medicago sativa subsp. X varia</i>	6688812	7E-54	J
CL124 contig1	3	<i>laccase (diphenol oxidase)</i>	<i>Arabidopsis thaliana</i>	10177761	1E-52	S
MDI41_D03	1	<i>laccase (diphenol oxidase)</i>	<i>Arabidopsis thaliana</i>	10177761	2E-60	S
MDI03-C07	1	<i>lactoylglutathione lyase</i>	<i>Arabidopsis thaliana</i>	18402873	4E-49	G
MDI39_B03	1	<i>L-asparaginase</i>	<i>Glycine max</i>	29420787	8E-22	EC
MDI33_D12	1	<i>l-asparaginase (l-asparagine amidohydrolase)</i>	<i>Arabidopsis thaliana</i>	9294462	3E-11	E
MDI18_B07	1	<i>late elongated hypocotyl</i>	<i>Castanea sativa</i>	51980218	6E-23	S
CL46 contig1	5	<i>Late embryogenesis abundant protein Lea14-A</i>	<i>Gossypium hirsutum</i>	1170745	2E-52	S
MDI40_B09	1	<i>latex profilin Hev b 8</i>	<i>Hevea brasiliensis</i>	6979167	2E-32	S
MDI13_H12	1	<i>latex profilin Hev b 8</i>	<i>Hevea brasiliensis</i>	6979167	3E-27	S
MDI46-C08	1	<i>latex profilin Hev b 8</i>	<i>Hevea brasiliensis</i>	6979167	2E-49	S
MDI17_H01	1	<i>latex-abundant protein</i>	<i>Hevea brasiliensis</i>	4235430	9E-60	S
MDI32-E09	1	<i>latex-abundant protein</i>	<i>Hevea brasiliensis</i>	4235430	6E-23	S
MDI34_G02	1	<i>latex-abundant protein</i>	<i>Hevea brasiliensis</i>	4235430	5E-38	S
MDI21_C08	1	<i>lectin-like receptor kinase 7;3</i>	<i>Oryza sativa</i>	77556195	5E-27	U
MDI25_A01	1	<i>leucine aminopeptidase, putative</i>	<i>Oryza sativa</i>	50915564	4E-98	S
CL344Contig1	2	<i>leucine rich repeat protein</i>	<i>Cicer arietinum</i>	38566726	9E-46	Z
MDI18_A04	1	<i>Leucine Rich Repeat, putative</i>	<i>Oryza sativa</i>	77549387	2E-26	Z
MDI02-E01	1	<i>Leucine Rich Repeat, putative</i>	<i>Oryza sativa</i>	77554041	2E-20	Z
MDI11_B08	1	<i>leucine zipper-containing protein, putative</i>	<i>Oryza sativa</i>	50908781	4E-36	T
MDI13_H05	1	<i>leucine-rich protein, putative</i>	<i>Oryza sativa</i>	50910025	3E-56	Z
MDI02-A02	1	<i>Leucine-rich repeats and calponin homology domain containing protein 1</i>	<i>Mus musculus</i>	82955323	5E-20	Z
MDI16_D03	1	<i>leucoanthocyanidin dioxygenase</i>	<i>Vitis labrusca x Vitis vinifera</i>	22266677	2E-82	Q
MDI42_H12	1	<i>leucyl-tRNA synthetase, putative</i>	<i>Arabidopsis thaliana</i>	62321577	3E-79	J
MDI22_B06	1	<i>LIM domain containing protein-like</i>	<i>Oryza sativa</i>	55771369	3E-31	R
MDI27_G09	1	<i>LIN2 (Lesion Initiation 2); coproporphyrinogen oxidase</i>	<i>Arabidopsis thaliana</i>	15218820	7E-26	R
MDI01-D01	1	<i>lipase, putative</i>	<i>Arabidopsis thaliana</i>	15293170	6E-27	I
MDI25_C01	1	<i>lipid binding</i>	<i>Arabidopsis thaliana</i>	15221585	1E-14	S
MDI23_G03	1	<i>lipid binding</i>	<i>Arabidopsis thaliana</i>	30685701	9E-41	S
CL61 contig1	4	<i>lipid transfer protein 3 precursor</i>	<i>Gossypium hirsutum</i>	11095210	8E-29	S
MDI32-A12	1	<i>lipid transfer protein; glossy1 homolog</i>	<i>Arabidopsis thaliana</i>	9758349	1E-45	S
MDI31_A12	1	<i>L-isoaspartyl methyltransferase</i>	<i>Arabidopsis thaliana</i>	1322021	7E-54	S
CL91 contig1	3	<i>low temperature and salt responsive protein</i>	<i>Solanum tuberosum</i>	24745954	6E-20	S
CL45 contig1	5	<i>low temperature and salt responsive protein</i>	<i>Solanum tuberosum</i>	24745954	3E-20	S

MDI23_E03	1	<i>lupeol synthase</i>	<i>Betula platyphylla</i>	18147594	5E-89	S
MDI32-G08	1	<i>lustrin A-like</i>	<i>Oryza sativa</i>	53791963	1E-19	S
MDI01-G04	1	<i>lysine decarboxylase-like protein</i>	<i>Oryza sativa</i>	53792247	4E-35	Q
MDI02-D07	1	<i>lysine ketoglutarate reductase trans-splicing related 1</i>	<i>Arabidopsis thaliana</i>	24636978	2E-44	E
MDI33_H05	1	<i>lysophospholipase-like protein</i>	<i>Arabidopsis thaliana</i>	7573379	7E-38	I
MDI41_B08	1	<i>lysosomal Pro-X carboxypeptidase - like protein</i>	<i>Arabidopsis thaliana</i>	2827710	2E-34	S
CL90 contig1	4	<i>LYTB-like protein precursor</i>	<i>Adonis palaestina</i>	10716963	1E-154	K
MDI27_G05	1	<i>macronuclear development protein, putative</i>	<i>Oryza sativa</i>	50582717	1E-33	S
MDI43-H12	1	<i>MADS box protein</i>	<i>Prunus persica</i>	70955228	4E-66	K
CL478 contig1	2	<i>MADS box protein AP2L</i>	<i>Eucalyptus globulus</i>	11120557	8E-57	K
MDI27_F11	1	<i>MADS box protein SEP3</i>	<i>Lotus corniculatus var. japonicus</i>	60100342	4E-35	K
MDI09-D01	1	<i>maize transposon MuDR mudrA-like protein, putative</i>	<i>Oryza sativa</i>	37530384	2E-10	S
CL1 contig1	19	<i>major allergen Mal d 1.06A02</i>	<i>Malus x domestica</i>	60290811	1E-41	S
CL130Contig1	2	<i>major latex-like protein</i>	<i>Rubus idaeus</i>	3164115	2E-30	S
MDI24_F10	1	<i>major latex-like protein</i>	<i>Rubus idaeus</i>	3164115	2E-29	S
MDI43_A06	1	<i>makorin Ring finger protein, putative</i>	<i>Oryza sativa</i>	54290921	5E-10	K
MDI43-F10	1	<i>Mal d 1-like</i>	<i>Malus x domestica</i>	41322974	2E-46	S
CL81 contig1	4	<i>malate dehydrogenase</i>	<i>Nicotiana tabacum</i>	10798652	5E-65	G
MDI01-B08	1	<i>Mangrin</i>	<i>Bruguiera sexangula</i>	16416376	2E-47	S
MDI10-F02	1	<i>mannan endo-1,4-beta-mannosidase, putative</i>	<i>Oryza sativa</i>	34912090	3E-51	G
MDI27_C03	1	<i>mannose-6-phosphate isomerase, putative</i>	<i>Oryza sativa</i>	34895502	1E-34	G
CL28 contig1	5	<i>MAP kinase</i>	<i>Arabidopsis thaliana</i>	6580145	3E-119	T
MDI52_C11	1	<i>MAP3K delta-1 protein kinase, putative</i>	<i>Oryza sativa</i>	50906405	3E-18	T
MDI06-E05	1	<i>MATE efflux family protein, putative</i>	<i>Oryza sativa</i>	77548302	5E-49	S
MDI48-E05	1	<i>MATE efflux protein-like</i>	<i>Oryza sativa</i>	56784985	6E-61	S
MDI42_F08	1	<i>MCT-1 protein-like</i>	<i>Oryza sativa</i>	34894570	2E-79	S
MDI03-E10	1	<i>MDR1 (multidrug resistance protein 1)</i>	<i>Coptis japonica</i>	14715462	3E-77	R
MDI30_H04	1	<i>membrane trafficking factor, putative</i>	<i>Arabidopsis thaliana</i>	7267177	2E-67	S
MDI13_E12	1	<i>Membrane-associated 30 kDa protein, chloroplast precursor (M30).</i>	<i>Pisum sativum</i>	729842	1E-41	S
MDI24_B01	1	<i>metal ion binding</i>	<i>Arabidopsis thaliana</i>	15223416	1E-18	S
MDI52_D08	1	<i>metal ion binding</i>	<i>Arabidopsis thaliana</i>	15223416	7E-20	S
MDI20_D03	1	<i>metal ion binding</i>	<i>Arabidopsis thaliana</i>	15242741	2E-42	S
MDI36_B04	1	<i>metal ion binding</i>	<i>Arabidopsis thaliana</i>	30688566	9E-20	S
CL340 contig1	2	<i>metal ion binding</i>	<i>Arabidopsis thaliana</i>	15237243	2E-19	S
MDI45-D05	1	<i>metal ion binding</i>	<i>Arabidopsis thaliana</i>	22327990	2E-24	S
MDI26_B05	1	<i>metalloendopeptidase</i>	<i>Pisum sativum</i>	799369	4E-24	R

MDI24_E02	1	<i>Metallopeptidase</i>	<i>Arabidopsis thaliana</i>	42566909	6E-63	R
MDI14_E09	1	<i>metallophosphatase, putative</i>	<i>Lupinus luteus</i>	18075960	3E-79	R
CL10 contig1	11	<i>Metallothionein</i>	<i>Casuarina glauca</i>	1223652	2E-10	R
MDI48-F05	1	<i>metallothionein-like protein class II</i>	<i>Fagus sylvatica</i>	3901014	1E-20	R
MDI25_F09	1	<i>methionyl-tRNA synthetase, putative</i>	<i>Arabidopsis thaliana</i>	16323444	3E-18	J
MDI13_D04	1	<i>methyl-binding domain protein MBD4</i>	<i>Arabidopsis thaliana</i>	15229373	1E-15	S
MDI07-F05	1	<i>methyltransferase, putative</i>	<i>Solanum tuberosum</i>	53749450	5E-68	J
MDI47-E05	1	<i>MHK; ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase</i>	<i>Arabidopsis thaliana</i>	30682312	1E-16	S
MDI38_F01	1	<i>mitogen-activated protein kinase</i>	<i>Nicotiana tabacum</i>	34979909	2E-17	S
CL376 contig1	2	<i>mitogen-activated protein kinase 1, putative</i>	<i>Medicago sativa</i>	32400274	6E-63	S
CL466 contig1	2	<i>ML domain protein, putative</i>	<i>Nicotiana glauca</i>	55294668	3E-29	S
CL78 contig1	4	<i>Mob1/phocein family protein, putative</i>	<i>Solanum bulbocastanum</i>	32470635	8E-108	D
MDI14_E03	1	<i>Mob1-like protein</i>	<i>Medicago sativa subsp. falcata</i>	47846249	8E-81	D
CL86 contig1	4	<i>monodehydroascorbate reductase</i>	<i>Mesembryanthemum crystallinum</i>	15865451	2E-73	G
MDI19_G08	1	<i>monofunctional lysine-ketoglutarate reductase 2</i>	<i>Gossypium hirsutum</i>	23193178	3E-32	S
MDI32-D02	1	<i>MRP-like ABC transporter</i>	<i>Oryza sativa</i>	27263148	7E-69	S
MDI19_A11	1	<i>multicatalytic endopeptidase complex, proteasome component, beta subunit</i>	<i>Arabidopsis thaliana</i>	2511598	6E-30	S
MDI26_B07	1	<i>multifunctional aminoacyl-tRNA ligase-like protein</i>	<i>Arabidopsis thaliana</i>	6899922	1E-52	S
MDI03-G01	1	<i>myb-like transcription factor 2</i>	<i>Gossypioides kirkii</i>	23476295	4E-10	K
MDI14_H05	1	<i>myb-like transcription factor 6</i>	<i>Gossypium raimondii</i>	23476313	8E-71	K
MDI09-G04	1	<i>myb-related transcription factor-like</i>	<i>Oryza sativa</i>	50726439	1E-37	K
MDI02-E04	1	<i>MYH9_6</i>	<i>Arabidopsis thaliana</i>	20147241	2E-40	K
MDI33_B05	1	<i>myosin heavy chain homolog, putative</i>	<i>Oryza sativa</i>	34908914	8E-43	ZD
MDI43-F02	1	<i>myosin heavy chain -related-like</i>	<i>Oryza sativa</i>	50948487	4E-17	ZD
MDI50-C06	1	<i>Na+/K+/Cl-cotransport protein, putative</i>	<i>Oryza sativa</i>	50944163	4E-26	P
MDI22_E01	1	<i>NAC2; transcription factor</i>	<i>Arabidopsis thaliana</i>	15237699	4E-30	K
MDI45-D10	1	<i>NAC2; transcription factor</i>	<i>Arabidopsis thaliana</i>	15237699	2E-15	K
MDI37_G09	1	<i>NAC2; transcription factor</i>	<i>Arabidopsis thaliana</i>	15237699	5E-15	K
MDI36_E01	1	<i>N-acetyltransferase</i>	<i>Arabidopsis thaliana</i>	15224996	1E-63	I
MDI11_E10	1	<i>NADH:cytochrome b5 reductase</i>	<i>Vernicia fordii</i>	55979115	7E-53	HC
MDI17_F11	1	<i>NADH-dehydrogenase, putative</i>	<i>Pisum sativum</i>	37725947	2E-103	C
MDI26_G02	1	<i>NADH-ubiquinone oxidoreductase subunit PSST</i>	<i>Lupinus luteus</i>	9652180	2E-50	IC
MDI38_C05	1	<i>NADK2; NAD+ kinase</i>	<i>Arabidopsis thaliana</i>	18395013	3E-107	S
MDI47-C10	1	<i>NADP malic enzyme</i>	<i>Oryza sativa</i>	37694731	4E-22	S
MDI12_C03	1	<i>NADP-dependent glyceraldehyde-3-phosphate dehydrogenase</i>	<i>Nicotiana plumbaginifolia</i>	1842115	7E-79	GC

