

UNIVERSIDADE DE RIBEIRÃO PRETO – UNAERP

DOUTORADO EM BIOTECNOLOGIA

Sandra Regina Costa Fantinato

**Sistema de tratamento e caracterização da expressão de genes
de fenilpropanoides em folhas de *Myracrodruon urundeuva*
(Aroeira-do-Sertão)**

Ribeirão Preto

2015

UNIVERSIDADE DE RIBEIRÃO PRETO – UNAERP

DOUTORADO EM BIOTECNOLOGIA

Sistema de tratamento e caracterização da expressão de genes de
fenilpropanoides em folhas de *Myracrodruon urundeuva* (Aroeira-
do-Sertão)

Tese apresentada ao curso de
Biotecnologia da Universidade de
Ribeirão Preto para obtenção do Título
de Doutora em Biotecnologia.

Orientadora: Profa. Dra. Suzelei de
Castro França

Ribeirão Preto

Abril - 2015

Ficha catalográfica preparada pelo Centro de
Processamento Técnico da Biblioteca Central da
UNAERP

- Universidade de Ribeirão Preto -

F216s Fantinato, Sandra Regina Costa, 1968-
Sistema de tratamento e caracterização da expressão de
genes de fenilpropanoides em folhas de *Myracrodruon urundeuva*
(Aroeira-do-Sertão) / Sandra Regina Costa Fantinato. - - Ribeirão
Preto, 2015.
189 f.: il. color.

Orientadora: Prof^a. Dr^a. Suzelei de Castro França

Tese (doutorado) - Universidade de Ribeirão Preto, UNAERP,
Biotecnologia. Ribeirão Preto, 2015.

1. Biologia molecular. 2. Bioinformática. 3. *Pipeline*. I. Título.

CDD 660

DEDICATÓRIA

Às minhas queridas, preciosas e amadas
filhas

Yasmin e

Alanis

e ao meu amor Danilo Fantinato.

AGRADECIMENTOS

Agradeço em primeiro lugar a **Deus** que iluminou o meu caminho durante esta caminhada.

À Universidade de Ribeirão Preto pela oportunidade do exercício científico e pelo apoio para a realização deste trabalho, principalmente na pessoa da Profª Drª **Neide de Souza Lehfeld**.

Agradeço minha orientadora Profª Drª **Suzelei de Castro França**, por sua cumplicidade e responsabilidade na construção dessa Tese.

Ao Prof. Dr. **Milton Faria Junior**, por me conduzir e contribuir com os primeiros passos da biotecnologia e da bioinformática.

Aos meus colegas **Caio, Carolina, Fernanda, Patricia G.Roberto, Renan e Vinicius** pelo apoio.

Aos colegas, professores e funcionários do Programa de Pós-Graduação *Strictu Sensu* em Biotecnologia da Unaerp, pela cooperação e companheirismo.

Agradeço também ao meu esposo, **Danilo**, meu companheiro incondicional, que de forma especial e carinhosa me deu força e incentivo, me apoiando sempre.

Às minhas filhas, **Yasmin** e **Alanis**, por iluminar minha vida e me encher de alegria. Perdoem-me pelos momentos de ausência e jamais se esqueçam do amor infinito que tenho por vocês.

Agradeço aos meus pais, **Oriovaldo** (*in memoriam*) e **Juraci**, meus maiores exemplos. Obrigada por cada incentivo e orientação, pelas orações em meu favor. Obrigada por serem a minha referência de tantas maneiras e estarem sempre presentes na minha vida de uma forma indispensável.

RESUMO

Por meio deste trabalho, foi criado um *pipeline* computacional usando uma combinação de ferramentas de *software* de código aberto, configuradas para agilizar a mineração de sequências expressas (ESTs), o corte de vetor, o agrupamento de EST, além de executar a busca *on-line* para categorização e anotação funcional de genes alvo. Foi modelado e implementado na linguagem de programação Java e banco de dados MySQL um sistema baseado na *web*, chamado WISE. Com os arquivos de *singlets* e *contigs* gerados pelo sistema WISE, foram feitos os alinhamentos, utilizando-se o algoritmo BLASTx de comparação de sequências de nucleotídeos com base de dados não redundante de proteínas. Implementou-se o programa Online Blast XLS Reporter em linguagem de programação C# (CSharp) para executar o BLASTx em batelada. Com o objetivo de verificar a eficiência do sistema desenvolvido, dados obtidos experimentalmente da biblioteca de cDNA de folhas da *Myracrodruon urundeuva* – uma espécie medicinal nativa do cerrado brasileiro – foram submetidos à análise computacional. Com a implantação do sistema WISE, proporcionaram-se aos pesquisadores interoperabilidade, usabilidade e agilidade na execução das análises, principalmente, ao ser comparado às técnicas de manipulação antes utilizadas.

Palavras-chave: Biologia molecular. *Pipeline*. Bioinformática.

ABSTRACT

Through this work, a computational pipeline using a combination of open source software tools was created, set to streamline the mining of expressed sequence tag (ESTs), vector cutting, grouping ESTs, and perform the search on-line for categorization and functional annotation of target genes. It was modeled and implemented in the Java programming language and MySQL database one web-based system, called WISE. With contigs and singlets files generated by WISE system, the alignments were made using the BLASTx algorithm for the comparison of nucleotide sequences based on the non-redundant protein data. It was implemented the XLS Blast Online Reporter program in programming language C# (CSharp) to run the BLASTx batch. In order to verify the efficiency of the system developed, experimentally obtained data from the cDNA library of leaves from Myracrodruon urundeuva - a native medicinal plant of the Brazilian Cerrado (savannah) - were subjected to computer analysis. With the implementation of the WISE system was provided to researchers interoperability, usability and agility in the execution of the tests, especially when compared to manipulation techniques used previously.

KEYWORDS: Molecular biology. Pipeline. Bioinformatics.

LISTA DE FIGURAS

Figura 1 – <i>Myracrodum urundeuva</i>	12
Figura 2 – Principais fatores que influenciam no metabolismo secundário das plantas.....	14
Figura 3 – Principais vias do metabolismo de fenilpropanoides e suas interligações	15
Figura 4 – Produção de ESTs e ORESTES	26
Figura 5 – <i>Pipeline</i> da montagem dos genomas/transcriptomas	29
Figura 6 – Índices de qualidade do PHRED no <i>Consed</i>	31
Figura 7 – Fotografia das folhas da <i>M. urundeuva</i>	38
Figura 8 – Diagrama de <i>use case</i> de negócio	39
Figura 9 – Diagrama de Ishikawa.....	41
Figura 10 – Interface do programa Sequence Name Converter	44
Figura 11 – Exemplo de um cromatograma gerado pelo sequenciador.....	45
Figura 12 – Representação das três principais fases do algoritmo de montagem do CAP3	48
Figura 13 – Requisição HTTP enviada para o servidor do NCBI, objetivando o processamento da sequência	50
Figura 14 – Estrutura do programa Online Blast XLS Reporter	51
Figura 15 – Interface do programa Online Blast XLS Reporter.....	51
Figura 16 – Esquema do <i>pipeline</i>	53
Figura 17 – Diagrama Entidade Relacionamento	56
Figura 18 – Arquitetura cliente/servidor – Sistema WISE	60
Figura 19 – Logotipo do sistema WISE	61
Figura 20 – Representação de arquitetura MVC em camadas	63
Figura 21 – <i>Template</i> padrão das páginas do sistema WISE	65
Figura 22 – Estrutura de pastas do sistema WISE	68

Figura 23 – Tela de <i>login</i> do sistema WISE	70
Figura 24 – Tela de cadastro de usuários	70
Figura 25 – Tela principal do sistema.....	71
Figura 26 – Tela de edição e administração de usuários do sistema	72
Figura 27 – Tela de meus projetos.....	72
Figura 28 – Tela de execução do <i>pipeline</i>	73
Figura 29 – Tela com informações de saída do PHRED	74
Figura 30 – Listagem dos arquivos após saída do PHRED	74
Figura 31 – Portal KEGG	75
Figura 32 – Mapa Metabólico da enzima e descrição.....	76
Figura 33 – Exemplo de predição de função de uma proteína	77
Figura 34 – Exemplo de criação de uma variável de ambiente	78
Figura 35 – Tela de autenticação do usuário para acesso ao aplicativo <i>mobile</i>	83
Figura 36 – Tela de Menu de Opções do aplicativo <i>mobile</i>	83
Figura 37 – Representação esquemática da execução do <i>pipeline</i> de programas no sistema WISE	86
Figura 38 – Percentual de <i>singlets no results</i> e <i>singlets</i> que apresentam resultados com plantas	88
Figura 39 – Percentual de <i>contigs no results</i> e <i>contigs</i> que apresentam resultados com plantas	89

LISTA DE QUADROS

Quadro 1: Relação de programas usados no <i>pipeline</i> com a descrição, funcionalidades e parâmetros.....	42
Quadro 2: Dados gerados após a execução do programa Online Blast XLS Reporter.....	52
Quadro 3: Descrições de todas as entidades e atributos criados no banco de dados do sistema WISE.....	57
Quadro 4: Tipos de teste.....	79

LISTA DE SIGLAS

A	Adenina
AAAA	Ano
ACeDB	<i>A Caenorhabdits elegans Data Base</i>
APP	<i>Application</i>
ARO	Aroeira
BISTI	<i>Biomedical Information Science and Technology Initiative</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
BOSC	<i>Bioinformatics Open Source Conference</i>
C	Citosina
C#	CSharp
cDNA	Molécula de DNA complementar a um mRNA transcrito
ICEP	Código de Endereço Postal
CPAN	<i>Comprehensive Perl Archive Network</i>
CSS	<i>Cascading Style Sheets</i>
DAO	<i>Data Access Object</i>
DD	Dia
dbEST	Banco de dados de EST
DDBJ	<i>DNA Data Bank of Japan</i>
DDNA	<i>Data Bank of Japan</i>
DNA	Ácido desoxirribonucléico
DNAr	Ácido desoxirribonucléico recombinante
EBI	<i>European Bioinformatics Institute</i>
EMBL	<i>European Molecular Biology Laboratory</i>
EST	<i>Expressed Sequence Tags</i>
EUA	Estados Unidos da América
FTP	<i>File Transfer Protocol</i>
FO	Folha
G	Guanina
GB	<i>Giga Byte</i>
GenBank	<i>National Center for Biotechnology Information</i>
GPL	<i>General Public License</i>
HD	<i>Hard Disk</i>
HTTP	<i>HyperText Transfer Protocol</i>
HTML	<i>Hypertext Markup Language</i>
IDE	<i>Integrated Development Environment</i>
JDK	<i>Java Developer Kit</i>
JSF	<i>Java Server Faces</i>
JSP	<i>Java Server Pages.</i>
JAR	<i>Java ARchive</i>
JVM	<i>Java Virtual Machine</i>
KB	<i>Kilobyte</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>

LNCC	Laboratório Nacional de Computação Científica
Log	Logaritmo
MEDLINE	<i>Medical Literature Analysis and Retrieval System Online</i>
MD5	<i>Message-Digest algorithm 5</i>
MM	Mês
mRNA	RNA mensageiro
MS	Milsegundos
M.urundeuva	Myracrodruon urundeuva.
MVC	<i>Model View Controller</i>
NCBI	<i>National Center for Biotechnology Information</i>
NFS	<i>Network File System</i>
NHI	<i>National Institute of Health</i>
NLM	National Library of Medicine
OBF	<i>Open Bioinformatics Foundation</i>
ONSA	Organization for Nucleotide Sequencing and Analysis
OMIM	Online Mendelian Inheritance in Man
PB	Pares de base
PCR	<i>Polymerase Chain Reaction</i>
PDB	<i>Protein Data Bank</i>
PERL	<i>Practical Extraction and Report Language</i>
PHP	<i>Hypertext Preprocessor</i>
PIR	<i>Protein Information Resource</i>
PL	Placa
RAM	<i>Random Access Memory</i>
RefSeq	<i>Reference Sequence</i>
RNA	Ácido ribonucleico
SS	Segundos
SEQ	Sequência
SGBD	Sistema Gerenciador de Banco de Dados
SHA	<i>Secure Hash Algorithm</i>
SQL	<i>Structured Query Language</i>
T	Timina
TCP	<i>Transmission Control Protocol</i>
UML	<i>Unified Modeling Language</i>
UNAERP	Universidade de Ribeirão Preto
UNICAMP	Universidade Estadual de Campinas
USP	Universidade de São Paulo
XML	<i>Extensible Markup Language</i>
WWW	<i>World Wide Web</i>