CL204 contig1	2	<i>NADP-dependent malic enzyme (NADP-ME)</i>	<i>Populus trichocarpa</i>	20469	2E-136	S
MDI40_D07	1	<i>NADP-dependent malic protein, putative</i>	<i>Oryza sativa</i>	57899543	1E-23	S
MDI12_E06	1	<i>NADPH-cytochrome P450 oxydoreductase isoform 3</i>	<i>Populus balsamifera subsp. trichocarpa</i> <i>X P. deltoides</i>	13183566	9E-11	Q
MDI34_A01	1	<i>NADPH-cytochrome P450 oxydoreductase isoform 3</i>	<i>Populus balsamifera subsp. trichocarpa</i> <i>X P. deltoides</i>	13183566	1E-33	Q
MDI39_C06	1	<i>NADPH-cytochrome P450 oxydoreductase isoform 3</i>	<i>Populus balsamifera subsp. trichocarpa</i> <i>X P. deltoides</i>	13183566	2E-90	Q
MDI50-G07	1	<i>NADPH-cytochrome P450 oxydoreductase isoform 3</i>	<i>Populus balsamifera subsp. trichocarpa</i> <i>X P. deltoides</i>	13183566	9E-21	Q
MDI08-B01	1	<i>NADPH-cytochrome P450 reductase, putative</i>	<i>Pisum sativum</i>	6503253	8E-78	Q
MDI06-A07	1	<i>NADPH-ferrihemoprotein reductase (ATR2)</i>	<i>Arabidopsis thaliana</i>	7269921	3E-57	S
MDI37_H10	1	<i>NADPH-protochlorophyllide oxidoreductase</i>	<i>Zea mays</i>	46019982	3E-59	S
MDI24_G08	1	<i>NADP-isocitrate dehydrogenase</i>	<i>Eucalyptus globulus</i>	1750380	2E-79	G
CL367 contig1	2	<i>nascent polypeptide associated complex alpha chain, putative</i>	<i>Oryza sativa</i>	50931251	2E-52	S
MDI34_E01	1	<i>nifU-like protein</i>	<i>Arabidopsis thaliana</i>	21595581	3E-32	O
MDI21_C09	1	<i>nitrate transporter, putative</i>	<i>Oryza sativa</i>	57899510	6E-30	S
MDI48-C11	1	<i>nitrate transporter, putative</i>	<i>Oryza sativa</i>	57899510	1E-11	S
MDI24_A12	1	<i>nitrate transporter, putative</i>	<i>Arabidopsis thaliana</i>	20465363	6E-39	S
CL484 contig1	2	<i>nitrate-induced NOI protein, putative</i>	<i>Arabidopsis thaliana</i>	20465673	9E-21	S
CL138 contig1	3	<i>nitrite transporter</i>	<i>Cucumis sativus</i>	15391731	5E-111	S
MDI33_G02	1	<i>nitrite transporter, putative</i>	<i>Oryza sativa</i>	53793165	3E-28	S
MDI12_D12	1	<i>NLI interacting factor (NIF) family protein-like</i>	<i>Oryza sativa</i>	56202314	2E-11	K
MDI27_B02	1	<i>nodulin protein, putative</i>	<i>Arabidopsis thaliana</i>	15028078	8E-43	S
MDI17_A08	1	<i>nodulin-like protein</i>	<i>Arabidopsis thaliana</i>	10177510	1E-33	S
MDI32-A05	1	<i>nodulin-like protein</i>	<i>Poncirus trifoliata</i>	24461856	3E-13	S
MDI08-A08	1	<i>nodulin-like protein</i>	<i>Gossypium hirsutum</i>	28932880	2E-56	S
CL135 contig1	3	<i>nodulin-like protein</i>	<i>Oryza sativa</i>	57900274	3E-15	S
MDI41_F08	1	<i>non-green plastid inner envelope membrane protein precursor.</i>	<i>Brassica oleracea var. botrytis</i>	1143711	2E-15	S
MDI34_C05	1	<i>Nramp metal transporter homolog</i>	<i>Thlaspi japonicum</i>	33235567	3E-23	S
MDI27_A11	1	<i>NRAMP2; metal ion transporter</i>	<i>Arabidopsis thaliana</i>	15220149	5E-18	S
MDI27_E02	1	<i>N-rich protein</i>	<i>Glycine max</i>	57898928	4E-17	S
MDI34_C08	1	<i>NRS/ER (Nucleotide-Rhamnose synthase/epimerase-reductase)</i>	<i>Arabidopsis thaliana</i>	18407710	1E-29	S
MDI45-C05	1	<i>NTL1 (nitrate transporter); calcium ion binding / transporter</i>	<i>Arabidopsis thaliana</i>	18409382	1E-49	S
MDI33_E10	1	<i>n-type ATP pyrophosphatase-like</i>	<i>Oryza sativa</i>	50913043	3E-33	S
MDI20_G01	1	<i>nucellin-like aspartic protease</i>	<i>Oryza sativa</i>	77553816	3E-17	S
MDI03-G06	1	<i>nuclear matrix protein 1</i>	<i>Lycopersicon esculentum</i>	15077144	1E-60	Y
MDI22_D11	1	<i>nuclear RNA binding protein A, putative</i>	<i>Oryza sativa</i>	34910712	5E-11	Y

MDI47-A12	1	Nuclear transport factor 2 (NTF-2)	<i>Schizosaccharomyces pombe</i>	19862916	1E-11	TY
CL21 contig1	4	nuclear transport factor 2, putative	<i>Arabidopsis thaliana</i>	21555203	2E-12	TY
CL21 contig2	4	nuclear transport factor 2, putative	<i>Arabidopsis thaliana</i>	21555203	1E-15	TY
MDI20_F02	1	nuclear transport factor 2, putative	<i>Arabidopsis thaliana</i>	21555203	4E-24	TY
MDI43-F04	1	nucleic acid binding	<i>Arabidopsis thaliana</i>	15228732	1E-27	YF
MDI20_E04	1	nucleic acid binding	<i>Arabidopsis thaliana</i>	42564243	2E-19	YF
CL353 contig1	2	nucleic acid binding	<i>Arabidopsis thaliana</i>	42569396	6E-52	YF
MDI49-A05	1	nucleic acid binding / transcription factor/ zinc ion binding	<i>Arabidopsis thaliana</i>	18402112	3E-18	KF
CL463 contig1	2	nucleic acid binding / zinc ion binding	<i>Arabidopsis thaliana</i>	18403023	1E-89	YF
MDI14_F11	1	nucleoporin PRECOZ, putative	<i>Arabidopsis thaliana</i>	22652297	2E-73	Y
MDI52_A02	1	nucleoside diphosphate kinase	<i>Nicotiana tabacum</i>	62114996	3E-69	YF
MDI31_B05	1	nucleotide binding	<i>Arabidopsis thaliana</i>	79313281	3E-32	YF
MDI22_A11	1	nucleotide binding	<i>Arabidopsis thaliana</i>	15220198	9E-61	YF
MDI40_H04	1	nucleotide binding	<i>Arabidopsis thaliana</i>	15237124	1E-29	YF
MDI21_B08	1	nucleotide binding	<i>Arabidopsis thaliana</i>	18379056	2E-46	YF
MDI10-F01	1	nucleotide-sugar transporter/ sugar porter	<i>Arabidopsis thaliana</i>	30689311	3E-38	OYFG
MDI10-C07	1	Nucleotidyltransferase	<i>Arabidopsis thaliana</i>	22328167	5E-42	Y
CL461 contig1	2	Oleosin	<i>Corylus avellana</i>	29170509	7E-28	S
MDI20_D08	1	oligopeptide transporter	<i>Arabidopsis thaliana</i>	15218799	2E-52	S
MDI05-C03	1	oligopeptide transporter OPT-like	<i>Oryza sativa</i>	50910954	9E-60	S
MDI06-H05	1	oligosaccharide transporter	<i>Arabidopsis thaliana</i>	15220609	3E-65	S
MDI17_F01	1	oligouridylate binding protein	<i>Nicotiana plumbaginifolia</i>	6996560	4E-18	S
MDI19_C09	1	oligouridylate binding protein, putative	<i>Arabidopsis thaliana</i>	21593280	1E-82	S
MDI14_F12	1	O-linked GlcNAc transferase, putative	<i>Arabidopsis thaliana</i>	21436428	3E-35	G
MDI36_C12	1	O-linked GlcNAc transferases, similarity	<i>Arabidopsis thaliana</i>	5732436	6E-23	G
MDI27_B01	1	O-linked N-acetyl glucosamine transferase	<i>Arabidopsis thaliana</i>	18139887	6E-15	OG
MDI22_C05	1	orcinol O-methyltransferase, putative	<i>Rosa gigantea</i>	55163142	1E-31	Q
MDI04-B03	1	ORG4 (OBP3-respsive gene 4)	<i>Arabidopsis thaliana</i>	18396223	1E-35	S
MDI48-G1	1	ornithine aminotransferase	<i>Vitis vinifera</i>	9957234	6E-22	E
CL87Contig1	3	O-succinylhomoserine sulfhydrylase, similar	<i>Arabidopsis thaliana</i>	15450932	7E-36	S
MDI12_E09	1	Outer envelope protein of 80 KDA	<i>Arabidopsis thaliana</i>	18419973	1E-15	S
MDI23_F11	1	oxalyl-CoA decarboxylase, putative	<i>Oryza sativa</i>	50901676	3E-85	S
MDI29_G06	1	Oxidoreductase	<i>Arabidopsis thaliana</i>	15233990	2E-47	R
MDI02-F01	1	Oxidoreductase	<i>Arabidopsis thaliana</i>	18391139	1E-69	R
MDI02-B10	1	Oxidoreductase	<i>Arabidopsis thaliana</i>	18402992	8E-54	R
CL213 contig1	2	oxidoreductase, 2OG-Fe(II) oxygenase family protein-like	<i>Oryza sativa</i>	55296462	2E-60	R
MDI21_A10	1	oxidoreductase, putative	<i>Oryza sativa</i>	77552296	2E-57	R

MDI05-H12	1	oxidoreductase/ zinc ion binding	<i>Arabidopsis thaliana</i>	42568692	3E-27	R
MDI03-D12	1	oxysterol binding	<i>Arabidopsis thaliana</i>	15238398	3E-53	TI
MDI47-F09	1	oxysterol binding protein, putative	<i>Oryza sativa</i>	57114406	6E-14	TI
MDI32-H07	1	oxysterol-binding protein	<i>Arabidopsis thaliana</i>	15238398	2E-70	TI
MDI22_E06	1	p23 co-chaperone, putative	<i>Oryza sativa</i>	50944819	2E-28	O
CL38 contig1	5	P23 tumor protein-like	<i>Solanum tuberosum</i>	78191404	4E-76	S
MDI51-B03	1	p32.2 annexin	<i>Nicotiana tabacum</i>	3219618	3E-28	S
MDI20_A11	1	p60 katanin, putative	<i>Oryza sativa</i>	55770193	1E-83	S
MDI21_G07	1	palmitoyl-protein thioesterase, putative	<i>Oryza sativa</i>	37536884	1E-65	OI
CL22 contig1	8	papain-like cysteine proteinase, putative	<i>Gossypium hirsutum</i>	57282617	5E-153	E
MDI33_C02	1	patellin 1 (PATL1)	<i>Cucurbita pepo</i>	82469975	6E-38	S
MDI15_C03	1	pathogenesis-related protein 10	<i>Vitis vinifera</i>	11182126	1E-56	S
MDI51-G11	1	PDI-like protein	<i>Zea mays</i>	4151124	4E-12	O
MDI32-B03	1	PDI-like protein	<i>Arabidopsis thaliana</i>	21592996	2E-47	O
MDI07-E05	1	PDI-like protein	<i>Quercus suber</i>	23304737	3E-43	O
MDI08-H01	1	PDI-like protein	<i>Quercus suber</i>	23304737	5E-44	O
MDI20_D01	1	PDI-like protein	<i>Quercus suber</i>	23304737	1E-23	O
MDI11-H04	1	PDR5-like ABC transporter	<i>Spirodela polyrhiza</i>	1514643	9E-56	Q
MDI05-D08	1	pectin methyl esterase	<i>Solanum tuberosum</i>	6689892	8E-79	S
MDI51-G08	1	pectinesterase, putative	<i>Arabidopsis thaliana</i>	7270826	6E-10	G
MDI25_G12	1	pentatricopeptide (PPR) repeat-containing protein, putative	<i>Oryza sativa</i>	50935049	8E-37	R
MDI02-G03	1	pentatricopeptide (PPR) repeat-containing protein, putative	<i>Oryza sativa</i>	50899520	1E-48	R
MDI27_C12	1	pepsin A	<i>Arabidopsis thaliana</i>	15229656	3E-44	S
MDI15_C07	1	peptidase/ pyroglutamyl-peptidase I	<i>Arabidopsis thaliana</i>	42562269	4E-53	S
CL57 contig1	4	peptidase/ serine-type peptidase	<i>Arabidopsis thaliana</i>	15222236	3E-12	S
MDI03-C04	1	peptidase/ serine-type peptidase	<i>Arabidopsis thaliana</i>	30695525	1E-48	S
MDI11-H06	1	peptide transport protein, putative	<i>Oryza sativa</i>	34895718	5E-16	S
MDI33_C11	1	peptidyl-prolyl cis-trans isomerase, similar	<i>Arabidopsis thaliana</i>	9797744	3E-34	O
MDI18_D04	1	Periplasmic solute binding protein	<i>Pseudomonas fluorescens PfO-1</i>	77456287	5E-33	S
MDI26_D10	1	permease 1, putative	<i>Arabidopsis thaliana</i>	21450871	2E-11	S
MDI12_A02	1	permease1	<i>Arabidopsis thaliana</i>	79600443	3E-74	S
MDI35_B07	1	Peroxidase	<i>Nicotiana tabacum</i>	5381253	1E-79	C
MDI46-F10	1	Peroxidase	<i>Arabidopsis thaliana</i>	15240737	2E-31	C
MDI49-H12	1	Peroxidase 63 precursor (Atperox P63) (ATP26a)	<i>Arabidopsis thaliana</i>	26397798	1E-22	R
MDI17_H03	1	peroxidase 7	<i>Triticum monococcum</i>	57635159	5E-43	R
MDI10-C12	1	Peroxisomal 2,4-dienoyl-CoA reductase	<i>Arabidopsis thaliana</i>	66774119	2E-88	S
MDI17_D04	1	P-glycoprotein	<i>Solanum tuberosum</i>	4204793	5E-96	S

MDI18_H10	1	<i>Phantastica</i>	<i>Lycopersicon esculentum</i>	5230656	5E-52	S
CL149 contig1	3	<i>Phenylalanine ammonia-lyase</i>	<i>Camellia sinensis</i>	1171998	3E-154	C
MDI15_A03	1	<i>phenylalanine ammonium lyase</i>	<i>Vitis vinifera</i>	1345583	6E-58	C
MDI15_G07	1	<i>phi-1</i>	<i>Nicotiana tabacum</i>	3759184	7E-45	S
MDI49-C06	1	<i>phi-1 (phosphate-induced protein 1), putative</i>	<i>Oryza sativa</i>	50933547	7E-51	S
MDI51-B02	1	<i>phi-1 protein, putative</i>	<i>Arabidopsis thaliana</i>	12642895	3E-25	S
CL202 contig1	2	<i>phi-1, putative</i>	<i>Oryza sativa</i>	50933547	1E-42	S
MDI26_A07	1	<i>phi-1-like phosphate-induced protein, putative</i>	<i>Arabidopsis thaliana</i>	21593043	1E-19	S
MDI33_F05	1	<i>phloroglucinol O-methyltransferase</i>	<i>Rosa chinensis var. spontanea</i>	47232556	1E-16	S
CL257Contig1	2	<i>phosphate translocator</i>	<i>Beta vulgaris</i>	5881143	9E-66	S
MDI07-G06	1	<i>phosphatase 2A regulatory A subunit</i>	<i>Oryza sativa</i>	50898976	1E-67	S
MDI44-A04	1	<i>phosphate translocator</i>	<i>Pisum sativum</i>	20853	7E-22	S
MDI07-A08	1	<i>Phosphate transport system permease protein 2</i>	<i>Pseudomonas fluorescens PfO-1</i>	77461834	2E-63	S
MDI14_C06	1	<i>phosphate transporter</i>	<i>Sesbania rostrata</i>	12697486	1E-88	S
MDI21_E12	1	<i>phosphate transporter</i>	<i>Sesbania rostrata</i>	12697486	1E-50	S
CL291Contig1	2	<i>phosphate transporter, mitochondrial</i>	<i>Arabidopsis thaliana</i>	3318617	2E-55	S
MDI12_C07	1	<i>phosphatidate cytidyltransferase</i>	<i>Arabidopsis thaliana</i>	15232383	6E-41	S
MDI44-F10	1	<i>phosphatidic acid phosphatase alpha</i>	<i>Vigna unguiculata</i>	9621903	2E-76	S
MDI30_A02	1	<i>phosphatidylglycerol specific phospholipase C, putative</i>	<i>Oryza sativa</i>	50904529	1E-27	S
MDI47-D06	1	<i>phosphoenolpyruvate carboxylase</i>	<i>Cucumis sativus</i>	20302419	3E-23	G
MDI07-G11	1	<i>phosphoenolpyruvate carboxylase, isoform 1</i>	<i>Mesembryanthemum crystallinum</i>	18073854	2E-106	G
MDI08-E04	1	<i>Phosphoglucomutase, chloroplast precursor (Glucose phosphomutase) (PGM)</i>	<i>Solanum tuberosum</i>	12585317	3E-111	G
MDI17_D07	1	<i>Phosphoglucomutase, cytoplasmic (Glucose phosphomutase) (PGM)</i>	<i>Mesembryanthemum crystallinum</i>	12585308	3E-107	G
CL139 contig1	3	<i>phosphoglycerate kinase 1, cytosolic</i>	<i>Populus nigra</i>	3738257	3E-173	G
MDI20_B01	1	<i>phosphoglycerate mutase-like protein</i>	<i>Glycine max</i>	21700765	2E-59	G
MDI19_A10	1	<i>phosphoinositide 5-phosphatase</i>	<i>Arabidopsis thaliana</i>	18400310	2E-41	I
MDI33_G10	1	<i>phosphoinositide 5-phosphatase</i>	<i>Arabidopsis thaliana</i>	18400310	2E-44	I
MDI24_D04	1	<i>phosphoinositide-specific phospholipase C-line</i>	<i>Arabidopsis thaliana</i>	8843789	2E-38	IT
MDI06-D09	1	<i>phospholipase C1</i>	<i>Nicotiana tabacum</i>	6969575	2E-67	I
CL288 contig1	2	<i>Phospholipase D alpha1 (Choline phosphatase 1)(Phosphatidylcholine-hydrolyzing phospholipase D1)</i>	<i>Nicotiana tabacum</i>	391461	8E-41	I
MDI05-A05	1	<i>phosphomannose isomerase</i>	<i>Cyamopsis tetragonoloba</i>	60617303	4E-44	G
MDI02-B07	1	<i>phosphoprotein phosphatase/ protein tyrosine/serine/threonine phosphatase</i>	<i>Arabidopsis thaliana</i>	15229482	1E-24	S
CL195 contig1	2	<i>phosphoprotein phosphatase/ protein tyrosine/serine/threonine phosphatase</i>	<i>Arabidopsis thaliana</i>	18403920	2E-59	S
MDI19_A03	1	<i>phosphoribosylamine-glycine ligase</i>	<i>Arabidopsis thaliana</i>	15218286	7E-25	S

MDI15_A04	1	<i>phosphoribosyl-ATP pyrophosphohydrolase At-IE, putative</i>	<i>Oryza sativa</i>	56783703	7E-36	S
CL413 contig1	2	<i>Photolyase/Blue-Light Receptor 2 - PHR2</i>	<i>Arabidopsis thaliana</i>	15226633	6E-54	S
MDI34_H01	1	<i>Photosystem I reaction center subunit II, chloroplast precursor (Photosystem I 20 kDa subunit) (PSI-D).</i>	<i>Spinacia oleracea</i>	19855891	4E-35	C
MDI44-E12	1	<i>photosystem I subunit V precursor, putative</i>	<i>Arabidopsis thaliana</i>	11908045	9E-59	C
CL415 contig1	2	<i>photosystem I subunit V precursor, putative</i>	<i>Arabidopsis thaliana</i>	12642864	2E-46	C
CL374 contig1	2	<i>photosystem II 10 kDa protein, chloroplast</i>	<i>Arachis hypogaea</i>	83776786	2E-55	C
MDI32-C11	1	<i>photosystem II stability/assembly factor HCF136, chloroplast, putative</i>	<i>Oryza sativa</i>	54291349	1E-19	C
MDI25_D09	1	<i>photosystem II type I chlorophyll a/b binding protein, putative</i>	<i>Arabidopsis thaliana</i>	12642854	2E-37	C
CL412 contig1	2	<i>photosystem I-N subunit</i>	<i>Phaseolus vulgaris</i>	28629385	3E-44	C
CL350 contig1	2	<i>PHR2 (Photolyase/Blue-Light Receptor 2)</i>	<i>Arabidopsis thaliana</i>	15226633	4E-55	S
MDI33_E03	1	<i>p-hydroxyphenylpyruvate dioxygenase</i>	<i>Arabidopsis thaliana</i>	3694811	3E-27	S
MDI27_H03	1	<i>phytoene desaturase</i>	<i>Momordica charantia var. abbreviata</i>	40456031	2E-53	S
MDI20_D02	1	<i>phytoene desaturase</i>	<i>Prunus armeniaca</i>	60672631	2E-74	S
MDI15_B09	1	<i>Phytosulfokine receptor precursor, putative</i>	<i>Oryza sativa</i>	51536231	5E-21	S
CL400 contig1	2	<i>PK12 protein kinase</i>	<i>Nicotiana tabacum</i>	2911280	6E-98	S
MDI24_F03	1	<i>plasma intrinsic protein 2,2</i>	<i>Juglans regia</i>	28395420	1E-66	S
MDI21_G10	1	<i>plasma membrane H⁺-ATPase</i>	<i>Sesbania rostrata</i>	31580857	5E-30	S
MDI48-B05	1	<i>plasma membrane intrinsic protein 2-1</i>	<i>Pyrus communis</i>	13486938	2E-58	S
MDI49-B02	1	<i>plasma membrane intrinsic protein, putative</i>	<i>Pisum sativum</i>	5139539	3E-39	S
CL104 contig1	3	<i>plasma membrane intrinsic protein, putative</i>	<i>Populus tremula x P. tremuloides</i>	57336888	4E-99	S
MDI33_A02	1	<i>plasma membrane intrinsic protein, putative</i>	<i>Artragalus membranaceus</i>	58700408	4E-34	S
MDI29_B10	1	<i>plastidic adenylate transporter, putative</i>	<i>Mesembryanthemum crystallinum</i>	61651614	2E-11	S
MDI23_C08	1	<i>plastidic ATP/ADP transporter</i>	<i>Citrus hybrid cultivar</i>	21063925	4E-81	S
MDI19_E10	1	<i>Plastocyanin</i>	<i>Lactuca sativa</i>	132070	2E-36	S
MDI22_D02	1	<i>pollen surface protein, putative</i>	<i>Arabidopsis thaliana</i>	21593810	2E-40	S
MDI19_D01	1	<i>poly(A) polymerase, putative</i>	<i>Arabidopsis thaliana</i>	26449498	1E-30	S
MDI21_F05	1	<i>poly(A) polymerase, putative</i>	<i>Oryza sativa</i>	53792977	8E-28	S
MDI09-H02	1	<i>poly(A)-binding protein</i>	<i>Cucumis sativus</i>	7528270	3E-33	S
CL449 contig1	2	<i>polygalacturonase isoenzyme 1 beta subunit homolog</i>	<i>Arabidopsis thaliana</i>	1762584	5E-73	G
MDI16_H04	1	<i>polygalacturonase, putative</i>	<i>Oryza sativa</i>	50932909	3E-57	G
MDI31_F12	1	<i>polyprotein, putative</i>	<i>Oryza sativa</i>	37534402	1E-58	S
MDI38_C07	1	<i>Polyubiquitin</i>	<i>Oryza sativa</i>	61223100	1E-22	S
MDI40_B02	1	<i>Pop3 protein, putative</i>	<i>Arabidopsis thaliana</i>	51316534	4E-27	S
CL275Contig1	2	<i>Porin</i>	<i>Prunus armeniaca</i>	5031279	1E-70	S
MDI05-E04	1	<i>potassium transporter HAK1p, putative</i>	<i>Mesembryanthemum crystallinum</i>	14091469	2E-57	S
MDI17_B11	1	<i>potyviral helper component protease-interacting protein 2, putative</i>	<i>Oryza sativa</i>	52076862	3E-17	S
MDI11-H08	1	<i>PP2A-3 (Protein Phosphatase 2A-3); protein phosphatase type 2A</i>	<i>Arabidopsis thaliana</i>	18410971	2E-78	TO