SUMÁRIO

1. INTRODUÇÃO	12
1.1 JUSTIFICATIVA.....	34
1.2 OBJETIVO.....	35
1.2.1 Objetivos Específicos	36
2. ESTRATÉGIA METODOLÓGICA	37
2.1 MODELAGEM E IMPLEMENTAÇÃO DO SISTEMA WISE.....	54
2.2 IMPLEMENTAÇÃO DO SISTEMA WISE.....	62
2.3 TESTE DE EXECUÇÃO DO SISTEMA WISE	78
2.4 IMPLEMENTAÇÃO DE CONSULTAS DO SISTEMA WISE VIA APLICAÇÃO <i>MOBILE</i>	81
3. RESULTADOS E DISCUSSÃO.....	84
4. CONCLUSÃO	93
5. REFERÊNCIAS.....	95

INTRODUÇÃO

A *M.urundeuva*, popularmente conhecida como aroeira-do-sertão, é uma espécie arbórea nativa, que tem preferência por solos calcários e de alta fertilidade. Seus troncos são altos, medindo de 10 a 15 metros, e podem atingir 1 metro de diâmetro (LORENZI, 1992). As flores são amarelo-alaranjadas com forma de estrelas. Floresce entre os meses de julho e setembro, e a maturação dos frutos ocorre de setembro a outubro (Figura 1) (NUNES et al., 2008). De acordo com Monteiro et al. (2012), o significado de “aroeira” é árvore arara; é uma simplificação do vocábulo araroeira, derivado de arara + eira.

Seu limite de distribuição natural se estende pelas regiões Nordeste, Sudeste e Centro-Oeste do Brasil até a região chaquenha da Bolívia, Paraguai e Argentina. Nas florestas estacionais decíduais do norte de Minas Gerais, a *M. urundeuva* ocorre frequentemente nas matas secas calcárias e na caatinga arbórea (NUNES et al., 2008).

Figura 1 – *Myracrodruon urundeuva*.



Fonte: <http://multimudas.com.br> acessado em 16/05/2013 às 01:30

No Estado do Ceará, é de uso popular a entrecasca da *M. urundeuva*, sob a forma de extrato aquoso ou simplesmente de cozimentos, como anti-inflamatório para várias afecções, por exemplo, ginecológicas (na prática de “banho de assento” usada por mulheres no tratamento ginecológico pós-parto) (MATOS, 1989; VIANA et al., 1995) e cicatrizante para ferimentos cutâneos tanto para adultos quanto para crianças (LORENZI; MATOS, 2002). As raízes são usadas no tratamento de reumatismo e as folhas são indicadas para o tratamento de úlceras (NUNES et al., 2008). Sua madeira, em função da durabilidade e resistência ao ataque de cupins e fungos, é muito usada na construção civil como postes e na confecção de móveis (NUNES et al., 2008). Bernardes (2011) ressalta, ainda, que as cascas de *M. urundeuva* além do uso medicinal são também utilizadas na fabricação de tintas para tecidos.

Os diversos usos *M. urundeuva* despertaram interesse na realização de estudos farmacológicos com vistas à comprovação de sua eficácia terapêutica, validada e correlacionada à presença de chalconas, metabólitos secundários da classe dos fenilpropanoides (CAMPOS, 2008).

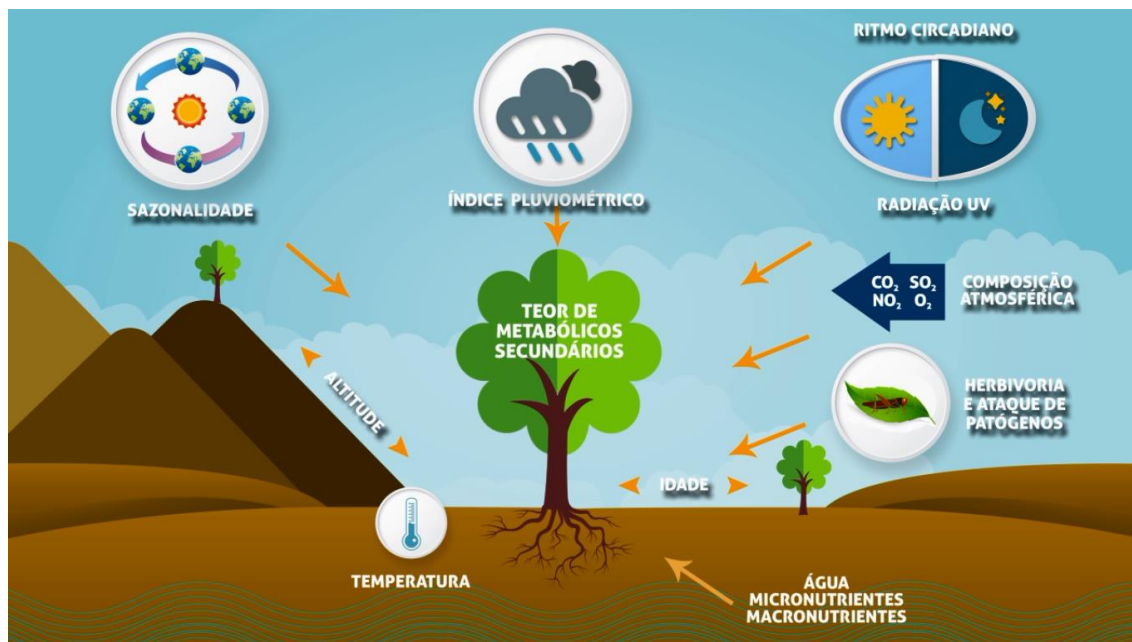
Segundo Kurkin 2003, a via de formação dos fenilpropanoides é considerada uma importante rota do metabolismo secundário em plantas. Eles são responsáveis por gerar uma grande variedade de substâncias, entre as quais podem ser citados os flavonoides, os isoflavonoides, os taninos, as lignanas, as cumarinas, os ácidos fenólicos bem como chalconas - composto responsável pela síntese de isoflavonas (com atividade antifúngica).

O metabolismo secundário de plantas é responsável pela biossíntese de inúmeras moléculas bio/farmacologicamente ativas, sendo constituído por vias ramificadas, catalizadas por enzimas e sujeitas à regulação transcricional de genes, que são ativadas ou desativadas em resposta a diferentes estímulos ambientais.

Metabólitos secundários são importantes para a sobrevivência e a propagação das plantas que os produzem; servem como sinais químicos e permitem que a planta responda a estímulos ambientais para a adaptação ao meio (BENNETT, WALLSGROVE, 1994). Esses compostos são responsáveis pela proteção das plantas contra os herbívoros e contra a infecção por microrganismos patogênicos. Também agem como atrativo para animais polinizadores (TAIZ, ZEIGER, 2010). A Figura 2, colocada a seguir, ilustra alguns

dos principais fatores que podem influenciar o acúmulo de metabólitos secundários em plantas.

Figura 2 - Principais fatores que influenciam no metabolismo secundário das plantas.



Fonte: *Medicinal plants: factors of influence on the content of secondary metabolites* (GOBBO-NETO, LOPES, 2007, Adaptado pela autora).

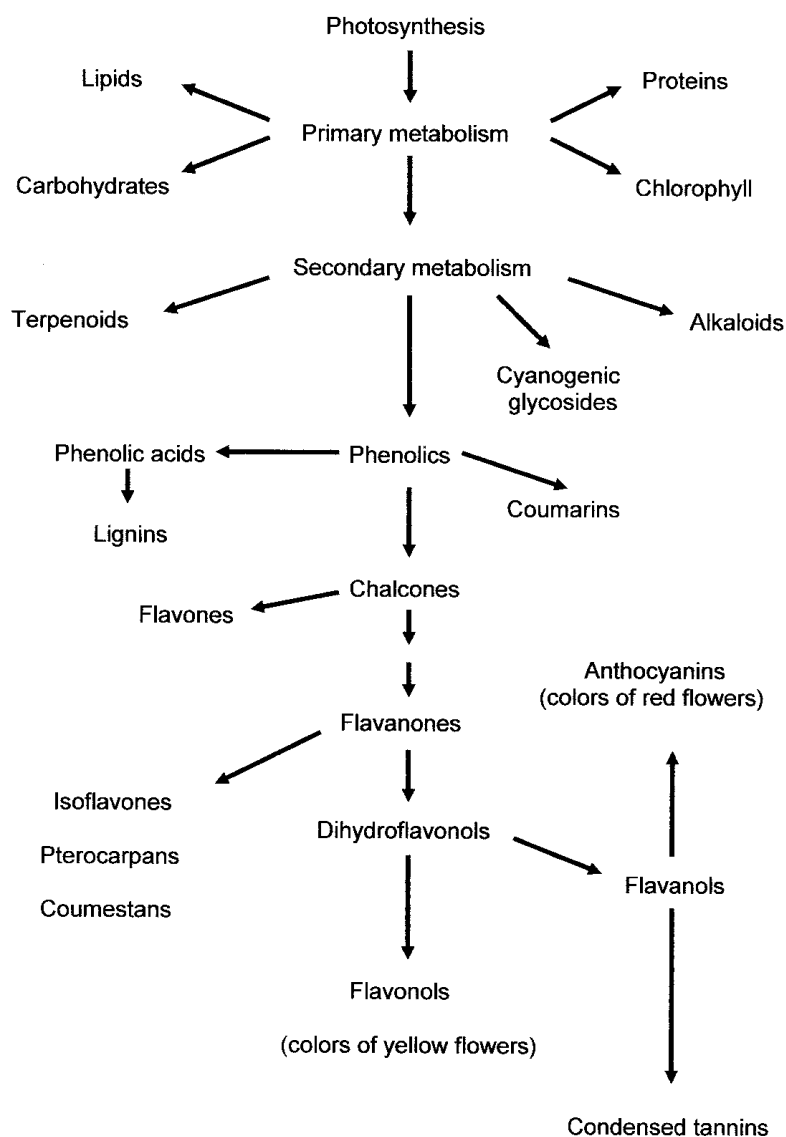
O metabolismo secundário pode ser dividido em classes estruturais, tais como: terpenos, lignanas, taninos, lactonas, esteroides, chalconas, flavonas, flavanonas, alcaloides e quinonas. A Figura 3 ilustrará os três grandes grupos de metabolismo secundário de plantas: terpenos, compostos fenólicos e alcaloides (PERES, 2015).

Os terpenos se originam a partir do ácido mevalônico ou do piruvato mais o 3-fosfoglicerato (PERES, 2015). Alguns terpenos têm ação no crescimento dos vegetais. Em razão de sua toxicidade, agem também na defesa de muitos vegetais contra animais herbívoros e insetos (TAIZ, ZEIGER, 2010).

Os fenóis são derivados do ácido chiquímico ou ácido mevalônico. Os compostos fenólicos de ervas e plantas medicinais dietéticos incluem ácidos fenólicos, flavonoides, taninos, cumarinas, lignanas e outros. Várias bioatividades de compostos fenólicos são responsáveis por suas propriedades quimiopreventivas, como por exemplo, efeitos antioxidantes, anticarcinogênicos e anti-inflamatórios (HUANG, CAI, ZHANG, 2010).