MDI15_A08	1	<i>PPI1 (Proton Pump Interactor 1)</i>	<i>Arabidopsis thaliana</i>	22328976	2E-12	S
MDI19_C11	1	<i>PR-10, similar</i>	<i>Vigna unguiculata</i>	7106538	3E-27	S
MDI49-A07	1	<i>prefoldin subunit 2, putative</i>	<i>Oryza sativa</i>	77555734	1E-16	S
MDI15_H02	1	<i>prefoldin-related KE2-like</i>	<i>Oryza sativa</i>	50908023	2E-44	S
MDI18_F08	1	<i>prenylated Rab receptor 2</i>	<i>Arabidopsis thaliana</i>	18072825	1E-36	S
MDI23_H07	1	<i>prenylated Rab receptor 2</i>	<i>Arabidopsis thaliana</i>	18072825	2E-21	S
MDI06-F09	1	<i>preprotein translocase, SecA subunit</i>	<i>Oryza sativa</i>	77549073	7E-39	S
MDI13_H06	1	<i>PRLI-interacting factor L-like</i>	<i>Oryza sativa</i>	51090705	3E-45	S
MDI20_H05	1	<i>probable oligopeptide transport permease protein AppC</i>	<i>Bradychizobium sp. BTai1</i>	78697945	2E-42	S
MDI24_D05	1	<i>Probable pectate lyase 8 precursor</i>	<i>Arabidopsis thaliana</i>	32129848	2E-46	
MDI04-E04	1	<i>prohibitin-like protein PHB1</i>	<i>Arabidopsis thaliana</i>	15235317	3E-36	O
MDI32-D03	1	<i>proline-rich protein</i>	<i>Glycine max</i>	8745402	4E-23	S
MDI06-D08	1	<i>proline-rich protein</i>	<i>Capsicum annuum</i>	3242079	5E-21	S
CL347 contig1	2	<i>prolyl 4-hydroxylase alpha subunit-like protein</i>	<i>Arabidopsis thaliana</i>	21618073	2E-59	S
MDI10-E02	1	<i>prolyl aminopeptidase-like protein</i>	<i>Arabidopsis thaliana</i>	15233141	5E-62	E
MDI13_D01	1	<i>proteasome 27 kDa subunit</i>	<i>Euphorbia esula</i>	6984140	3E-63	O
MDI52_C08	1	<i>proteasome protein-like</i>	<i>Oryza sativa</i>	50900108	4E-28	O
MDI30_D12	1	<i>proteasome regulatory subunit, putative</i>	<i>Arabidopsis thaliana</i>	15450839	1E-66	O
MDI19_C02	1	<i>Proteasome subunit alpha type 7 (20S proteasome alpha subunit D) (20S proteasome subunit alpha-4).</i>	<i>Cicer arietinum</i>	12229936	7E-76	O
MDI09-C07	1	<i>Proteasome subunit beta type 1 (20S proteasome alpha subunit F) (20S proteasome subunit beta-6)</i>	<i>Petunia x hybrida</i>	17380185	1E-88	O
MDI06-D04	1	<i>protein binding</i>	<i>Arabidopsis thaliana</i>	42568802	2E-48	S
MDI37_E07	1	<i>protein binding</i>	<i>Arabidopsis thaliana</i>	15219092	1E-21	S
MDI24_F04	1	<i>protein binding</i>	<i>Arabidopsis thaliana</i>	15229233	1E-41	S
MDI24_F07	1	<i>protein binding</i>	<i>Arabidopsis thaliana</i>	30687245	1E-64	S
MDI10-C03	1	<i>protein binding / ubiquitin-protein ligase/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	15231003	1E-28	S
CL152 contig1	3	<i>protein binding / ubiquitin-protein ligase/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	15241188	1E-19	S
MDI11_F11	1	<i>protein binding / ubiquitin-protein ligase/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	22331355	3E-35	S
MDI19_A07	1	<i>protein binding / ubiquitin-protein ligase/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	30687725	7E-52	S
MDI19_A01	1	<i>protein binding / ubiquitin-protein ligase/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	30686999	1E-24	S
MDI02-D01	1	<i>protein binding / zinc ion binding</i>	<i>Arabidopsis thaliana</i>	15238905	7E-69	S
MDI37_G01	1	<i>protein carrier</i>	<i>Arabidopsis thaliana</i>	42565125	1E-42	S
CL456 contig1	2	<i>protein disulfide isomerase</i>	<i>Quercus suber</i>	41393676	4E-81	O
MDI33_A10	1	<i>protein disulfide isomerase</i>	<i>Zea mays</i>	59861273	2E-34	O
MDI03-A04	1	<i>Protein disulfide isomerase precursor (PDI)</i>	<i>Datisca glomerata</i>	11133818	5E-91	O
MDI35_B11	1	<i>protein disulfide isomerase-related protein, putative</i>	<i>Oryza sativa</i>	51536089	6E-26	O
CL377Contig1	2	<i>protein kinase 1 - stpk1</i>	<i>Solanum tuberosum</i>	1200256	2E-54	TLDB
CL458 contig1	2	<i>protein kinase C binding / zinc ion binding [Arabidopsis thaliana]</i>	<i>Arabidopsis thaliana</i>	18410510	1E-56	TLDB

CL230 contig1	2	<i>protein kinase Ck2 regulatory subunit 2</i>	<i>Nicotiana tabacum</i>	21261784	4E-21	TLDB
MDI18_F01	1	<i>protein kinase Pti1</i>	<i>Glycine max</i>	29838544	4E-39	TLDB
MDI35_C01	1	<i>protein kinase, putative</i>	<i>Oryza sativa</i>	55769725	7E-50	TLDB
MDI42_A07	1	<i>protein kinase-like</i>	<i>Oryza sativa</i>	51535592	2E-30	TLDB
MDI19_D05	1	<i>protein MFP-b (Multifunctional Protein)</i>	<i>Cucumis sativus</i>	551444	8E-67	S
MDI29_C12	1	<i>protein phosphatase 2C (PP2C)</i>	<i>Fagus sylvatica</i>	7768153	2E-53	TO
MDI29_B01	1	<i>protein phosphatase type 2C, putative</i>	<i>Arabidopsis thaliana</i>	21539551	8E-48	TO
MDI13_C10	1	<i>protein phosphatase, putative</i>	<i>Arabidopsis thaliana</i>	4587597	6E-36	TO
MDI25_F04	1	<i>protein phosphatase, putative</i>	<i>Rubus idaeus</i>	6468695	1E-34	TO
CL280 contig1	2	<i>Protein TOC75, chloroplast precursor (75 kDa translocon at the outer-envelope-membrane of chloroplasts)</i>	<i>Oryza sativa</i>	82582304	1E-76	S
MDI19_H08	1	<i>protein transport protein SEC12p, putative</i>	<i>Arabidopsis thaliana</i>	17473926	7E-44	S
MDI34_C04	1	<i>protein transport protein Sec24-like, putative</i>	<i>Arabidopsis thaliana</i>	13124520	2E-24	S
CL462 contig1	2	<i>protein transport protein subunit - like, complete cds</i>	<i>Arabidopsis thaliana</i>	26451289	1E-21	S
MDI25_D03	1	<i>protein transporter</i>	<i>Arabidopsis thaliana</i>	42568712	2E-42	S
MDI29_C11	1	<i>protein-methionine-S-oxide reductase</i>	<i>Arabidopsis thaliana</i>	15227955	2E-10	S
MDI08-E02	1	<i>PRT1 (Proteolysis 1); ubiquitin-protein ligase</i>	<i>Arabidopsis thaliana</i>	15230192	5E-21	S
MDI23_C05	1	<i>PRT1 (Proteolysis 1); ubiquitin-protein ligase</i>	<i>Arabidopsis thaliana</i>	15230192	4E-36	S
MDI30_H02	1	<i>Pto-like kinase SG5-3d</i>	<i>Phaseolus vulgaris</i>	14010523	9E-28	S
MDI18_G06	1	<i>PTS protein, putative</i>	<i>Cicer arietinum</i>	7208777	1E-46	S
MDI25_E02	1	<i>pumilio/Mpt5 family RNA-binding protein, putative</i>	<i>Arabidopsis thaliana</i>	3980412	6E-32	S
MDI03-F11	1	<i>PUR ALPHA-1; nucleic acid binding</i>	<i>Arabidopsis thaliana</i>	30685174	4E-59	K
MDI02-A03	1	<i>purple acid phosphatase, putative</i>	<i>Oryza sativa</i>	34896986	9E-19	G
MDI07-H04	1	<i>purple acid phosphatase, putative</i>	<i>Arabidopsis thaliana</i>	49036419	5E-42	G
MDI25_H07	1	<i>putative ML domain protein</i>	<i>Nicotiana glauca</i>	55294668	7E-30	S
MDI04-G10	1	<i>PVR3-like protein</i>	<i>Ananas comosus</i>	25044847	6E-13	S
MDI26_D02	1	<i>pyridoxal-phosphate-dependent aminotransferase - like protein</i>	<i>Arabidopsis thaliana</i>	2827713	6E-75	E
MDI19_B06	1	<i>pyridoxine biosynthesis protein</i>	<i>Glycine max</i>	72256519	1E-52	S
MDI21_D12	1	<i>pyrophosphate-dependent phosphofructokinase alpha subunit</i>	<i>Citrus x paradisi</i>	3790102	2E-16	G
MDI26_G06	1	<i>pyrophosphate-dependent phosphofructokinase alpha subunit</i>	<i>Citrus x paradisi</i>	3790102	4E-48	G
MDI27_C09	1	<i>pyrophosphate-dependent phosphofructokinase beta subunit</i>	<i>Citrus x paradisi</i>	3790100	8E-57	G
MDI32-F10	1	<i>pyruvate decarboxylase 1</i>	<i>Lotus corniculatus</i>	29373077	2E-58	S
MDI04-B10	1	<i>pyruvate decarboxylase 1</i>	<i>Petunia x hybrida</i>	60656565	3E-84	S
CL221 contig1	2	<i>pyruvate kinase</i>	<i>Glycine max</i>	22296820	5E-91	GC
MDI18_G03	1	<i>pyruvate kinase</i>	<i>Citrus sinensis</i>	68138979	3E-66	GC
CL424 contig1	2	<i>pyruvate kinase-like</i>	<i>Deschampsia antarctica</i>	20562907	7E-43	GC
MDI38_G03	1	<i>QM family protein</i>	<i>Caragana jubata</i>	50313215	6E-86	S
MDI49-H10	1	<i>quercetin 3-O-glucoside-6''-O-malonyltransferase</i>	<i>Verbena x hybrida</i>	45789980	1E-15	S