Os alcaloides são derivados de aminoácidos aromáticos, os quais são derivados do ácido chiquímico (PERES, 2015). De acordo com o portal science.howstuffworks (2014), alguns alcaloides, assim como a nicotina, são utilizados em pesticidas e sua principal utilização, no entanto, é na medicina, porque eles podem agir rapidamente em áreas específicas do sistema nervoso. Os alcaloides são os componentes ativos de muitos anestésicos, sedativos, estimulantes, relaxantes e tranquilizantes.

Figura 3 - Inter-relações entre o metabolismo primário e secundário em plantas.



Fonte: (GIADA, 2013, Pág.88).

É importante destacar que a casca da *M.urundeuva* é rica em tanino e outras substâncias fenólicas; contém flavonoides e duas chalconas diméricas de forte ação anti-inflamatória (BANDEIRA, 2002).

Monteiro, Araújo e Amorin (2005) em suas pesquisas descrevem que o tanino é um metabólito secundário, que possui propriedades antimicrobianas; antídotos em intoxicações por metais pesados e alcaloides; adstringentes; via externa: cicatrizantes; via interna: antidiarreico, antisépticos, antioxidantes.

Os estudos de Yoshihara, Minho e Yamamura (2013) apresentam a utilização de plantas taniníferas, cientificamente testadas para o controle das parasitoses gastrintestinais em ruminantes. De acordo com os estudos, duas hipóteses são sugeridas para explicar o efeito anti-helmíntico dos taninos contra uma população de nematódeos gastrintestinais. A primeira é a capacidade que os taninos têm de agir no parasito, afetando severamente os processos biológicos dos nematódeos. A segunda hipótese é a de que os taninos poderiam atuar indiretamente, melhorando a utilização proteica pelo hospedeiro e, conseqüentemente, obtendo uma melhor resposta imunológica deste aos parasitos.

Nas pesquisas de Albuquerque (2003), analisou-se o efeito das chalconas na conjuntivite alérgica induzida por ovalbumina, em cobaias, e concluiu-se que as chalconas têm efeito terapêutico sobre esse tipo de inflamação.

Viana, Bandeira e Matos (2003), com experimentos em ratos, comprovaram atividades analgésicas e anti-inflamatórias por chalconas diméricas presentes na entrecasca da *M.urundeuva*.

Bianco (2004) apud Machado e Oliveira (2014) comprovou que o uso do extrato vegetal aquoso e hidroalcoólico das folhas da *M.urundeuva* é capaz de diminuir o crescimento bacteriano.

Os trabalhos de Sá et al. (2009), demonstram que, após isolar proteína lectina da casca *M.urundeuva*, ela apresentou a função larvicida contra o mosquito *Aedes aegypti*.

Nobre et al. (2009) avaliaram os efeitos neuroprotetores das chalconas de *M.urundeuva* sobre a citotoxicidade induzida por 6-hidroxidopamina em células mesencefálicas de ratos. Os resultados sugerem que chalconas poderiam proporcionar benefícios, juntamente com outras terapias, em lesões neurodegenerativas, como a doença de Parkinson.

Nos estudos de Ferreira et al. (2011), o extrato etanólico extraído a partir de sementes de *M.urundeuva* apresentou traços de esteroides (alcaloides e fenóis em sua composição), demonstrando atividade citotóxica em células tumorais humanas.

Os compostos fenólicos compreendem estruturas químicas que podem ser classificados pelo número e arranjo de seus átomos de carbono, sendo divididos em pelo menos dez grupos: fenóis simples, ácidos fenólicos, cumarinas, isocumarinas, naftoquinonas, xantonas, estilbenos, antraquinonas, flavonoides e lignanas (FRANÇA et al, 2012).

Lignanas são substâncias fenólicas, presentes nas plantas e que conseguem se ligar aos receptores de estrógeno nas células, funcionando como fitoestrógenos. Entre as suas propriedades farmacológicas estão, ainda, atividades antifúngica, antitumoral, anti-inflamatória, hepato-regeneradora (TRAZZI, 2008).

Flavonoides (flavonas e flavanóis) são compostos de origem natural do grupo dos fenilpropanoides, abundantes em algumas plantas e agrupados em diferentes classes de acordo com suas características químicas e biossintéticas, possuindo relevantes propriedades farmacológicas, tais como: ação antiviral, anti-inflamatória, atividade antimicrobiana e antioxidante, anticarcinogênicas, antialérgicas, além de serem potentes inibidores de enzimas-chaves como várias quinases e terem papel na ativação de enzimas glicolíticas (CAMPOS, 2008). Mais de 4000 variedades de flavonoides foram identificadas, muitas das quais são responsáveis pelas cores das flores, frutos e folhas (ROBERT et al., 2001).

Muitos benefícios fisiológicos são atribuídos aos flavonoides, incluindo a proteção contra doenças cardiovasculares e câncer, em razão de seus efeitos antioxidantes e propriedades de eliminação de radicais livres. Os flavonoides têm várias propriedades que são benéficas para a prevenção e tratamento de doenças oculares, particularmente, aquelas que envolvem a perda de células nervosas (MAHER, HANNEKEN, 2005).

Nos trabalhos de Yang et al. (2013), é destacado que os flavonoides presentes na planta medicinal *Glycyrrhiza uralensis*, também conhecida como alcaçuz chinês (planta nativa da Ásia), são eficientes no tratamento de

asma alérgica, pois reduziram significativamente inflamações nos pulmões causadas pela doença.

As chalconas são consideradas como intermediárias essenciais na biossíntese dos flavonoides e despertam grande interesse químico e farmacológico por apresentarem diversas atividades biológicas, as quais variam conforme os diferentes substituintes presentes nas moléculas. Elas apresentam, também, atividade antiviral e antioxidante, ação anti-histaminica, além de ação antitumoral, antimalárica, antifúngica e atividade antileishmania (CAMPOS, 2008).

A produção de moléculas bioativas pode ser regulada em nível transcricional com a indução ou repressão da expressão de genes específicos. Assim sendo, procedimentos biotecnológicos para isolamento identificação e categorização funcional de genes expressos são úteis não só para a elucidação de vias de biossíntese, como também para a manipulação genética da produção de metabólitos secundários bioativos.

A partir desse contexto, é importante destacar as técnicas e ferramentas para análise das etiquetas de genes expressos (*Expressed Sequence Tags* - ESTs). Parkinson e Blaxter (2009) explicam que ESTs, são pedaços de sequências de RNAs mensageiros (mRNA) derivados por meio de reações de sequenciamento únicas e efetuados em clones selecionados aleatoriamente a partir de bibliotecas de cDNA. São descritas como etiquetas, pois correspondem apenas a pedaços dos genes que um organismo expressa em uma determinada situação e permitem identificar os genes que estão sendo induzidos ou reprimidos e produzidos por uma célula numa determinada condição, bem como permitem a realização de estudos comparativos (WOLFSBERG, LANDSMAN, 1997).

Atualmente há diversos bancos de dados de ESTs e entre eles destacam-se: o NCBI – *National Center for Biotechnology Information* - Centro Nacional para Informação sobre Biotecnológica dos EUA, o qual fornece acesso à informação biomédica e genômica (NCBI, 2013), e é considerado o banco de dados central sobre informações genômicas; o GenBank (*National Center for Biotechnology Information*), que faz parte do Banco de Dados de Colaboração Internacional; o *Nucleotide Sequence*, que compreende o DNA DataBank do

Japão (DDBJ); o EMBL - *European Molecular Biology Laboratory* - Laboratório Europeu de Biologia Molecular.

De acordo com o NCBI (2013), o RefSeqGene é a coleção de sequências genômicas de referência específica para gene humanos; é um subconjunto do banco de dados do NCBI RefSeq - *Reference Sequence*, o qual é baseado na revisão de especialistas de bancos de dados específicos do local e da comunidade de testes genéticos. Com a proposta de produzir um conjunto não redundante de sequências de DNA genômico, transcritos (cDNA) e de proteínas para diversos organismos, em sua base de dados, o RefSeq é um banco de dados muito utilizado pelos cientistas por não conter dados redundantes.

O SWISS-PROT é um banco de dados de sequência curada de proteínas, (tais como a descrição da função de uma proteína, a sua estrutura de domínios, modificações pós-traducionais, variantes, etc), com um nível mínimo de redundância e alto nível de integração com outras bases de dados (BAIROCH, APWEILER, 2010).

Os bancos de sequências também podem ser classificados como bancos estruturais ou bancos funcionais. Sequências de nucleotídeos, de aminoácidos e a estrutura de proteínas apresentam informações diferentes, apesar de representarem um produto de um gene e devem ser armazenadas em bancos específicos. Nos bancos de dados estruturais, são mantidos apenas dados relativos à estrutura de proteínas, já em bancos funcionais, são fornecidos *links* para mapas metabólicos de organismos com genoma parcialmente ou totalmente sequenciado. As buscas são realizadas a partir de palavras-chave ou de sequências. O principal banco funcional é o *Kyoto Encyclopedia of Genes and Genomes*, conhecido como KEGG (PROSDOCIMI et al., 2003).

Tal como descrito anteriormente, o arquivamento de dados passou a ser de responsabilidade de projetos nacionais e internacionais. Os bancos de dados incluem em suas bases: sequências de ácidos nucleicos (incluindo genomas completos), sequências de aminoácidos de proteínas, estruturas de proteínas e ácidos nucleicos, estruturas cristalográficas de pequenas moléculas, funções de proteínas, padrões de expressão de genes, vias metabólicas e redes de interação e controle (LESK, 2008).

No caso das anotações de sequências de DNA, o GenBank é um banco de dados de referência. Localizado no *National Institutes of Health (NIH)*, faz parte da *International Nucleotide Sequence Database Collaboration*, que compreende o DNA *DataBank of Japan (DDBJ)*, o *European Molecular Biology Laboratory (EMBL)*, e o NCBI GenBank, com atualizações diárias que mantêm suas bases de dados sempre sincronizadas e atualizadas (NCBI, 2013).

Esses três bancos formam a *International Nucleotide Sequence Database Collaboration (INSDC)*, armazenando e trocando informações para reunir as sequências nucleotídicas depositadas nesses bancos e garantir que essas sequências sejam acessadas em todo o mundo. Esses três bancos de dados trocam informações sobre nucleotídeos diariamente, portanto sequências encontradas em um banco também serão encontradas nos outros bancos (GENEBIO, 2014).

Lançamentos bimestrais completos e atualizações diárias do banco de dados GenBank estão disponíveis por FTP - *File Transfer Protocol* (protocolo que faz transferência de arquivos pela Internet). Na *home page* NCBI: www.ncbi.nlm.nih.gov é possível acessar GenBank e seus serviços de recuperação e análise (BENSON et al., 2012).

No contexto de sequenciamento de proteínas, destaca-se o UniProt (*Universal Protein Resource*), que é um catálogo de informação sobre proteínas. Funciona como um depósito de sequências e de funções de proteínas, além de reunir a informação contida na TrEMBL, na Swiss-Prot e na PIR (UNIPROT, 2013).

Existem, ainda, bancos de dados de estruturas de macromoléculas biológicas, cujo objetivo consiste em arquivar, anotar e distribuir conjuntos de coordenadas atômicas. O *Protein Data Bank (PDB)* é o principal banco de dados de estruturas de macromoléculas e possui como conteúdo estruturas de proteínas, ácidos nucleicos e uns poucos carboidratos. Atualmente, o PDB é gerenciado por uma organização distribuída e chamada *Research Collaboratory for Structural Bioinformatic*. Bases na Rutgers University – New Jersey em San Diego Supercomputer Center – California e no National Institute of Standards and Technology – Mayland (LESK, 2008).

Santos e Ortega (2003) relatam que outros bancos de dados são específicos de um organismo, tal como o OMIM (*Online Mendelian Inheritance*

in Man), que foi criado para catalogar todos genes e alelos relacionados a doenças e outras características humanas, bem como proporcionar um detalhamento técnico e bibliografia referente a cada característica.

Com o passar dos anos, programas como o BLAST (*Basic Local Alignment Search Tool*) foram criados para realizar buscas rápidas de semelhanças entre uma sequência qualquer e as sequências já armazenadas nas bases de dados. O BLAST encontra regiões de similaridade local entre sequências e compara as sequências de nucleótidos ou as bases de dados de sequências de proteínas, além de calcular a significância estatística dos resultados. BLAST pode ser usado para inferir relações funcionais e evolutivas entre sequências, bem como ajudar a identificar similaridade de genes. As sequências a serem comparadas podem ser tanto de nucleotídeos, como de aminoácidos e devem estar no formato FASTA (BLAST, 2013), o qual é composto por um texto com quebras de linha e aproximadamente cinquenta caracteres por linha. A primeira linha de cada sequência inicia com o sinal '>' e o nome da sequência (FORMIGHIERI, 2002).

Por exemplo:

>sequência

```
TACAGTCTGGTGGTGGTTGACTGAGAGTGGTGGTATTTCCGTC
GCCCTAGATGAACTCCAGCAAGCGACGGCGGCTGTCGACGGT
GGTGGTGGCGGTGGCGGCTGGGGTGTGAGAGAGATACCCTTA
TATA
```

O BLAST passa pelas seguintes etapas: primeiramente, são criadas pequenas sequências de tamanho fixo contendo, mais ou menos 3 ou 4 letras. Essa pequena sequência, também chamada de palavra, possui grande similaridade com partes da sequência de entrada. Em seguida, todas as sequências do banco são validadas, realizando alinhamentos exatos entre as palavras e as sequências que estão no banco de dados. Por último, são estendidos os alinhamentos exatos encontrados na etapa anterior, efetuando alinhamentos locais (NORONHA, 2006). O BLAST também possui outros subprogramas que se diferenciam de acordo com o tipo de comparação que se deseja realizar. Um deles é o BLASTP e é utilizado para comparar uma sequência de consulta aminoácido contra um banco de dados de sequências de proteínas. Quando se quer comparar uma sequência de pesquisa nucleotídeos

contra uma base de dados de sequência de nucleotídeos, utiliza-se o BLASTN (PIROOZNIJA, PERKINS, DENG, 2008). No BLASTX são inseridos nucleotídeos como entrada, os quais são traduzidos em sequências de aminoácidos e, depois, comparados em um banco de dados de proteínas. Geralmente, é utilizado para determinar o quadro de leitura no qual uma determinada sequência de nucleotídeos deve ser traduzida em proteínas, além de dar valor biológico às possíveis proteínas (NORONHA, 2006).

Outro subprograma desse grupo é o TBLASTN. Na entrada, é inserida uma sequência de aminoácido que será comparada em banco de dados de nucleotídeos. Geralmente, ele é usado para encontrar genes que ainda não foram descritos devido ao grande número de proteínas desconhecidas.

Finalmente, tem-se o subprograma TBLASTX, em que na entrada é inserida uma sequência de nucleotídeos com a finalidade de obter um valor biológico para sequências de nucleotídeos desconhecidas (NORONHA, 2006).

Uma nova ferramenta para classificar várias sequências de consulta e descobrir a sua relação com outra é o MOLE-BLAST. Ele trabalhar com um *locus* específico a partir de um conjunto de organismos, em vez de sequências, como o genoma inteiro de um organismo ou *contigs* anotadas. Atualmente, MOLE-BLAST só aceita consultas de nucleotídeos (BORATYN, 2015).

Lifschitz (2006) descreve que o algoritmo utilizado no tipo FAST apresenta maior precisão em seus resultados, porém, o BLAST é mais rápido devido a sua heurística, sendo mais popular e mais utilizado.

Escolher um sistema gerenciador de banco de dados para armazenamento e gerenciamento de dados biológicos pode ser uma ótima opção devido à vasta quantidade de informações geradas diariamente nos laboratórios de bioinformática. Bases de dados como EMBL, SWISS-PROT, TrEMBL utilizam como repositório de dados o SGBD Oracle. Quando utilizados como repositório de sequências, SGBD relacionais armazenam tais sequências como cadeias de caracteres. Esses arquivos de texto são armazenados como objetos grandes. O CLOB, por exemplo, é um tipo de dado utilizado para armazenar grandes objetos, com tamanhos superiores a 4GB, baseados em caracteres (FREEMAN, 2005). A versão 10g do Oracle possui como vantagem a incorporação do BLAST e consultas em *Structured Query Language* (SQL)

para pré-filtragem de sequências ou pós-processamento dos resultados (LIFSCHITZ, 2006).

Entre as vantagens da utilização de um SGBD, Elmasri e Navathe (2011) destacam: o controle de redundância de dados, isto é, se for preciso atualizar algum campo sobre um referido dado, essa operação precisaria ser feita em vários arquivos e não em apenas um, além disso, com o armazenamento de dados repetidos, o espaço de armazenamento é desperdiçado; a redundância poderia ainda causar inconsistência nos dados, caso apenas um dado fosse atualizado e o outro não; a restrição que controla o acesso aos dados armazenados, restringindo o acesso ou a alteração de alguns dados importantes; as técnicas de pesquisas para o processamento eficiente de consulta, isto é, os SGBDs geralmente possuem *buffers* ou *caches* de armazenamento próprios para melhorar seus desempenhos e módulos de processamento e otimização de consulta, que são responsáveis por escolher um plano de execução para as consultas; os *Backups* e recuperação, que oferecem recursos para recuperação de dados, quando ocorrem falhas de *softwares* ou *hardwares*; as múltiplas interfaces do usuário; a imposição de restrições de integridade, as quais impossibilitam a inserção de dados com tipos de dados diferentes dos tipos requeridos ou a inserção de maiores quantidades de caracteres do que as especificadas; a possibilidade de garantir padrões, permitindo a definição de padrões para dados, datas, nomes etc.; a utilização de ações de interferências: em bancos de dados relacionais pode-se utilizar *triggers* (um objeto de banco de dados que está associado a uma tabela e que é ativado quando um determinado evento ocorre para a tabela); o *stored procedures* (conjunto de comandos, que ajuda o usuário ou administrador de banco de dados a diminuir uma série de tarefas repetitivas).

Lifschitz (2006) ainda cita outro SGBD que possui a incorporação do BLAST, o *freeware* PostgreSQL. A incorporação do PostgreSQL com o BLAST possui o nome de BlastGres, que representa um segmento de sequência com o tipo de dados *range*, a indicação do início e do fim do segmento da sequência por números e a criação de um tipo *location*, com o qual é associado um valor de *range* ao identificador de uma sequência, podendo, ainda, utilizar um índice para realizar a busca de uma região de uma dada sequência. Contudo, a

desvantagem é que essa sequência será guardada como uma cadeia de caracteres.

Entre os SGBDs com modelo orientado a objeto, destaca-se o Caenorhabditis elegans Data Base (ACeDB). Que possui um módulo de gerenciamento de banco de dados projetado para manipular dados biológicos. Internamente, esse SGBD representa os dados em forma de árvore com formato binário. Nesse banco de dados, a entrada e saída de dados são feitas por arquivos de texto, com sintaxe parecida a linguagem XML (*extensible markup language*), chamados ACE files (LIFSCHITZ, 2006).

O Microsoft SQL Server é um banco de dados robusto, muito utilizado comercialmente em grandes corporações, no entanto, por se tratar de um banco de dados proprietário, ele apresenta alto custo para a aquisição de suas licenças.

Outro SGBD *freeware* muito utilizado em meios acadêmicos e em projetos genomas é o MySQL. E é muito utilizado por ser gratuito, possuir código aberto e possibilitar a armazenagem de dados em tabelas separadas, proporcionando velocidade, flexibilidade, excelente desempenho e estabilidade. Destaca-se pela alta compatibilidade com linguagens como PHP, Java, Python, C# (CSharp), Ruby, C/C++, Visual Basic, Visual C++ (MANZANO, 2011). É pouco exigente quanto a recursos de *hardware*. Em razão das referidas características e vantagens, é importante destacar que o SGBD modelado no presente trabalho é o MySQL.

Um SGBD fornece funções importantes a um banco de dados, como a proteção contra defeitos e/ou falhas de *software* ou *hardware*, a proteção contra acesso não autorizado ou malicioso ao banco de dados e manipulação do banco de dados, possibilitando a realização de consultas, armazenamento e geração de relatórios. Outra função importante de um SGBD consiste em permitir a manutenção do sistema, fazendo com que, com o passar do tempo, o banco de dados evolua à medida que os requisitos mudam (ELMASRI; NAVATHE, 2011).

Entre as ferramentas desenvolvidas utilizando-se SGBD, destaca-se o Eimeria TDB, que é um banco de dados criado como repositório *web* para permitir o acesso a dados de sequência, anotações e análises comparativas. É um recurso integrado de sequenciamento de cDNA e dos dados de anotação de

parasitas do gênero *Eimeria*, que infecta muitas espécies de animais silvestres e domésticos, incluindo aves. A interface *web* da ferramenta foi desenvolvida utilizando linguagem de programação PHP, HTML e linguagens JavaScript, e o banco de dados é MySQL. A ferramenta *Eimeria TDB* é mantida pelo Grupo de Pesquisa em Biologia Molecular de coccídeos no Instituto de Ciências Biomédicas da Universidade de São Paulo, Brasil (RANGEL *et al.*, 2013).

Os bancos de dados da literatura científica, como PUBMED e MEDLINE, também fornecem funcionalidade adicional. Por exemplo, eles podem procurar artigos semelhantes com base na análise de palavra. Sistemas de reconhecimento de texto estão sendo desenvolvidos para extrair automaticamente o conhecimento sobre a função da proteína a partir dos resumos de artigos científicos (NILGES, LINGE, 2014).

Segundo Okura (2002), o estudo de um genoma envolve uma série de etapas até que se consiga chegar a resultados mais concretos para serem aplicados na prática (por exemplo, prevenção ou tratamento de doenças). O primeiro procedimento é determinar a sequência de bases do genoma, identificar e caracterizar seus genes. Os projetos de sequenciamento de genoma atuam exatamente nesse procedimento e seguem, em geral, duas abordagens. A primeira é determinar a sequência completa do genoma alvo e obter seus genes, e a outra, conhecida como sequenciamento EST, procura determinar somente segmentos do genoma que expressam genes.