MDI19_B09	1	Quinone oxidoreductase-like protein	<i>Arabidopsis thaliana</i>	21553644	3E-29	S
MDI33_A01	1	r40c1 protein, putative	<i>Oryza sativa</i>	50940085	1E-21	S
MDI35_F05	1	RAB6; GTP binding protein	<i>Arabidopsis thaliana</i>	15224916	4E-51	OU
CL34 contig1	6	raffinose synthase or seed imbibition protein, putative	<i>Oryza sativa</i>	50540754	1E-111	G
MDI05-B01	1	RAN	<i>Nicotiana sylvestris</i>	48249482	3E-58	UA
MDI06-C10	1	Ran binding protein-1	<i>Lycopersicon esculentum</i>	14091665	2E-39	UA
MDI05-G01	1	RAN BP1 (RAN Binding protein 1)	<i>Arabidopsis thaliana</i>	30697102	5E-40	UA
MDI51-F10	1	Ran GTPase binding	<i>Arabidopsis thaliana</i>	30686578	3E-88	YA
MDI09-F09	1	Ran GTPase binding	<i>Arabidopsis thaliana</i>	15223423	3E-16	YA
CL110 contig1	3	rapid alkalization factor 2	<i>Solanum chacoense</i>	37695573	1E-21	S
MDI01-G03	1	ras-GTPase-activating protein SH3-domain binding protein-like	<i>Arabidopsis thaliana</i>	21553535	5E-39	TV
MDI09-G10	1	ras-like small GTP binding ptoein, putative	<i>Zea mays</i>	46946656	3E-46	TD
CL307 contig1	2	Ras-related protein Rab7	<i>Prunus armeniaca</i>	3914550	2E-50	S
CL73 contig1	4	RD22-like protein (BURP domain-containing protein)	<i>Vitis vinifera</i>	54645911	1E-82	S
MDI23_B08	1	RD26; transcription factor	<i>Arabidopsis thaliana</i>	18416983	6E-81	K
MDI22_F09	1	receptor associated protein, putative	<i>Capsicum chinense</i>	61889383	2E-44	S
MDI17_E03	1	receptor protein kinase PERK1-like protein	<i>Oryza sativa</i>	50905775	7E-79	S
MDI24_A06	1	receptor protein kinase-like	<i>Oryza sativa</i>	57900294	1E-83	S
MDI40_C06	1	receptor-like protein CLAVATA2	<i>Arabidopsis thaliana</i>	6049568	7E-19	S
MDI12_E03	1	receptor-like protein kinase homolog RK20-1	<i>Phaseolus vulgaris</i>	4530126	7E-13	S
MDI30_A03	1	regulatory protein NPR1-like; transcription factor inhibitor I kappa B-like	<i>Arabidopsis thaliana</i>	9758986	3E-30	S
MDI14_F06	1	resistance protein, putative	<i>Lycopersicon esculentum</i>	16974114	2E-49	S
MDI48-D04	1	respiratory burst oxidase protein D (RbohD)	<i>Arabidopsis thaliana</i>	3242788	4E-21	S
MDI04-G02	1	retroelement pol polyprotein, putative	<i>Arabidopsis thaliana</i>	5306241	1E-13	S
MDI24_E03	1	retrotransposon protein, putative, unclassified	<i>Oryza sativa</i>	77554865	3E-27	S
MDI25_B03	1	RHO GDP-dissociation inhibitor 1, putative	<i>Arabidopsis thaliana</i>	6648200	2E-58	T
MDI18_B11	1	ribonuclease NGR2, putative	<i>Oryza sativa</i>	56785270	9E-28	L
MDI26_C01	1	ribonuclease PH	<i>Pseudomonas fluorescens Pf-5</i>	70733342	4E-43	L
MDI12_G07	1	ribonuclease/ transcriptional repressor	<i>Arabidopsis thaliana</i>	15220828	7E-20	K
MDI04-G03	1	ribophorin I homologue, putative	<i>Hordeum vulgare subsp. vulgare</i>	2894378	5E-77	S
MDI25_A03	1	ribosomal Pr 117	<i>Triticum aestivum</i>	31400871	2E-67	J
MDI39_D04	1	ribosomal protein L13a like protein	<i>Arabidopsis thaliana</i>	7268019	5E-93	J
MDI37_G04	1	ribosomal protein L2	<i>Arabidopsis thaliana</i>	3036817	5E-96	J
MDI04-F10	1	ribosomal protein L22	<i>Eucalyptus globulus subsp. Globulus</i>	60460846	8E-31	J
CL321 contig1	2	ribosomal protein L27 homolog	<i>Pisum sativum</i>	498906	1E-63	J
MDI08-D03	1	ribosomal protein L3	<i>Lycopersicon esculentum</i>	38327504	2E-63	J
MDI26_B11	1	ribosomal protein L3, putative	<i>Oryza sativa</i>	62734368	2E-74	J

CL200Contig1	2	ribosomal protein L32, putative	<i>Oryza sativa</i>	50947773	2E-35	J
MDI07-C12	1	ribosomal protein L34 precursor, putative plastid	<i>Arabidopsis thaliana</i>	15218810	8E-12	J
MDI19_G04	1	ribosomal protein L37	<i>Glycine max</i>	44662866	1E-09	J
MDI26_A03	1	Ribosomal protein L7Ae-like	<i>Arabidopsis thaliana</i>	21592866	5E-45	J
MDI05-F06	1	ribosomal protein L9, putative	<i>Arabidopsis thaliana</i>	21554945	1E-65	J
CL259 contig1	2	ribosomal protein S10, putative	<i>Arabidopsis thaliana</i>	21617950	1E-36	J
MDI34_D01	1	ribosomal protein S12, putative	<i>Oryza sativa</i>	50934241	3E-26	J
MDI44-E10	1	ribosomal protein S13	<i>Glycine max</i>	44662862	4E-70	J
MDI07-D07	1	ribosomal protein S14, putative	<i>Oryza sativa</i>	50905421	6E-57	J
MDI07-D09	1	ribosomal protein S19	<i>Petunia sp.</i>	14200	1E-28	J
MDI18_E05	1	ribosomal protein S26	<i>Pisum sativum</i>	5706704	5E-21	J
MDI35_G03	1	ribosomal protein S26	<i>Pisum sativum</i>	5706704	1E-38	J
CL132Contig1	3	Ribosomal protein S27 - RS27A; structural constituent of ribosome	<i>Arabidopsis thaliana</i>	15233044	5E-32	J
MDI33_H02	1	ribosomal protein S3a	<i>Cicer arietinum</i>	24817254	5E-60	J
MDI25_B05	1	ribosomal protein S4	<i>Oryza sativa</i>	77551805	4E-26	J
MDI33_C05	1	ribosomal protein S6 plastid, putative	<i>Arabidopsis thaliana</i>	21553658	1E-33	J
CL95 contig1	2	ribosomal protein S7 (rps7) and ribosomal protein S12 (rps12)	<i>Lactuca sativa</i>	27728739	4E-69	J
MDI12_D07	1	ribosomal protein S7 (rps7) and ribosomal protein S12 (rps12)	<i>Lactuca sativa</i>	27728739	5E-36	J
CL99 contig1	3	ribosomal protein, putative	<i>Oryza sativa</i>	50901182	8E-71	J
MDI23_C09	1	Ribosomal protein S18 - RPS18C ; RNA binding / structural constituent of ribosome	<i>Arabidopsis thaliana</i>	15234790	7E-65	J
MDI16_A10	1	ribulose-1,5-bisphosphate carboxylase/oxygenase activase	<i>Malus x domestica</i>	415852	1E-53	C
CL26 contig1	6	ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit	<i>Panax ginseng</i>	77157637	3E-74	C
MDI21_B11	1	ribulose-phosphate 3-epimerase	<i>Spinacia oleracea</i>	3264788	7E-27	GC
MDI19_C10	1	Rieske iron-sulfur protein Tic55	<i>Pisum sativum</i>	2764524	2E-66	S
MDI05-B08	1	RIN4 (RPM1 Interacting Protein 4)	<i>Arabidopsis thaliana</i>	30687738	5E-15	S
CL224 contig1	2	ring zinc finger protein	<i>Artemisia desertorum</i>	62869545	1E-14	K
MDI01-D04	1	RING zinc finger protein-like	<i>Arabidopsis thaliana</i>	11994767	4E-17	K
MDI30_B08	1	ring zinc finger protein-like protein	<i>Oryza sativa</i>	29367335	5E-55	K
CL394 contig1	2	ring-box 1 protein-like	<i>Arabidopsis thaliana</i>	21389679	4E-50	K
MDI16_H10	1	RING-H2 Finger Protein 2B - RHA2B; protein binding / zinc ion binding	<i>Arabidopsis thaliana</i>	18379162	3E-29	K
MDI20_C01	1	RING-H2 finger protein RHG1a, putative	<i>Oryza sativa</i>	50948261	2E-16	K
MDI31_B08	1	ring-H2 finger protein, putative	<i>Oryza sativa</i>	52353630	1E-11	K
MDI03-D11	1	ripening-related bZIP protein, putative	<i>Vitis vinifera</i>	7406677	1E-39	K
MDI34_C12	1	ripening-related pectate lyase	<i>Mangifera indica</i>	62550728	2E-70	S
CL80 contig1	4	ripening-related protein, putative	<i>Vitis vinifera</i>	7406663	6E-44	S
MDI26_G05	1	ripening-related protein, putative	<i>Vitis vinifera</i>	7406669	1E-44	S
MDI36_G05	1	ripening-related protein-like	<i>Arabidopsis thaliana</i>	21553812	2E-34	S

MDI30_H12	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	15237311	4E-18	R
MDI07-E01	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	18409491	5E-37	R
MDI26_F02	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	18409491	3E-31	R
MDI15_B04	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	18418072	4E-65	R
MDI18_E11	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	30694922	2E-34	R
MDI40_F06	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	15231284	9E-87	R
MDI26_H08	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	15232459	6E-11	R
MDI13_D12	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	18423239	1E-20	R
MDI05-D10	1	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	22329999	4E-61	R
CL390 contig1	2	RNA binding / nucleic acid binding	<i>Arabidopsis thaliana</i>	42573682	1E-22	R
MDI21_E05	1	RNA binding / nucleic acid binding / protein transporter	<i>Arabidopsis thaliana</i>	15240057	4E-12	R
MDI16_F05	1	RNA binding / structural constituent of ribosome	<i>Arabidopsis thaliana</i>	15232515	7E-28	J
MDI17_A03	1	RNA binding protein, putative	<i>Oryza sativa</i>	50919311	2E-54	O
MDI38_H06	1	RNA helicase	<i>Arabidopsis thaliana</i>	3775993	4E-45	S
MDI05-C10	1	RNA helicase	<i>Vigna radiata</i>	7211427	9E-70	S
MDI35_C09	1	RNA polymerase II transcription factor	<i>Arabidopsis thaliana</i>	15233829	3E-43	K
MDI09-E09	1	RNA-binding protein-like protein	<i>Arabidopsis thaliana</i>	21618121	2E-62	O
MDI06-H07	1	RNase H domain-containing protein-like	<i>Oryza sativa</i>	51535805	4E-12	S
MDI32-F08	1	RNase H, putative	<i>Arabidopsis thaliana</i>	6692261	1E-12	S
MDI11_E03	1	Rop4 small GTP binding protein	<i>Zea mays</i>	4959465	2E-43	T
MDI08-E08	1	RPS15AB; structural constituent of ribosome	<i>Arabidopsis thaliana</i>	15224834	3E-55	J
MDI12_G06	1	RSW3 (Radial Swelling 3); hydrolase, hydrolyzing O-glycosyl compounds	<i>Arabidopsis thaliana</i>	15237538	1E-61	S
MDI42_C03	1	RUB1 conjugating enzyme	<i>Arabidopsis thaliana</i>	6635457	3E-79	S
MDI03-A01	1	Rubisco activase 2	<i>Mesembryanthemum crystallinum</i>	71834884	4E-65	S
CL327 contig1	2	S6 ribosomal protein	<i>Asparagus officinalis</i>	7649387	1E-117	S
MDI40_D03	1	S-adenosyl methionine synthase-like	<i>Solanum tuberosum</i>	78191442	5E-52	H
MDI04-F09	1	S-adenosyl-L-methionine synthetase	<i>Elaeagnus umbellata</i>	13540318	3E-74	H
MDI01-G09	1	S-adenosyl-L-methionine synthetase 1	<i>Catharanthus roseus</i>	1555576	2E-77	H
MDI48-E10	1	S-adenosyl-L-methionine synthetase 1	<i>Catharanthus roseus</i>	1655576	1E-70	H
MDI29_F06	1	S-adenosyl-L-methionine:delta24-sterol-C-methyltransferase	<i>Glycine max</i>	1399380	2E-23	H
MDI16_G01	1	S-adenosylmethionine decarboxylase proenzyme	<i>Ipomoea nil</i>	3913439	2E-46	H
MDI47-B10	1	S-adenosylmethionine synthetase-2	<i>Pisum sativum</i>	37051117	2E-35	H
CL242 contig1	2	S-adenosylmethionine-dependent methyltransferase/ catalytic	<i>Arabidopsis thaliana</i>	42565060	4E-44	H
MDI49-G08	1	S-adenosylmethionine-dependent methyltransferase/ phosphoethanolamine N-methyltransferase	<i>Arabidopsis thaliana</i>	42572097	5E-47	H
MDI12_D11	1	S-adenosyl-methionine-sterol-C-methyltransferase, 5' partial; 1-1344, putative	<i>Arabidopsis thaliana</i>	12324209	4E-28	H
MDI26_C10	1	salt tolerance protein	<i>Sesuvium portulacastrum</i>	56112332	3E-49	S