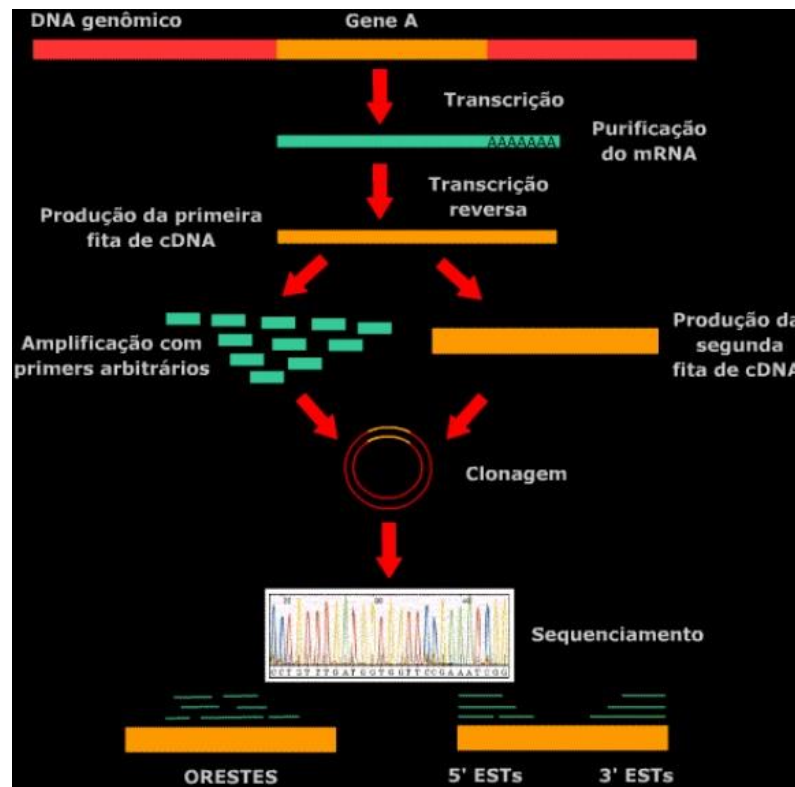
Uma das abordagens empregadas para sequenciar um genoma é determinar a sua sequência de bases, podendo então ser identificados os seus genes (fazê-lo por completo). Sequenciar um genoma significa sequenciar cada uma de suas moléculas de DNA e esse sequenciamento é dividido em duas etapas, sendo a primeira essencialmente biológica, que visa fragmentar a molécula, além de sequenciar individualmente cada um dos fragmentos. Já a segunda etapa envolve a área da computação e objetiva determinar a ordem dos fragmentos, de onde pode ser então deduzida a sequência de bases completas da molécula. Com a sequência de DNA determinada, pode-se, então, passar para o processo de anotação, isto é, para o processo de identificação e análise dos genes. O *pipeline* seguido pelos projetos de genoma completos compreende as etapas de sequenciamento, de montagem e anotação (OKURA, 2002).

Okura (2002) relata que o referido tipo de tecnologia torna o processo de determinação de genes mais rápido, se comparado com a abordagem de genoma completo, principalmente para genomas de organismos complexos, como animais e plantas, em que uma parcela geralmente pequena de seu genoma é compreendida por genes. Um projeto EST pode ser dividido quase que nas mesmas etapas de um projeto de genoma completo, embora possuam características bem distintas. A diferença básica entre eles está no fato de que num projeto EST nada se pode inferir sobre a ausência de genes, pois não há garantia de que todos os genes de interesse serão sequenciados.

Adotam-se, também, abordagens de fragmentar todo o genoma de organismos em pedaços menores e sequencia-los utilizando programas computacionais para montá-los, além de reconstituir a informação genômica inicial (ZHOU et al., 2003). O inconveniente deste método se refere à ação de lidar com sequências de repetição. Muitas vezes, não há nenhuma maneira de saber o tamanho da sequência de repetição ou em qual de muitas posições possíveis de diferentes fragmentos estão sobrepostas (DNAI, 2014).

Em organismos eucariotos, que possuem uma quantidade grande de DNA, utiliza-se uma técnica conhecida como *shotgun* hierárquico, a qual consiste em inserir o DNA inteiro em grandes vetores de clonagem, como cromossomos artificiais de bactérias ou leveduras, e realizar um *shotgun* desses grandes fragmentos de vetores plasmidiais para o sequenciamento. Outra opção é realizar o sequenciamento apenas das regiões gênicas, utilizando informações vindas do RNA mensageiro (mRNA). Nessa abordagem, é realizado uma biblioteca de cDNA, representando o conjunto de mRNAs de uma célula, que é clonado em vetores plasmidiais. Os insertos de cDNA presentes nos referidos vetores são sequenciados a partir de suas extremidades 5' ou 3' (vide a Figura 4), produzindo pequenas sequências que irão representar pedaços dos genes expressos no momento da extração do mRNA da célula em questão (PROSDOCIMI et al., 2003).

Figura 4 - Produção de ESTs.



Fonte: PROSDOCIMI, (2007, p. 36).

As etiquetas de genes expressos (ESTs) são a representação desses pedaços sequenciados e a análise desses genes é uma abordagem bastante utilizada para entender o metabolismo de diversos organismos. Essa estratégia de produção de ESTs, também denominada estratégia de sequenciamento do transcriptoma, pode ser utilizada quando o genoma do organismo não é tão grande. Ela é importante, pois, por meio dela, consegue-se montar bibliotecas de diferentes fases de desenvolvimento do organismo e observar quais genes são expressos em cada momento. Outro fator positivo nessa abordagem é a possibilidade do estudo de como ocorre a expressão diferencial de genes em diferentes órgãos de um mesmo organismo, podendo, assim, entender a função desses órgãos ou como eles realizam funções desconhecidas (PROSDOCIMI *et al.*, 2003).

De acordo com Prodocimi (2007), o sequenciamento de moléculas de DNA em um sequenciador MegaBACE 1000, ocorre, primeiramente, com as moléculas sendo polimerizadas, durante a reação de sequenciamento numa máquina termocicladora, após a eletroforese capilar, e a leitura dos padrões de fluorescência é realizada pelo sequenciador (MegaBACE 1000). Além disso, o

dado bruto é lido por um programa nomeador de bases, como por exemplo, o PHRED, de forma a gerar a sequência de bases e qualidade associada a elas.

O sequenciador MegaBACE é muito utilizado em diversos laboratórios e centros de pesquisa brasileiros. Possui diferentes modelos como: o MegaBACE 500, que aceita uma, duas ou três matrizes de 16 capilares; o *Flexible* MegaBACE 1000, que aceita até 6 matrizes de 16 capilares; o MegaBACE 1000, que possui capacidade de 96 capilares agrupados em matrizes de 16 capilares. Esse equipamento consegue sequenciar 96 capilares em até 2 horas. Já o MegaBACE 4000 consegue analisar, simultaneamente, 384 amostras de DNA. E entre suas características pode-se destacar a produção da saída de dados do *base calling* em formatos compatíveis com PHRED e o processo automatizado de carregamento e reposição da matriz de gel da eletroforese capilar.

O MegaBACE também disponibiliza um *software* com uma interface gráfica amigável, a qual permite o controle e a otimização dos parâmetros de execução como a substituição da matriz de gel, controle de temperatura do instrumento, salvar várias condições de execução e apresentar dados de execução. Os componentes que acompanham o MegaBACE são: instrumento de eletroforese capilar, matrizes de capilar, matrizes de separação para análise do DNA, computador e monitor, e o *software* para manipulação do sequenciador (AMERSHAM BIOSCIENCES, 2003a)

Outros sequenciadores muito usados são os da Applied Biosystems. Os modelos ABI 3730xL, 3730, 3500xL, 3500, 3130xL, 3130 e 310 diferenciam-se pelo número de capilares, sendo o maior número suportado pelo ABI 3730xL (96 capilares). Como características podem-se destacar os recursos de automação do processo, a sensibilidade ótica e polímeros avançados, que permitem a obtenção de dados do sequenciamento de alta qualidade. Outra vantagem se dá por conta da ferramenta de análise que vem integrada no sequenciador. Tal ferramenta possibilita a realização de diversas tarefas como: coleta de dados, gerenciamento das configurações do instrumento, visualização dos dados em tempo real, ferramentas para realização do *base calling*, atribuição de valores de qualidade, editar e imprimir dados do sequenciamento usando o KB *basecaller* (LIFE TECHNOLOGIES, 2013).

No método de sequenciamento *Next Generation Sequencing*, existem mais dois sequenciadores muito utilizados, o Solid, também da Applied Biosystems, e o Illumina. Ambos sequenciam grandes quantidades de pequenos fragmentos em um tempo muito pequeno (ILLUMINA, 2013; LIFE TECHNOLOGIES, 2013).

Após os dados saírem dos sequenciadores de DNA, eles passam por um processo de leitura para a decodificação de cada nucleotídeo e identificação da sequência de DNA gerada e é atribuído um valor de qualidade para cada posição nucleotídica identificada. Esse processo também é chamado de *base calling*. Depois disso, inicia-se a análise computacional com o intuito de gerar cromatogramas para converter cada um em uma sequência de bases (chamadas de *reads*). Para tanto, utilizam-se algoritmos que realizam essa tarefa e entre eles destaca-se o algoritmo PHRED, desenvolvido pelos autores Ewing e Green em 1998, que foi adotado pela comunidade científica como padrão quase unânime.

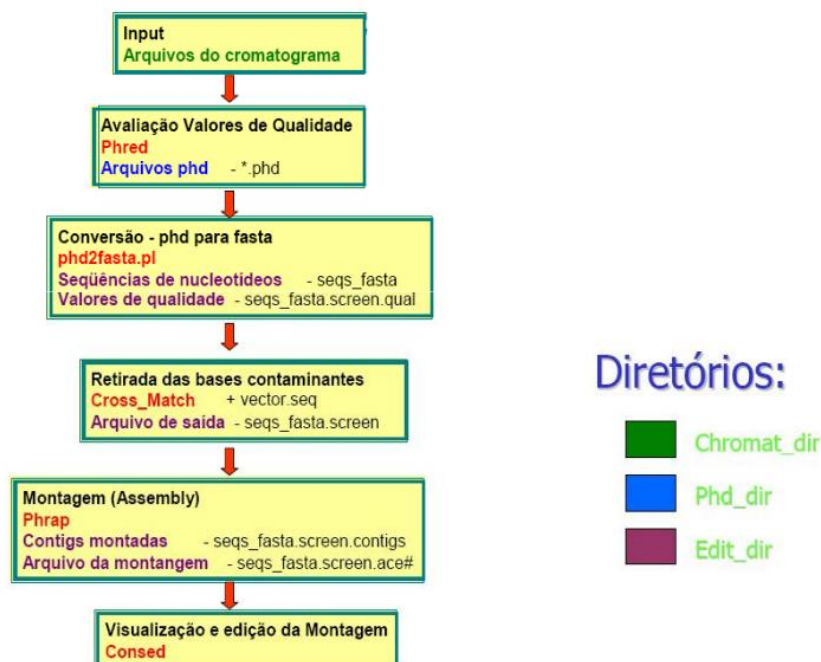
O *pipeline* de montagem dos programas PHRED/PHRAP/CONSED poderá ser observado, a seguir, na Figura 5.

O PHRED além de reconhecer a sequência de nucleotídeos, lendo os dados de fluorescência oriundos do sequenciador, também atribui valores de qualidade às bases constituintes da sequência nucleotídica para indicação da qualidade do *base calling*, e gera arquivos de saída contendo informações sobre o *base call*. Como entrada de dados, o PHRED suporta os arquivos cromatogramas nos formatos SCF (gerados pelos sequenciadores da LI-COR), ABI (gerados pelos sequenciadores da Applied Biosystems) e ESD (gerados pelos sequenciadores da MegaBACE). Como saída, o PHRED consegue gerar arquivos nos formatos FASTA, XBAP, SCF e PHD (PROSDOCIMI *et al.*, 2003; MADEIRA, 2004).

No próximo passo, é necessário transformar os arquivos PHD que saíram do PHRED em arquivos FASTA. Para isso, utiliza-se um *script* desenvolvido em linguagem de programação Perl, chamado *phd2fasta*. Esse *script* é fornecido quando se faz *download* do *script* *phredPhrap*, oferecido no *site* do CONSED (<http://bozeman.mbt.washington.edu/consed/consed.html#howToGet>).

Recomenda-se a utilização desse *script* para geração correta dos dados para posterior utilização do CONSED (GORDON; ABAJIAN; GREEN, 2013).

Figura 5 - Pipeline da montagem dos genomas/transcriptomas.



Fonte: GRUBER, (2001, p. 29).

O Phd2fasta lê arquivos phd e escreve sequência e qualidade em arquivos FASTA (PHD2FASTA, 2013).