CL116 contig1	2	<i>salt tolerance protein</i>	<i>Sesuvium portulacastrum</i>	56112332	2E-95	S
MDI29_H03	1	<i>salt tolerance protein 5, putative</i>	<i>Oryza sativa</i>	51535412	1E-30	S
CL338 contig1	2	<i>salt-tolerance protein</i>	<i>Glycine max</i>	78173056	6E-61	S
MDI02-B08	1	<i>SAR DNA binding protein</i>	<i>Oryza sativa</i>	3288883	6E-32	S
MDI46-E03	1	<i>SAR1/GTP-binding secretory factor</i>	<i>Arabidopsis thaliana</i>	2104532	4E-24	S
MDI52_D06	1	<i>Scarecrow transcriptional regulator-like</i>	<i>Arabidopsis thaliana</i>	22136380	1E-23	K
MDI51-E12	1	<i>Scarecrow-like protein</i>	<i>Lilium longiflorum</i>	32813435	4E-13	K
MDI20_B04	1	<i>SCPL27; catalytic/serine carboxypeptidase</i>	<i>Arabidopsis thaliana</i>	15231911	2E-109	S
MDI05-B02	1	<i>SDL-1 protein</i>	<i>Nicotiana glauca</i>	18307504	5E-85	S
CL464 contig1	2	<i>secretory carrier membrane protein, putative</i>	<i>Oryza sativa</i>	50911775	9E-97	U
MDI42_F01	1	<i>secretory laccase</i>	<i>Gossypium arboreum</i>	40218371	9E-21	U
MDI29_H01	1	<i>sedoheptulose-1,7-bisphosphatase; SBPase</i>	<i>Arabidopsis thaliana</i>	786466	3E-41	S
CL313 contig1	3	<i>Seed maturation protein PM23</i>	<i>Glycine max</i>	6648962	1E-30	S
CL71 contig1	4	<i>Seed maturation protein PM37</i>	<i>Glycine max</i>	5802244	3E-138	S
MDI52_C05	1	<i>selenium binding</i>	<i>Arabidopsis thaliana</i>	18390562	1E-63	P
MDI34_G08	1	<i>Senescence-associated gene 21 - SAG21</i>	<i>Arabidopsis thaliana</i>	18411926	5E-20	S
CL40 contig2	3	<i>senescence-associated protein 5</i>	<i>Hemerocallis hybrid cultivar</i>	3551954	4E-27	S
CL216Contig1	2	<i>senescence-associated protein 5, putative</i>	<i>Arabidopsis thaliana</i>	4191796	6E-49	S
MDI07-B09	1	<i>senescence-associated protein DH</i>	<i>Zea mays</i>	54208706	1E-31	S
MDI52_G05	1	<i>senescence-associated protein SAG102, putative</i>	<i>Malus x domestica</i>	33772238	2E-12	S
CL59 contig1	4	<i>senescence-associated protein, putative</i>	<i>Pisum sativum</i>	13359451	1E-66	S
CL216 contig1	2	<i>senescence-associated protein, putative</i>	<i>Oryza sativa</i>	54291245	6E-45	S
MDI46-C07	1	<i>senescence-associated protein, similar</i>	<i>Arabidopsis thaliana</i>	51968454	9E-41	S
CL40 contig1	2	<i>senescence-associated protein, similar to</i>	<i>Arabidopsis thaliana</i>	51968454	1E-56	S
MDI05-C07	1	<i>senescence-associated protein-like</i>	<i>Oryza sativa</i>	54291569	1E-12	S
MDI25_D08	1	<i>senescence-associated protein-like</i>	<i>Oryza sativa</i>	54291569	3E-16	S
CL326Contig1	2	<i>serine /threonine kinase similar to NAK, putative</i>	<i>Oryza sativa</i>	37534182	5E-51	S
MDI36_E02	1	<i>serine decarboxylase, putative</i>	<i>Oryza sativa</i>	50252214	2E-37	S
MDI07-F12	1	<i>Serine Hydromethyltransferase 4 - SHM4; glycine hydroxymethyltransferase</i>	<i>Arabidopsis thaliana</i>	15236375	2E-84	S
MDI20_C02	1	<i>serine O-acetyltransferase 1</i>	<i>Glycine max</i>	39841350	2E-45	C
MDI33_D02	1	<i>serine protease, putative</i>	<i>Populus euramericana</i>	29786399	1E-13	S
MDI51-E06	1	<i>serine protease, putative</i>	<i>Oryza sativa</i>	50918929	5E-51	S
MDI34_C01	1	<i>serine/threonine kinase, putative</i>	<i>Vitis vinifera</i>	28629126	1E-32	TDV
MDI01-F05	1	<i>serine/threonine kinases, similar</i>	<i>Arabidopsis thaliana</i>	6056375	2E-60	TDV
MDI15_B08	1	<i>serine/threonine protein kinase, putative</i>	<i>Oryza sativa</i>	52077183	3E-24	TDV
MDI15_D07	1	<i>serine/threonine protein kinase, putative</i>	<i>Mesembryanthemum crystallinum</i>	4884974	4E-32	TDV

MDI30_F11	1	<i>serine/threonine protein kinase-like</i>	<i>Solanum tuberosum</i>	53689728	8E-31	TDV
MDI34_A11	1	<i>serine/threonine-specific protein kinase, putative</i>	<i>Oryza sativa</i>	50253149	1E-40	S
MDI51-B06	1	<i>serine/threonine-specific protein kinase, putative</i>	<i>Oryza sativa</i>	50940951	2E-18	S
MDI19_C12	1	<i>Serk2 protein</i>	<i>Zea mays</i>	13897310	1E-14	S
MDI45-G03	1	<i>shaggy related kinase</i>	<i>Arabidopsis thaliana</i>	460832	1E-23	S
MDI32-E03	1	<i>shikimate kinase, putative</i>	<i>Fagus sylvatica</i>	76782196	2E-25	S
MDI17_H07	1	<i>short-chain alcohol dehydrogenase A</i>	<i>Pisum sativum</i>	37051111	1E-48	S
MDI47-B08	1	<i>sigma-54 dependent transcriptional regulator RtcR</i>	<i>Pseudomonas fluorescens Pf-5</i>	70729670	3E-34	S
MDI11_C11	1	<i>signal peptide peptidase, putative</i>	<i>Oryza sativa</i>	77550327	9E-51	U
MDI08-D02	1	<i>signal recognition particle 72, putative</i>	<i>Oryza sativa</i>	56784353	2E-61	U
MDI15_H03	1	<i>single myb histone 6</i>	<i>Zea mays</i>	34105723	1E-17	K
MDI17_C10	1	<i>SKS17; copper ion binding</i>	<i>Arabidopsis thaliana</i>	18425129	6E-68	P
MDI32-B12	1	<i>SKS17; copper ion binding</i>	<i>Arabidopsis thaliana</i>	18425129	5E-24	P
MDI22_B08	1	<i>small GTP-binding protein</i>	<i>Nicotiana glauca</i>	1053063	6E-78	T
CL196 contig1	2	<i>small GTP-binding protein</i>	<i>Nicotiana glauca</i>	1616614	2E-66	T
CL281 contig1	2	<i>small rubber particle protein</i>	<i>Hevea brasiliensis</i>	37622210	1E-57	S
MDI19_H02	1	<i>small zinc finger-like protein</i>	<i>Mesembryanthemum crystallinum</i>	5107212	1E-30	K
CL360 contig1	2	<i>snakin-1</i>	<i>Solanum tuberosum</i>	14627134	3E-18	E
CL153 contig1	3	<i>snakin2</i>	<i>Solanum tuberosum</i>	14625945	7E-21	E
MDI20_E05	1	<i>SNF1-related protein kinase</i>	<i>Lycopersicon esculentum</i>	37901484	1E-51	KB
CL197 contig1	2	<i>SNF1-related protein kinase</i>	<i>Cucumis sativus</i>	1743009	0.0	KB
CL9 contig2	2	<i>S-norcochlorine synthase 1</i>	<i>Papaver somniferum</i>	61814642	1E-17	S
MDI17_E11	1	<i>sodium:hydrogen antiporter/ solute:hydrogen antiporter</i>	<i>Arabidopsis thaliana</i>	22330742	4E-55	P
MDI08-F04	1	<i>sorbitol dehydrogenase-like protein</i>	<i>Arabidopsis thaliana</i>	21553353	4E-12	G
MDI06-A01	1	<i>sorbitol transporter</i>	<i>Malus x domestica</i>	51849623	3E-71	S
MDI20_D11	1	<i>Sorting nexin 1</i>	<i>Brassica oleracea</i>	37936234	1E-29	UZV
MDI13_E05	1	<i>SOS2-like protein kinase</i>	<i>Glycine max</i>	21954717	2E-56	S
CL380 contig1	2	<i>Soul Protein, similar (heme-binding protein)</i>	<i>Arabidopsis thaliana</i>	5734756	2E-56	S
CL29 contig1	6	<i>Spermidine synthase (Putrescine aminopropyltransferase) (SPDSY)</i>	<i>Coffea arabica</i>	6094335	1E-109	E
MDI36_B03	1	<i>sphingosine-1-phosphate phosphatase</i>	<i>Arabidopsis thaliana</i>	78126073	5E-29	S
MDI25_D11	1	<i>spliceosomal protein</i>	<i>Solanum tuberosum</i>	169589	7E-49	S
MDI25_D06	1	<i>splicing factor 3B subunit 5-like protein</i>	<i>Oryza sativa</i>	50946241	8E-35	S
CL79 contig1	4	<i>squalene monoxygenase</i>	<i>Datura innoxia</i>	62945915	2E-86	I
MDI40_E07	1	<i>squalene synthase</i>	<i>Centella asiatica</i>	55710094	3E-79	I
CL182 contig1	2	<i>SSR alpha subunit, putative</i>	<i>Oryza sativa</i>	53791783	2E-71	S
MDI07-B11	1	<i>Steroid 22-alpha-hydroxylase; DWF4; CYP90B1</i>	<i>Arabidopsis thaliana</i>	2935342	3E-78	I
CL325 contig1	2	<i>Steroid membrane binding protein, putative</i>	<i>Oryza sativa</i>	1432912	5E-69	I