Os dados vindos da saída do PHD2FASTA precisam passar por um processo de descontaminação, o qual consiste na remoção ou mascaramento de parte dos vetores de clonagem, onde as sequências de interesse foram inseridas e/ou na remoção de pedaços de DNA adaptadores, utilizados durante a construção das bibliotecas que não representam o DNA ou o cDNA que se deseja analisar. O *software* mais utilizado nessa etapa é o Cross_match, cujo algoritmo consiste em comparar a sequência desejada com um arquivo de sequência de vetores, sendo necessário realizar a entrada desse arquivo que apresenta a sequência dos vetores, onde o programa encontrar similaridade entre as sequências. Ele realizará o mascaramento, acrescentando a letra X na sequência de entrada. Assim os nucleotídeos contendo regiões contaminadas serão alterados para X, não atrapalhando os processos posteriores (PROSDOCIMI *et al.*, 2003).

Após os arquivos estarem sem o vetor e já contendo a identificação das bases e suas respectivas qualidades, é necessário realizar a montagem dos pequenos fragmentos de DNA em sequências maiores, denominadas *contigs*. E para realizar essa tarefa, utiliza-se um *software* de montagem como o PHRAP. Existem outros *softwares*, como CAP3 e o TIGR Assembler, porém, o mais utilizado é o PHRAP, por gerar *contigs* maiores. Como saída do programa, é esperado em projetos genomas a obtenção da sequência montada do *contig* genômico; já em projetos transcriptomas, espera-se como saída as sequências de cada um dos genes expressos. As sequências podem ser visualizadas e editadas, após a montagem, em *softwares* como Consed ou Phrapview (PROSDOCIMI *et al.*, 2003; MADEIRA, 2004). Como saída, o PHRAP solta arquivos com as extensões: **.contigs* (arquivo FASTA que contém os *contigs*), **.singlets* (arquivo FASTA das leituras dos *singlets*) , **.ace* (arquivo que permite a montagem para visualização no Consed) (GORDON; ABAJIAN; GREEN, 1998).

Na Figura 6, abaixo, as cores de fundo indicam a qualidade do *base call* realizado pelo PHRED. Os fundos claros indicam alta qualidade e os escuros indicam baixa qualidade. Ainda no exemplo da Figura 6, apenas os *contigs* A612.r e A333.r possuem alta qualidade, entretanto, a parte amarela da sequência A612.r foi comprimida pelo PHRAP (PHRAP, 2013). Tal Figura mostra os índices de qualidade do PHRED e nota-se que a melhor sequência (A333.r) possui ondas mais altas. Já as ondas mais longas e baixas representam sequências com baixa qualidade (PHRAP, 2013).

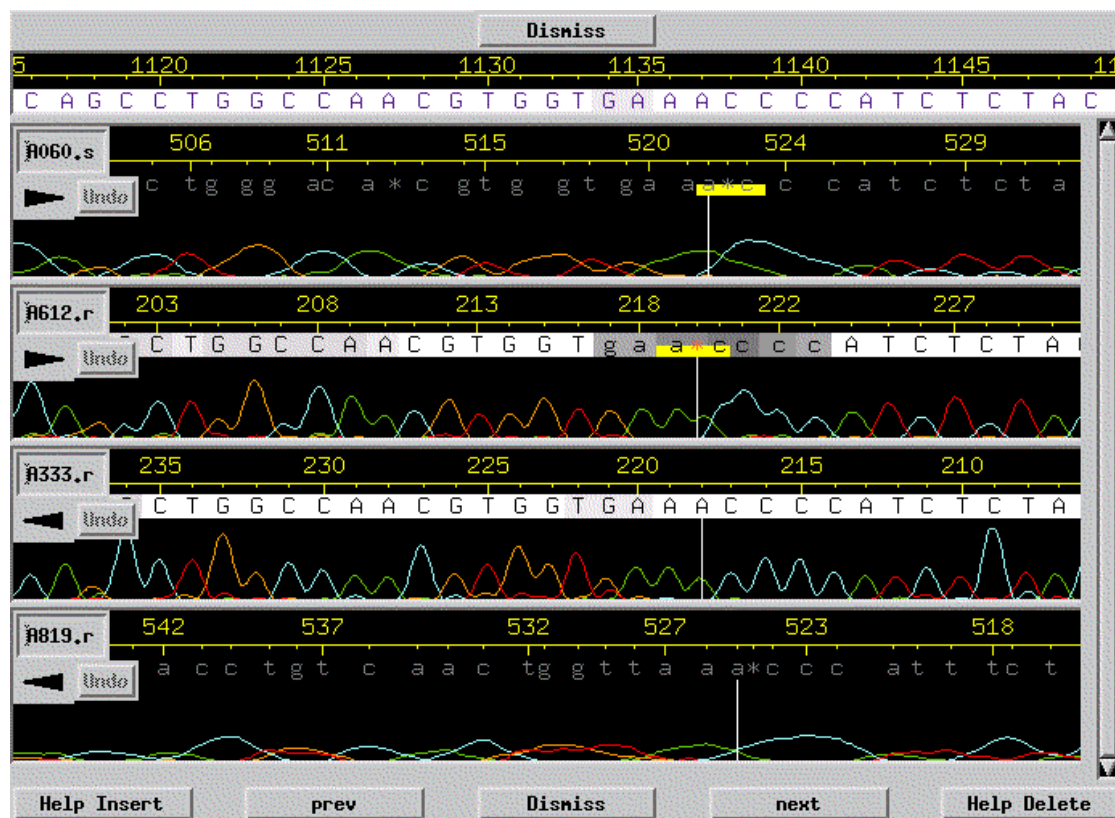
Os programas mais utilizados para montagem de sequências são os PHRED, PHRAP e CONSED, todavia, existem outros *softwares* que realizam essa tarefa, como é o caso do Sequencher, do Chromas e do SeqMan. É de costume dar preferência à utilização de pacotes de *softwares* livres, como os citados anteriormente.

No contexto dos sistemas operacionais, destaca-se o Linux por sua característica acessível com custo de licença zero. Caracteriza-se, também, pela estabilidade e robustez, pois um servidor Linux pode permanecer em funcionamento por meses ou anos.

Atualmente, existem algumas distribuições Linux voltadas para trabalhos em bioinformática, como é o caso do Bio-Linux, que é um sistema

operacional de código aberto, derivado da distribuição Ubuntu Linux e possui mais de 500 *softwares* de bioinformática (FIELD *et al.*, 2006). Outra ferramenta disponível para bioinformática é o DNALinux - uma máquina virtual pré-configurada, com o sistema operacional Linux e vários *softwares* de bioinformática já instalados. A vantagem da utilização dessa máquina virtual é a possibilidade de utilizar o sistema operacional Linux a partir de um computador com sistema operacional Microsoft Windows instalado (BASSI; GONZALEZ, 2007).

Figura 6 - Índices de qualidade do PHRED no Consed.



Fonte: <http://www.phrap.com/consed/>. Acessado em 27/05/2013.

Os profissionais da área de bioinformática além de saberem utilizar os *softwares* já desenvolvidos, precisam desenvolver seus próprios aplicativos para auxiliar nas soluções dos problemas enfrentados no dia a dia. Existem diversas linguagens de programação, cada uma com um conjunto de comandos específicos utilizados para criar uma interface homem *versus* máquina.

No ramo da bioinformática, uma das linguagens de programação muito utilizada é a Perl (PROSDOCIMI *et al.*, 2003) - criada por Larry Wall. Quando lhe foi delegada a função de construir um sistema de controle e

administração com o objetivo de produzir relatórios para rede de computadores Unix, Larry não estava satisfeito com as linguagens de programação da época e decidiu construir a sua (WALL, CHRISTIANSEN, ORWANT, 2001). Unindo características das linguagens awk e sed e com a estrutura oferecida pelos *scripts shell*, ambos oriundos do sistema operacional Unix. Larry tinha como objetivo construir uma linguagem de programação simples (para que programas fossem prototipados rapidamente), além de poderosa e flexível, podendo manipular os recursos dos computadores. Uma característica interessante sobre essa linguagem é o fato de ela ser *open source*, o que permite desenvolvedores cooperarem modificando e estendendo a linguagem, com seu uso totalmente gratuito. Outra característica é a linguagem ser portátil, ou seja, ela pode ser portada de um sistema operacional para outro, sem nenhuma modificação (DEITEL *et al.*, 2002).

Varias funcionalidades podem ser adicionadas à linguagem por meio de módulos que são escritos por desenvolvedores pelo mundo. Os módulos e suas respectivas documentações são armazenados em um centro de distribuição, o *Comprehensive Perl Archive Network* (CPAN) (DEITEL *et al.*, 2002) e, dentro desse contexto, foi desenvolvido um módulo Perl para auxiliar na bioinformática, chamado BioPerl (STAJICH *et al.*, 2013). Hoje em dia, existem também *toolkits* para outras linguagens de programação com foco na bioinformática, como é o caso do BioPerl, BioJava, BioPython, BioPHP, BioRuby.

É importante ressaltar também os trabalhos relacionados à criação de *pipeline* para análise genômica, no qual ocorre a divisão de um processamento sequencial em etapas. Os dados gerados pelo processamento de uma etapa são entrada para a próxima etapa, formando-se, desse modo, uma corrente de elementos de processamento (HOHPE, WOOLF, 2003).

O BioNotes é um sistema distribuído de anotações biológicas que possui como objetivo facilitar a interpretação dos dados que a comunidade da área de biologia dispõe. Ele foi implementado na Pontifícia Universidade Católica do Rio de Janeiro em linguagem de programação Java. A interface *web* foi escrita em JSP e o banco utilizado foi o Oracle 9i. As anotações são obtidas de diversas fontes externas, como o GenBank, PIR, SWISS-PROT, e de diversos aplicativos como BLAST, PHRED, PHRAP, GLIMMER, tRNAScan, Crossmatch. As anotações geradas pela execução desses aplicativos são transformadas para

o formato XML e, depois, armazenadas em um *data warehouse* (LEMOS; SEIBEL; CASANOVA, 2003).

O MAGIC-SPP é um pacote de processamento de sequência de DNA que consiste em um banco de dados relacional, um *pipeline* em Perl, e interfaces implementadas ou como JavaServer Pages ou como uma interface de usuário gráfica Java. O banco de dados é um repositório e também controla o processamento dos arquivos de rastreamento. Outro exemplo é o MAGIC-SPP inclui uma interface administrativa, um sistema de gerenciamento de informações laboratoriais e interfaces para sequências, explorando monitoramento de controle de qualidade e resolução de problemas relacionados com as atividades de sequenciamento. O algoritmo de corte sequência emprega novas funcionalidades concebidas para melhorar o desempenho em relação a preocupações como ligantes concatenadas, a identificação da posição inicial esperada de um vetor de inserção e estender o comprimento útil de sequências aparadas, completando as regiões curtas de baixa qualidade. Ele oferece combinação única de características que permitem a administração por um biólogo com pouco ou nenhum conhecimento de informática (LIANG *et al.*, 2006).

O ChromaPipe é um *pipeline* para análise, controle de qualidade e gestão de uma linha de sequenciamento de DNA. O usuário pode interagir em cada passo do *pipeline* e configurar os parâmetros que serão usados. O serviço inclui a execução de amostras em grande escala para usuários, realizando reações de sequenciamento de DNA e, para usuários de médio e pequeno porte, participação em projetos genomas parciais ou totais. A aplicação também inclui o acompanhamento de dados para a melhoria contínua da qualidade (relatórios por placa, o mês e o usuário) pela equipe de sequenciamento. Permite diferentes análises dos cromatogramas, assim como a visualização de regiões de boa qualidade (OTTO *et al.*, 2008).