MDI33_F02	1	<i>sterol 4-alpha-methyl-oxidase, putative</i>	<i>Gossypium arboreum</i>	27448145	1E-70	I
MDI22_H04	1	<i>Sterol-4alpha-methyl oxidase 1-1 - SMO1-1 ; catalytic</i>	<i>Arabidopsis thaliana</i>	15234416	2E-35	I
MDI08-A11	1	<i>stpk1 protein kinase</i>	<i>Solanum tuberosum</i>	1200256	2E-30	TLDB
MDI36_H01	1	<i>stress related protein, putative</i>	<i>Arabidopsis thaliana</i>	22530918	1E-18	S
CL276 contig1	2	<i>striated muscle activator-like protein</i>	<i>Oryza sativa</i>	50838962	4E-18	S
CL283 contig1	2	<i>strictosidine synthase</i>	<i>Arabidopsis thaliana</i>	30694556	3E-24	S
MDI25_E11	1	<i>Stromal cell-derived factor 2-like protein</i>	<i>Oryza sativa</i>	50943411	2E-28	S
MDI21_H03	1	<i>structural constituent of nuclear pore</i>	<i>Arabidopsis thaliana</i>	30689895	2E-30	J
MDI35_B01	1	<i>structural constituent of ribosome</i>	<i>Arabidopsis thaliana</i>	15241026	3E-28	J
MDI07-A03	1	<i>structural constituent of ribosome</i>	<i>Arabidopsis thaliana</i>	30690926	1E-14	J
MDI08-D10	1	<i>structural maintenance of chromosomes 1 protein</i>	<i>Arabidopsis thaliana</i>	45594277	7E-46	S
MDI43_A04	1	<i>structural molecule</i>	<i>Arabidopsis thaliana</i>	18407205	2E-52	S
MDI21_E09	1	<i>Stylosa protein</i>	<i>Antirrhinum majus</i>	49523567	3E-87	S
CL342 contig1	2	<i>Subtilase</i>	<i>Casuarina glauca</i>	29028287	9E-60	O
MDI15_F09	1	<i>subtilisin proteinase-like</i>	<i>Arabidopsis thaliana</i>	18418552	4E-55	O
MDI12_H12	1	<i>subtilisin-like proteinase, putative</i>	<i>Oryza sativa</i>	50931541	3E-10	O
CL131Contig1	3	<i>succinate dehydrogenase iron-sulphur subunit, mitochondrial</i>	<i>Arabidopsis thaliana</i>	12049600	2E-17	G
MDI11_C01	1	<i>succinyl-CoA ligase alpha 2 subunit</i>	<i>Lycopersicon esculentum</i>	49617539	1E-37	G
MDI12_G10	1	<i>succinyl-CoA ligase alpha subunit, putative</i>	<i>Oryza sativa</i>	50938629	8E-68	G
MDI41_H04	1	<i>sucrose synthase</i>	<i>Citrus urshiu</i>	6682843	2E-52	G
MDI03-G11	1	<i>Sucrose-phosphate synthase 1 (UDP-glucose-fructose-phosphate glucosyltransferase 1)</i>	<i>Citrus unshiu</i>	3915023	5E-58	G
MDI30_E09	1	<i>Sugar carrier protein C</i>	<i>Ricinus communis</i>	3915039	3E-46	S
MDI20_C05	1	<i>sugar porter/ transporter</i>	<i>Arabidopsis thaliana</i>	18408421	4E-60	G
MDI05-F01	1	<i>SUI1 protein</i>	<i>Coffea arabica</i>	28193048	6E-15	S
MDI20_E09	1	<i>sulfite reductase</i>	<i>Nicotiana tabacum</i>	3738234	7E-43	P
MDI29_H12	1	<i>SULTR4;2; sulfate transporter</i>	<i>Arabidopsis thaliana</i>	15230510	7E-16	P
MDI29_F01	1	<i>superoxide dismutase</i>	<i>Fagus sylvatica</i>	38228697	2E-53	P
CL83 contig1	4	<i>superoxide dismutase [Fe]</i>	<i>Lycopersicon esculentum</i>	33413303	3E-51	P
CL93 contig1	3	<i>suppressor of kinetochore protein 1 - Skp1</i>	<i>Medicago sativa</i>	4959710	2E-57	S
MDI30_H09	1	<i>Supressor of fatty acid desaturase deficiency 1 - SFD1; glycerol-3-phosphate dehydrogenase (NAD+)/ oxidoreductase, acting on CH-OH group of donors</i>	<i>Arabidopsis thaliana</i>	18405476	3E-30	I
MDI43_D06	1	<i>synaptobrevin 1</i>	<i>Oryza sativa</i>	60543737	3E-73	U
MDI13_F01	1	<i>synaptobrevin 1</i>	<i>Oryza sativa</i>	60543737	3E-52	U
MDI41_F07	1	<i>Syntaxin</i>	<i>Glycine max</i>	22597174	1E-58	S
MDI33_B10	1	<i>syntaxin of plants</i>	<i>Medicago sativa</i>	71534985	1E-36	S
MDI33_F03	1	<i>syringolide-induced protein 19-1-5</i>	<i>Glycine max</i>	19911573	1E-64	S
MDI52_E02	1	<i>syringolide-induced protein 19-1-5</i>	<i>Glycine max</i>	19911573	1E-26	S

MDI03-C01	1	<i>T7N9 14 protein, putative</i>	<i>Oryza sativa</i>	77556605	2E-25	S
MDI32-F01	1	<i>TAFII21 (TATA Binding protein associated factor 21KDA subunit); DNA binding / transcription initiation factor</i>	<i>Arabidopsis thaliana</i>	15221048	4E-31	K
MDI44-A05	1	<i>TATA box-binding protein associated factor 10</i>	<i>Flaveria trinervia</i>	41529318	2E-40	K
MDI07-D11	1	<i>TATA-box binding protein</i>	<i>Phaseolus vulgaris</i>	4102725	2E-54	K
MDI10-G01	1	<i>tbcl domain family protein, putative</i>	<i>Oryza sativa</i>	50726241	1E-50	S
MDI29_A04	1	<i>temperature-induced lipocalin</i>	<i>Populus balsamifera</i>	77744871	3E-63	S
MDI05-H09	1	<i>terpene synthase</i>	<i>Vitis vinifera</i>	45478053	1E-51	S
MDI01-C10	1	<i>tethering factor SEC34, putative</i>	<i>Oryza sativa</i>	50946351	7E-17	S
MDI47-A09	1	<i>TGB12K interacting protein 3</i>	<i>Nicotiana tabacum</i>	29826244	5E-28	S
CL30 contig1	6	<i>thaumatin-like protein</i>	<i>Actinidia deliciosa</i>	71057064	3E-96	S
MDI09-E04	1	<i>thaumatin-like protein</i>	<i>Arabidopsis thaliana</i>	7269268	1E-43	S
CL352 contig1	2	<i>Thiazole biosynthetic enzyme, chloroplast precursor</i>	<i>Citrus sinensis</i>	6094476	1E-80	S
MDI46-A08	1	<i>Thiazole biosynthetic enzyme, chloroplast precursor</i>	<i>Citrus sinensis</i>	6094476	7E-51	S
MDI07-H08	1	<i>thioesterase type B</i>	<i>Cuphea lanceolata</i>	12057063	1E-17	S
MDI04-F11	1	<i>thioredoxin</i>	<i>Ricinus communis</i>	1255954	3E-34	O
MDI05-B03	1	<i>Thioredoxin domain 2</i>	<i>Anabaena variabilis ATCC 29413</i>	75700587	6E-12	O
CL346 contig1	2	<i>thioredoxin H</i>	<i>Populus tremula x P. tremuloides</i>	198519	1E-37	O
MDI42_A08	1	<i>thioredoxin m2, putative</i>	<i>Pisum sativum</i>	15594012	4E-22	O
CL74 contig1	4	<i>thioredoxin peroxidase</i>	<i>Capsicum annuum</i>	18654477	1E-71	O
MDI49-A01	1	<i>thioredoxin peroxidase, putative</i>	<i>Oryza sativa</i>	50251981	4E-74	O
CL300 contig1	2	<i>thioredoxin-like U5 small ribonucleoprotein particle protein, putative</i>	<i>Oryza sativa</i>	31432762	5E-70	O
MDI33_F12	1	<i>thylakoid luminal protein-like</i>	<i>Oryza sativa</i>	50911146	2E-18	S
MDI21_H07	1	<i>Thylakoid membrane phosphoprotein of 14 KDA - TMP14</i>	<i>Arabidopsis thaliana</i>	18407178	1E-20	S
MDI17_H05	1	<i>TIR-NBS disease resistance-like protein</i>	<i>Populus trichocarpa</i>	82542039	6E-30	S
MDI12_E10	1	<i>tonoplast intrinsic protein</i>	<i>Kandelia candel</i>	23957320	1E-36	S
MDI52_G06	1	<i>tonoplast intrinsic protein</i>	<i>Kandelia candel</i>	23957320	6E-65	S
MDI11_G03	1	<i>tonoplast intrinsic protein</i>	<i>Prunus persica</i>	38678108	1E-41	S
MDI01-D10	1	<i>topoisomerase I</i>	<i>Pisum sativum</i>	2330649	1E-37	L
MDI42_G08	1	<i>transcription factor</i>	<i>Arabidopsis thaliana</i>	15222433	6E-10	K
MDI26_F09	1	<i>transcription factor</i>	<i>Arabidopsis thaliana</i>	18395292	2E-12	K
MDI43-E10	1	<i>transcription factor</i>	<i>Arabidopsis thaliana</i>	15226170	1E-57	K
CL165 contig1	3	<i>transcription factor</i>	<i>Arabidopsis thaliana</i>	15234741	1E-21	K
MDI01-H03	1	<i>transcription factor</i>	<i>Arabidopsis thaliana</i>	18399836	1E-13	K
MDI15_F07	1	<i>transcription factor 16</i>	<i>Arabidopsis thaliana</i>	71660789	5E-25	K
MDI20_G02	1	<i>transcription factor, putative</i>	<i>Arabidopsis thaliana</i>	11079528	1E-54	K
MDI05-G07	1	<i>transcription factor, putative</i>	<i>Oryza sativa</i>	31432723	1E-49	K
MDI05-F11	1	<i>transcription factor, putative</i>	<i>Vitis vinifera</i>	14582465	5E-46	K

MDI32-H09	1	<i>transcription factor/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	15237049	2E-21	K
MDI15_D02	1	<i>transcription factor/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	30689668	1E-39	K
MDI22_G09	1	<i>transcription regulator</i>	<i>Arabidopsis thaliana</i>	18397316	4E-53	K
MDI19_F10	1	<i>Transcriptional Regulator, AraC family</i>	<i>Pseudomonas fluorescens PfO-1</i>	77458690	4E-66	K
MDI20_A06	1	<i>Transcriptional Regulator, GntR family</i>	<i>Pseudomonas fluorescens PfO-1</i>	77457234	7E-42	K
MDI21_B10	1	<i>transducin/WD-40 repeat protein</i>	<i>Oryza sativa</i>	40218073	9E-54	K
MDI29_E03	1	<i>transferase, transferring glycosyl groups</i>	<i>Arabidopsis thaliana</i>	15233046	6E-26	S
MDI46-C06	1	<i>transferase, transferring glycosyl groups</i>	<i>Arabidopsis thaliana</i>	30697244	1E-32	S
MDI22_F03	1	<i>transferase, transferring glycosyl groups / transferase, transferring hexosyl groups</i>	<i>Arabidopsis thaliana</i>	30685369	1E-32	S
MDI03-H06	1	<i>transferase, transferring glycosyl groups / transferase, transferring hexosyl groups</i>	<i>Arabidopsis thaliana</i>	79562781	1E-49	S
CL24 contig1	5	<i>translation elongation factor 1A-9</i>	<i>Gossypium hirsutum</i>	74486744	0.0	J
CL24 contig2	2	<i>translation elongation factor 1A-9</i>	<i>Gossypium hirsutum</i>	74486744	5E-125	J
CL32 contig1	6	<i>translation elongation factor 1A-9</i>	<i>Gossypium hirsutum</i>	74486744	3E-166	J
MDI26_D04	1	<i>translation elongation factor 1-gamma</i>	<i>Prunus avium</i>	10444415	2E-27	J
CL163 contig1	3	<i>translation factor, putative</i>	<i>Pinus pinaster</i>	20218809	1E-48	J
MDI01-D07	1	<i>translation initiation factor</i>	<i>Pisum sativum</i>	23452071	4E-80	J
CL143 contig1	3	<i>translation initiation factor</i>	<i>Arabidopsis thaliana</i>	18421544	3E-73	J
MDI24_F09	1	<i>translation initiation factor</i>	<i>Arabidopsis thaliana</i>	22326739	2E-74	J
MDI35_G09	1	<i>translation initiation factor (eIF-4A)</i>	<i>Nicotiana tabacum</i>	475216	6E-41	J
CL76 contig1	4	<i>translation initiation factor 5A isoform VIII, eukaryotic</i>	<i>Hevea brasiliensis</i>	33325131	3E-80	J
CL173 contig1	3	<i>translation initiation factor eIF-1A, putative</i>	<i>Arabidopsis thaliana</i>	4587600	7E-47	J
CL54 contig1	4	<i>transmembrane protein, putative</i>	<i>Oryza sativa</i>	50933293	6E-39	S
MDI29_B03	1	<i>transport inhibitor response-like protein</i>	<i>Arabidopsis thaliana</i>	17064984	5E-39	S
MDI22_C09	1	<i>Transporter</i>	<i>Arabidopsis thaliana</i>	18395980	3E-82	S
CL208Contig1	2	<i>transporter-like protein</i>	<i>Arabidopsis thaliana</i>	15222153	2E-115	S
MDI21_E08	1	<i>trehalose-6-phosphate phosphatase</i>	<i>Arabidopsis thaliana</i>	20260496	4E-11	G
MDI32-G05	1	<i>triacylglycerol lipase</i>	<i>Arabidopsis thaliana</i>	15238889	2E-19	I
CL141 contig1	3	<i>triosephosphate isomerase 1</i>	<i>Zea mays</i>	168647	4E-82	G
MDI29_D12	1	<i>truncated acetyl Co-A acetyltransferase-like protein</i>	<i>Hevea brasiliensis</i>	16417952	6E-50	S
CL358 contig1	2	<i>truncated copper-binding protein CUTA (CUTA)</i>	<i>Arabidopsis thaliana</i>	51971807	6E-58	S
MDI22_A04	1	<i>tryptophan synthase beta subunit</i>	<i>Nostoc sp</i>	17129755	2E-16	EC
MDI08-E07	1	<i>Tsi1-interacting protein TSIP1 (TSIP1)</i>	<i>Nicotiana tabacum</i>	4337000	4E-51	S
MDI37_F08	1	<i>TUB1 (Tubulin beta-1); GTP binding / GTPase/ structural molecule</i>	<i>Arabidopsis thaliana</i>	15222873	2E-39	Z
MDI06-B11	1	<i>tubulin alpha-5 chain-like protein</i>	<i>Solanum tuberosum</i>	83283967	6E-104	Z