Segundo Forment *et al.* (2008), EST2uni é uma ferramenta para análise EST automatizada e para criação de banco de dados com uma interface *web* de mineração de dados. É uma ferramenta de análise EST capaz de converter, de uma forma automática, um conjunto de cromatogramas ou sequências simples em um banco de dados EST estruturado e com uma interface *web*. A análise do *pipeline* de EST inclui programas de pré-processamento, agrupamento e anotação padrão, e o *software* é modular,

facilitando, assim, a incorporação de novos métodos e análises. As opções de serviço são facilmente adaptadas às necessidades locais, basta modificar um arquivo de texto de configuração. Uma vez configurado, o *pipeline* é executado e o banco de dados preenchido com as análises de resultados em tempo real.

Uma das principais contribuições do presente trabalho consiste no desenvolvimento de um *pipeline* utilizando uma combinação de ferramentas de *software* de código aberto, configurada para elucidação da regulação genética do metabolismo secundário de plantas medicinais. Os programas que compõem o referido *pipeline* são responsáveis por agilizar a extração do Vetor de etiquetas de sequências expressas (ESTs) de busca de dados *on-line* para identificar ESTs e funcionalidades de genes-alvo.

Neste trabalho, foi escolhida para validar a eficiência do *pipeline* a biblioteca de cDNA da *Myracrodouon urundeuva* (Aroeira-do-Sertão), sendo alvos da investigação os genes expressos em folhas dessa espécie medicinal. O interesse pelo estudo da *M. urundeuva* é decorrente de ensaios farmacológicos, nos quais foi comprovada eficácia terapêutica de metabólitos biossintetizados nas folhas -atividades estas correlacionadas à presença de chalconas, metabólitos secundários da classe dos fenilpropanoides (CAMPOS, 2008).

1.1 JUSTIFICATIVA

O crescimento exponencial do número de novos genes descobertos no projeto genoma e a construção de arranjos de DNA em membranas propiciaram uma nova abordagem nos estudos da regulação gênica, tornando possível o monitoramento dos níveis de transcritos de um grande número de genes simultâneos.

No sequenciamento do genoma de uma espécie torna-se possível comparar os resultados com genomas de outras espécies, identificando-se seus parentescos bem como os detalhes da sua evolução. Neste contexto, as ferramentas de informática dão suporte aos projetos no campo da biologia molecular, proporcionando agilidade e otimização na decodificação de genes e de suas características regulatórias, ou seja, suas funcionalidades.

Uma das principais contribuições do presente trabalho consiste no desenvolvimento de um *pipeline* computacional, composto por ferramentas configuradas para o recebimento de arquivos de sequenciamento, identificação

da qualidade das sequências, filtragem do vetor, montagem dos fragmentos de DNA e comparações dos referidos fragmentos com sequências de DNA de organismos conhecidos.

Sobretudo as técnicas de informática aliadas às ciências como biologia, física, biotecnologia, química e genética propiciam a obtenção de informações detalhadas da herança genética de uma determinada planta e sua manipulação a favor da saúde humana com a produção biotecnológica de ativos. Dessa forma, a tecnologia da informação tornou-se essencial no meio científico, pois o sucesso da pesquisa é baseado no conjunto de informações valiosas que viabilizam a tomada de decisão. Pode-se dizer que a informação representa inteligência competitiva, auxilia na identificação de nichos de oportunidades para pesquisa e é um diferencial aos pesquisadores que desejam realizar contribuição intelectual relevante e inovadora aos seus trabalhos científicos.

1.2 OBJETIVO

Objetiva-se automatizar os processos de cada programa que é executado no *pipeline* e também criar um sistema para *web* baseado em linguagem de programação Java e banco de dados MySQL. Tal sistema proporcionará ao usuário final a usabilidade bem como a simplicidade na análise dos resultados gerados pelo *pipeline*, proporcionando um ambiente integrado a partir de uma base de dados.

1.2.1 Objetivos Específicos

- Elaborar um *pipeline* para tratar e caracterizar a expressão de genes de fenilpropanoides em folhas de *M.urundeuva* (aroeira-do-sertão).
- Instalar e configurar ferramentas de bioinformática já existentes com o intuito de compreender a análise das sequências eletroesferograma..
- Criar um programa para automatizar o processo manual de renomear arquivos gerados pelo sequenciador automático.
- Modelar e criar um de banco de dados para armazenamento, manipulação e gerenciamento das sequências.

- Implementar um sistema *web* que fará a função de automatizar cada uma das etapas do *pipeline* e disponibilizará consultas e informações do banco de dados e ESTs.
- Criar uma interface de comunicação via *web* entre o banco de dados local e o banco de dados Blastx do NCBI com o intuito de traduzir sequências de nucleotídeos dos *contigs* e *singlets* para proteínas.
- Implementar uma aplicação *mobile* (APP) para Android com a finalidade de viabilizar acessos aos resultados dos projetos em *smartphones* e *tablets*.
- Criar um Webservice (Servidor) na plataforma C# (CSharp) com o objetivo de publicar as consultas geradas no sistema WISE em uma aplicação *mobile*.

2. ESTRATÉGIA METODOLÓGICA

A análise das informações geradas nos experimentos de biologia molecular é desafiante não só pelo grande volume de dados, mas também, pela complexidade das tarefas de identificação de genes expressos e predição de suas respectivas funcionalidades. Nesse cenário, para cada projeto de acordo com o volume de dados, há aplicação de um *pipeline* em bioinformática específico com o intuito de facilitar a fase de análise de sequências geradas. Partindo desse princípio, este capítulo descreve a estruturação de um *pipeline* que utiliza a combinação de ferramentas de *software* com o objetivo de montar sequências de *M. urundeuva* geradas por um sequenciador automático Megabace 1000. Faz referência ao processo de automatização dessas montagens de sequências, que fornecem ao biólogo molecular facilidade e agilidade na interpretação e predição de suas funcionalidades. Aborda as fases de modelagem e implementação de um sistema denominado WISE, o qual tem a função de automatizar o referido *pipeline* e armazenar os resultados em um SGDB (Sistema Gerenciador de Banco de Dados), facilitando a recuperação das informações pelos usuários, bem como, descreve a implementação da versão *mobile* do sistema WISE.

As sequências para análise foram adquiridas por meio do projeto Construção de Macroarranjos e Rastreamento Randômico de Biblioteca cDNAs para avaliação da expressão coordenada de genes de fenilpropanoides em órgãos de *M. urundeuva*. Para tanto, os biólogos moleculares da Unidade de Biotecnologia da Universidade de Ribeirão Preto (UNAERP) extraíram o ácido ribonucleico (RNA) total das folhas de *M. urundeuva* (veja a Figura 7 a seguir), desses RNAs foram selecionados os RNAs mensageiros (mRNAs) e convertidos em fragmentos de DNA (cDNAs), que são complementares a esses mensageiros. Os fragmentos de cDNAs foram clonados em vetor pGEM-T *vector* (Promega) e inseridos por meio de transformação química em células bacterianas *Escherichia coli* DH5 α . Após a extração plasmidial e amplificação por reação em cadeia da polimerase (PCR - *Polymerase Chain Reaction*), tais fragmentos foram submetidos a uma reação para marcação com cromóforos adequados ao sequenciamento.

Figura 7 – Fotografia das folhas da *M. urundeuva*.



Fonte:

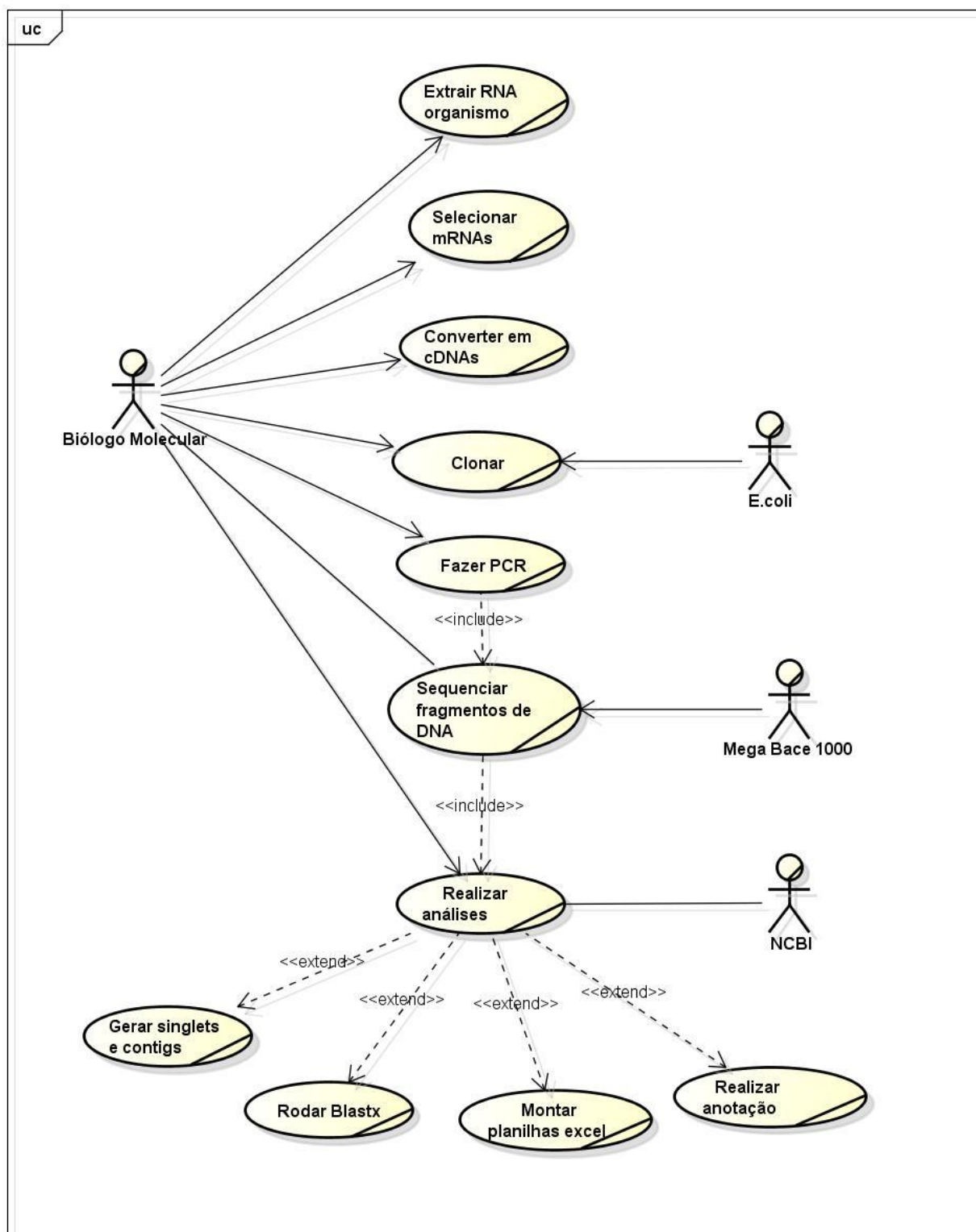
http://www.cnip.org.br/banco_img/Aroeira/myracrodrunurundeuvaallemo3.html

Acessado março 2015.

Dispostos numa placa, vários fragmentos de DNA das folhas de *M. urundeuva* foram sequenciados de forma simultânea por um sequenciador capilar automático de DNA MegaBACE 1000. Cada fragmento gerou um cromatograma, que se trata de um gráfico contendo quatro curvas coloridas, cada uma correspondendo a uma das quatro bases nitrogenadas, A (Adenina), C (Citosina), G (Guanina) e T (Timina). Um pico no gráfico em determinada posição do fragmento identificou uma base. Às bases não identificadas, a letra N (*uNknown*) foi associada à posição correspondente. Nesse momento, foram gerados os arquivos cromatograma de 10 placas sequenciadas da *M. urundeuva*. E, neste projeto foram geradas 1.150 sequências para submissão das análises computacionais.

Na Figura 8, é possível observar o diagrama de *use case* de negócio, representando os processos realizados pelo biólogo molecular. O diagrama de *use case* de negócios tem o propósito de representar um processo que fornece o valor para o ator de negócios, e descreve que o processo faz (JACOBSON, SPENCE, BITTNER, 2011).