CL146 contig1	3	tubulin folding cofactor B	<i>Arabidopsis thaliana</i>	20514259	4E-60	Z
MDI07-H05	1	Tubulin/FtsZ family, GTPase domain, putative	<i>Oryza sativa</i>	77549556	9E-16	Z
MDI35_D10	1	tubulin-like protein	<i>Arabidopsis thaliana</i>	4006861	3E-63	Z
MDI03-A07	1	TWD1 (Twisted Dwarf 1); FK506 binding / peptidyl-prolyl cis-trans isomerase	<i>Arabidopsis thaliana</i>	30686321	7E-99	O
CL483Contig1	2	type 2A protein phosphatase-1	<i>Vicia faba</i>	7248359	6E-31	TO
MDI24_F11	1	type IIIa membrane protein cp-wap13	<i>Vigna unguiculata</i>	2218152	1E-103	S
MDI26_F03	1	tyrosine phosphatase, putative	<i>Oryza sativa</i>	55770331	4E-62	S
MDI24_D11	1	U2 auxiliary factor small subunit	<i>Arabidopsis thaliana</i>	15723291	2E-23	A
MDI52_E05	1	ubiquinol-cytochrome c reductase iron-sulfur subunit, mitochondrial precursor, putative	<i>Oryza sativa</i>	50909025	1E-21	C
MDI41_C05	1	Ubiquitin	<i>Antirrhinum majus</i>	16071	5E-32	S
MDI34_B12	1	ubiquitin / ribosomal protein CEP52	<i>Oryza sativa</i>	52077170	5E-16	S
CL278 contig1	2	ubiquitin carrier protein	<i>Populus tomentosa</i>	56236462	6E-71	O
MDI08-C10	1	ubiquitin conjugating enzyme/ ubiquitin-like activating enzyme	<i>Arabidopsis thaliana</i>	18410856	8E-77	O
MDI37_B01	1	ubiquitin conjugating enzyme/ ubiquitin-like activating enzyme	<i>Arabidopsis thaliana</i>	79320672	8E-77	O
CL482 contig1	2	ubiquitin conjugating enzyme/ ubiquitin-like activating enzyme UBC10	<i>Arabidopsis thaliana</i>	30696309	3E-77	O
MDI31_E09	1	ubiquitin conjugating enzyme/ ubiquitin-like activating enzyme, UBC10	<i>Arabidopsis thaliana</i>	18423494	1E-47	O
MDI07-G08	1	ubiquitin extension protein-like protein	<i>Solanum tuberosum</i>	76573361	1E-23	O
MDI16_G07	1	ubiquitin fusion-degradation protein-like	<i>Solanum tuberosum</i>	78191476	4E-68	O
MDI07-H03	1	Ubiquitin-conjugating enzyme E2 UBC14	<i>Arabidopsis thaliana</i>	21554385	3E-36	O
MDI16_H05	1	Ubiquitin-conjugating enzyme E2 UBC14, putative	<i>Arabidopsis thaliana</i>	18410268	2E-11	O
CL150 contig1	2	Ubiquitin-conjugating enzyme E2-17 kDa (Ubiquitin-protein ligase) (Ubiquitin carrier protein)	<i>Lycopersicon esculentum</i>	464981	3E-75	O
CL383 contig1	2	ubiquitin-conjugating enzyme family protein-like	<i>Solanum tuberosum</i>	78191460	1E-70	O
MDI32-E05	1	ubiquitin-like protein	<i>Phaseolus vulgaris</i>	1684854	1E-12	OD
CL351Contig1	2	ubiquitin-like protein	<i>Arabidopsis thaliana</i>	62318558	7E-30	OD
MDI46-H05	1	ubiquitin-protein ligase	<i>Arabidopsis thaliana</i>	18397921	5E-31	O
MDI44-C11	1	ubiquitin-protein ligase - UPL5	<i>Arabidopsis thaliana</i>	15235410	5E-30	O
MDI23_H02	1	ubiquitin-protein ligase/ zinc ion binding	<i>Arabidopsis thaliana</i>	18379022	8E-58	O
MDI37_D05	1	Ubiquitin-specific protease 2 -UBP2; ubiquitin-specific protease	<i>Arabidopsis thaliana</i>	18390446	5E-29	S
MDI03-D05	1	ubiquitin-specific protease 23	<i>Arabidopsis thaliana</i>	11993486	1E-26	S
MDI32-A02	1	ubiquitin-specific protease 7	<i>Arabidopsis thaliana</i>	11993467	4E-60	S
MDI10-A05	1	UDP-D-Glucuronate 4-Epimerase 1 - GAE1; NAD binding / catalytic	<i>Arabidopsis thaliana</i>	15234745	2E-32	M
MDI42_C04	1	UDP-galactose 4-epimerase	<i>Cyamopsis tetragonoloba</i>	3021357	2E-54	G
MDI18_F07	1	UDP-glucose glucosyltransferase	<i>Fragaria x ananassa</i>	51705411	2E-17	S
MDI36_D06	1	UDP-glucose glucosyltransferase	<i>Fragaria x ananassa</i>	51705411	1E-45	S
MDI25_E03	1	UDP-glucose:sterol glucosyltransferase, putative	<i>Arabidopsis thaliana</i>	5080759	2E-59	S
MDI33_F09	1	UDP-glycosyltransferase 85A8	<i>Stevia rebaudiana</i>	37993655	2E-27	S

MDI13_E02	1	UDP-glycosyltransferase	<i>Jatropha curcas</i>	17484017	4E-32	S
MDI40_B08	1	UDP-N-acetylglucosamine pyrophosphorylase-like	<i>Oryza sativa</i>	53792734	2E-45	MG
MDI43-E01	1	UDP-sugar pyrophosphorylase	<i>Pisum sativum</i>	54650280	1E-90	G
MDI21_C07	1	UMP-kinase, putative	<i>Oryza sativa</i>	57900163	3E-16	S
MDI32-C04	1	UPL6; ubiquitin-protein ligase	<i>Arabidopsis thaliana</i>	79405297	6E-51	O
MDI11_A12	1	urea active transport protein, putative	<i>Oryza sativa</i>	37537200	3E-62	S
MDI30_B12	1	urease accessory protein D, putative	<i>Lycopersicon esculentum</i>	31096385	3E-13	S
MDI04-A03	1	UTP:alpha-D-glucose-1-phosphate uridylyltransferase	<i>Solanum tuberosum</i>	28863913	1E-82	G
MDI23_F12	1	UTP-glucose-1-phosphate uridylyltransferase/ nucleotidyltransferase	<i>Arabidopsis thaliana</i>	30686293	3E-112	G
CL85 contig1	4	vacuolar acid invertase	<i>Prunus cerasus</i>	31324469	2E-65	S
CL425 contig1	2	Vacuolar ATP synthase 16 kDa proteolipid subunit	<i>Gossypium hirsutum</i>	2493146	6E-53	S
MDI51-D09	1	vacuolar ATP synthase subunit C	<i>Arabidopsis thaliana</i>	6636332	1E-61	S
MDI47-A06	1	Vacuolar ATP synthase subunit G (V-ATPase G subunit) (Vacuolar proton pump G subunit)	<i>Citrus limon</i>	12585491	3E-24	S
CL440 contig1	2	vacuolar defense protein, putative	<i>Triticum aestivum</i>	78096543	2E-53	S
MDI14_H09	1	vacuolar H ⁺ -pyrophosphatase	<i>Prunus persica</i>	15982958	3E-92	P
MDI03-D04	1	Vacuolar invertase	<i>Citrus sinensis</i>	19705584	6E-66	S
CL111 contig1	3	vacuolar protein sorting; Vps29p, putative	<i>Oryza sativa</i>	50251955	2E-41	O
MDI08-D12	1	Vacuolar sorting receptor 1 precursor (BP-80) (80 kDa proaleurein binding protein)	<i>Pisum sativum</i>	71153401	9E-94	OUZ
MDI19_D04	1	valyl tRNA synthetase	<i>Arabidopsis thaliana</i>	1890130	2E-91	S
MDI10-B02	1	Vernalization Independence 4	<i>Arabidopsis thaliana</i>	21929715	4E-11	S
MDI32-E11	1	vicilin seed storage protein	<i>Juglans nigra</i>	31321944	1E-10	S
MDI15_B02	1	Vitamin C Defective 2 - VTC2	<i>Arabidopsis thaliana</i>	18416877	2E-28	S
CL136 contig1	3	Vitamin C Defective 2 - VTC2	<i>Arabidopsis thaliana</i>	18416877	5E-81	S
CL293Contig1	2	Vitamin C Defective 2 -VTC2	<i>Arabidopsis thaliana</i>	18416877	4E-12	S
MDI36_E07	1	VMA10 (Vacuolar Membrane ATPase 10)	<i>Arabidopsis thaliana</i>	30678328	1E-22	S
MDI01-D03	1	WD repeat domain 5 protein, putative	<i>Oryza sativa</i>	50938509	4E-45	R
MDI06-B08	1	WD40 protein Ciao1, putative	<i>Oryza sativa</i>	50935503	9E-81	R
MDI37_F01	1	WD-40 repeat protein MSI1 (Multicopy Suppressor of IRA1)	<i>Arabidopsis thaliana</i>	21387123	8E-57	R
MDI09-E05	1	WD-40 repeat protein-like	<i>Oryza sativa</i>	50911058	5E-55	R
CL246 contig1	2	WD-40 repeat protein-like	<i>Oryza sativa</i>	50941369	6E-78	R
MDI50-A05	1	WD-repeat protein	<i>Capsella rubella</i>	38260650	2E-53	R
CL186 contig1	2	wound induced protein	<i>Lycopersicon esculentum</i>	19320	9E-17	S
CL475 contig1	2	wound-induced protein WI12	<i>Mesembryanthemum crystallinum</i>	4574320	2E-23	S
MDI13_D05	1	wound-responsive protein, putative	<i>Arabidopsis thaliana</i>	21593594	1E-44	S
MDI16_C12	1	WRKY17; transcription factor	<i>Arabidopsis thaliana</i>	18400580	3E-27	K
MDI16_B11	1	Xaa-Pro aminopeptidase 1	<i>Lycopersicon esculentum</i>	15384989	3E-86	E

MDI29_B08	1	<i>Xanthine/uracil permease</i>	<i>Pseudomonas fluorescens PfO-1</i>	77460950	2E-45	F
CL145 contig1	3	<i>YABBY2-like transcription factor YAB2</i>	<i>Antirrhinum majus</i>	41745658	1E-51	K
CL299 contig1	2	<i>YGL010w-like protein</i>	<i>Picea mariana</i>	2982301	6E-55	K
MDI36_B10	1	<i>Yippee-like protein</i>	<i>Oryza sativa</i>	50938763	1E-24	K
MDI07-G02	1	<i>yippee-like protein at4g27740, putative</i>	<i>Oryza sativa</i>	77555722	7E-13	K
MDI40_E10	1	<i>ZFP2 (Zinc Finger Protein 2); nucleic acid binding / transcription factor/ zinc ion binding</i>	<i>Arabidopsis thaliana</i>	15242152	1E-16	K
MDI02-C01	1	<i>zinc finger (C3HC4-type RING finger) protein-like</i>	<i>Oryza sativa</i>	52076906	6E-49	K
MDI13_F07	1	<i>zinc finger DNA-binding protein</i>	<i>Catharanthus roseus</i>	55734108	4E-10	K
CL133 contig1	2	<i>zinc finger protein LSD1</i>	<i>Oryza sativa</i>	42406064	5E-31	K
CL174 contig1	2	<i>zinc finger protein, putative</i>	<i>Oryza sativa</i>	34908682	7E-50	K
MDI19_A09	1	<i>zinc finger protein, putative</i>	<i>Oryza sativa</i>	50931917	9E-11	K
MDI16_E02	1	<i>zinc finger protein-like</i>	<i>Oryza sativa</i>	51091322	1E-26	K
MDI11_H03	1	<i>zinc finger transcription factor-like protein</i>	<i>Arabidopsis thaliana</i>	22136426	1E-75	K
MDI22_G07	1	<i>zinc ion binding</i>	<i>Arabidopsis thaliana</i>	15233935	6E-20	K
MDI15_G10	1	<i>zinc transporter ZIP6</i>	<i>Arabidopsis halleri subsp. halleri</i>	40782197	6E-35	KP
MDI06-C02	1	<i>zinc transporter, putative</i>	<i>Oryza sativa</i>	28209495	8E-55	KP
CL49 contig1	5	<i>zinc-finger protein</i>	<i>Oryza sativa</i>	35187687	7E-46	K
MDI17_G07	1	<i>zinc-finger protein</i>	<i>Oryza sativa</i>	35187687	5E-31	K
CL102 contig1	3	<i>ZIP family metal transporter</i>	<i>Chengiopanax sciadophylloides</i>	82581289	3E-123	KP

Tabela 8 - Relação geral de prováveis genes identificados expressos em frutos de camu-camu.