Antes de continuar descrevendo o processo de análise computacional, é importante ressaltar que os biólogos moleculares da Unidade de Biotecnologia da UNAERP, além de fazerem os experimentos em bancada, faziam também: a análise e a avaliação da qualidade dos picos de cada amostra gerada pelo eletroferograma (sequenciador), o agrupamento de sequências similares *contigs*, *singlets*, entre outras. Informações sobre sequências gênicas e de proteínas, com funções conhecidas ou não, são oriundas desses estudos, no entanto, não havia uma padronização no armazenamento dessas informações. Elas eram armazenadas individualmente no computador do pesquisador, dessa forma, não ocorria o compartilhamento delas. Os dados eram gerados em diferentes formatos, havia duplicidade de informação e ineficiência no sistema de *backup*. Na maioria dos casos, tais informações não estavam disponíveis e organizadas em uma base de dados, com controle de acesso restrito aos pesquisadores. Esses fatos, muitas vezes, produziam a redundância de informações e de ações previstas em novas propostas de projetos, gerando dificuldade no avanço de alguns estudos devido à falta de um sistema integrado para gerenciamento.

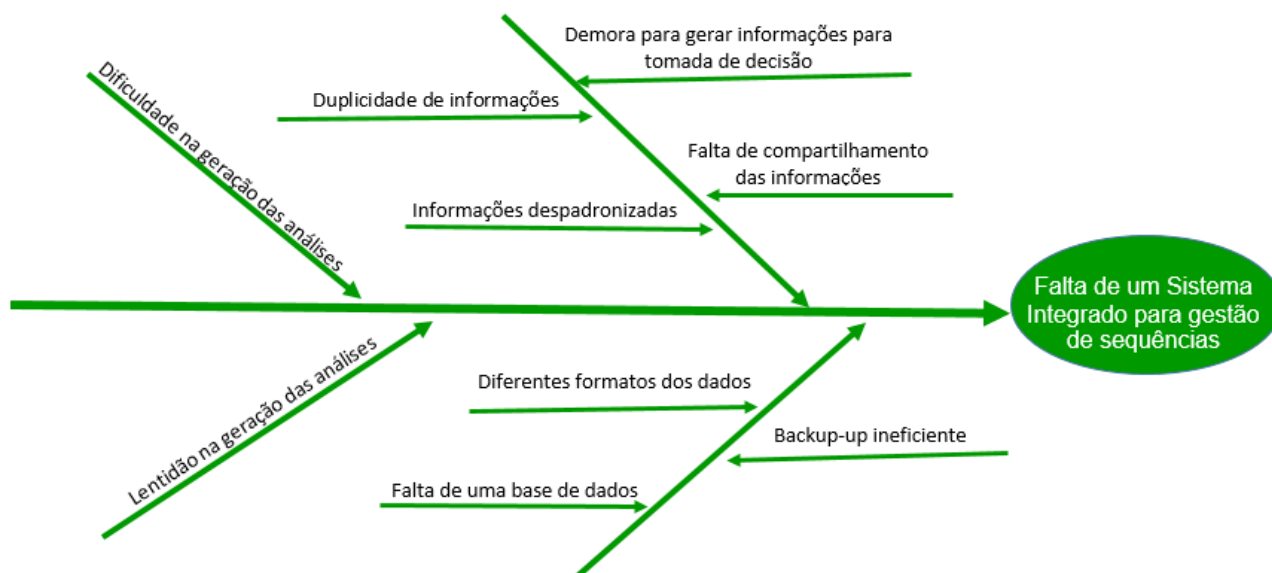
Figura 8: Diagrama de *use case* de negócio.

Fonte: Autoria própria.

Na Figura 9 o Diagrama de Ishikawa demonstra a representação das causas para o problema identificado na rotina dos biólogos moleculares na fase de análise das sequências. O diagrama de Causa e Efeito, ou diagrama "espinha de peixe", é uma ferramenta que tem a finalidade de descobrir as possíveis causas para um efeito. Ela pode ajudar na identificação imediata de causas e apontar as possíveis áreas potenciais para uma maior exploração e análise. Um dos passos para a construção do referido diagrama é anotar o efeito a ser investigado e desenhar a seta "espinha dorsal" para ele. Após, o próximo passo é ramificando em mais direções até identificar cada possível causa (VORLEY, 2008).

A descrição, a seguir, está relacionada às próximas etapas realizadas após o sequenciamento no MegaBace1000.

Figura 9 – Diagrama de Ishikawa.



Fonte: Autoria própria.

Os arquivos de sequências no formato cromatograma, gerados pelo sequenciador MegaBace 1000, foram encaminhados para a bioinformática. A partir deles, deu-se início aos testes com os programas (*software*) de montagem de sequências com o intuito de definir o *pipeline* do projeto *M. urundeuva*. Foram testados vários programas e *scripts*, no entanto, optou-se por realizar as análises

das sequências por meio de programas e *scripts* desenvolvidos em linguagem de programação *Practical Extraction and Report Language*, acrônimo Perl, sob a plataforma do sistema operacional Linux distribuição Fedora.

A linguagem de programação Perl foi escolhida por ser simples executá-la e ser multiplataforma, além de possuir diversos módulos específicos para as áreas científicas e de permite implementar aplicações personalizadas para a resolução de problemas relacionados à bioinformática. O sistema Operacional Linux caracteriza-se pela segurança, confiabilidade e por seu uso gratuito pelo consórcio GNU's Not Unix (GNU). A distribuição do sistema operacional Linux, utilizada para executar o *pipeline* criado na presente pesquisa, foi o Linux baseado na distribuição Fedora versão 17. Utilizou-se, também, o sistema Operacional Microsoft Windows 7.0 para a implementação dos programas Sequence Name Converter e o Online Blast XLS Reporter, os quais foram implementados na linguagem de programação C# (CSharp), que é uma linguagem orientada a objetos criada pela empresa Microsoft. Na criação do *pipeline*, utilizou-se os programas e *scripts* descritos no quadro 1.

Antes de iniciar as análises com o *pipeline* de programas, foi necessário designar todos os arquivos de sequências com diferentes nomes ou extensão de arquivo, os quais foram gerados pelo sequenciador MegaBACE 1000. No processo de nomeação dos arquivos de sequências, implementou-se o programa Sequence Name Converter em linguagem C# (CSharp), que é capaz de renomear sequências de um padrão para outro desejado, como por exemplo: SEQ491A_A12_BASF_096.ab1 renomear para SEQ_A12_096.ab1. Para tal procedimento, o programa demanda da utilização de três entradas: a primeira é o arquivo no formato original, denominado *input*, onde se dá a entrada padrão do nome das sequências, antes do procedimento de renomeação; a segunda é a entrada do formato de saída, denominado *output*, que é o padrão do nome das sequências após a renomeação; a terceira é a indicação do diretório (local), onde as sequências serão armazenadas. Ambos os padrões (*input* e *output*) podem conter constantes iguais para todas as sequências do diretório, tal como descrito no exemplo anterior (SEQ), bem como variáveis que alteram de sequência para sequência. Para definir uma variável, é necessário utilizar o nome da variável entre chaves e ela pode ser de três tipos: apenas números - esse tipo é definido com "N=" antes do nome da variável (Ex:

{N=VARIABEL1}); apenas letras - “L=” antes do nome da variável define esse tipo (Ex: {L=VARIABEL2}); ou variável mista - esse tipo de variável armazena uma cadeia de caracteres com letras e números e é definida apenas pelo nome entre chaves (Ex: {VARIABEL3}).

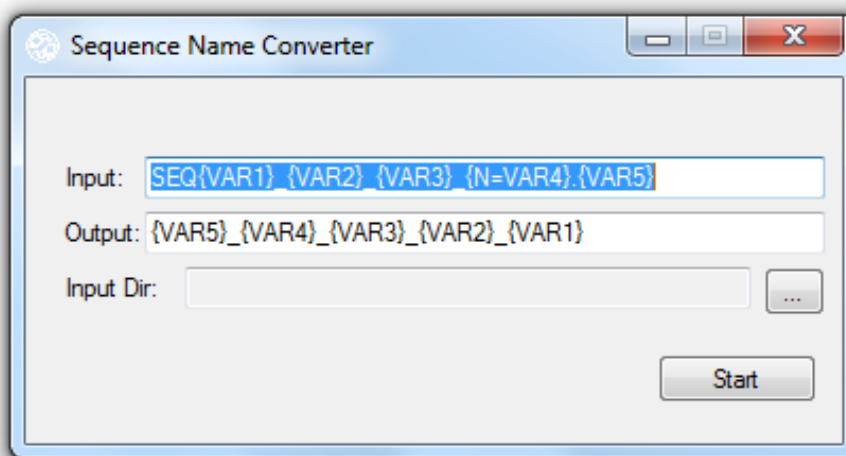
Quadro1: Relação de programas usados no *pipeline* com a descrição, funcionalidades e parâmetros.

Programa utilizado no <i>pipeline</i>	Descrição	Funcionalidade	Parâmetros utilizados
PHRED	Os sinais de fluorescência são detectados e convertidos na forma de sequências de bases nucleotídicas; ao final, são geradas sequências no formato FASTA	A função do programa PHRED é realizar a trimagem que avalia os eletroferogramas gerados nos sequenciamentos dos DNAs, observando-se a qualidade dos picos correspondentes a cada base sequenciada e conferindo um valor de probabilidade de erro a cada uma das amostras (GREEN, EWING, 2002).	-id (leitura dos arquivos de entrada do diretório) -pd (gravar os arquivos .phd.1 para o diretório de destino)
PHD2Fasta	As sequências FASTA geradas são agrupadas em um único arquivo	Nessa fase, as regiões de maior qualidade são identificadas. As sequências são, então, processadas em dois arquivos de texto, sendo um arquivo de <i>reads</i> e outro de qualidade.	-id (leitura dos arquivos de entrada do diretório) -os (nome do arquivo de sequência de saída) -oq (nome do arquivo de sequência de qualidade)
CROSSMATCH	Tem a função de delimitar as regiões de cDNA pertencentes ao vetor.	Permite o reconhecimento da sequência do vetor, no qual estão inseridos os fragmentos de cDNA da espécie alvo de análise.	-minmatch 10 -minscore 20
MASK.pl	Executa o <i>script</i> de retirada do vetor	Uma vez demarcado com um X, o vetor pode ser extraído, obtendo-se assim, sequências FASTA dos insertos limpos, ou seja, livres do vetor.	Arquivo de sequência e o Arquivo de qualidade
CAP3	Realiza o agrupamento das sequências transcritas, representando unigenes.	Agrupamento de sequências similares (<i>contigs</i>) e os <i>singlets</i> . A partir daí, têm-se as sequências de genes expressos que devem ser correlacionados às proteínas para predição de funcionalidade.	<i>Options default values</i> onde o valor -b foi setado em 20

Fonte: Autoria própria.

No presente projeto, o padrão utilizado para etiquetar as sequências foram iniciais de identificação do organismo, tipo do organismo, número da placa de depósito do DNA, letra e número de identificação do poço da placa, e a extensão .scf. Por exemplo, o nome original da sequência era SEQ495A_A01_B015.AB1, após executar o programa para renomear o nome atual, passou a ser AROFOPL01B15.scf sendo, ARO = aroeira, FO = folha, PL01 = placa01, B15 = letra e número de identificação do poço da placa .scf = extensão do arquivo. A Figura 10 apresenta a interface do programa Sequence Name Converter.

Figura 10: Interface do programa Sequence Name Converter



Fonte: Autoria própria.

Todas as sequências do diretório têm que ter o nome no formato do padrão definido no *input*, caso contrário será gerado um erro. A única diferença entre o padrão de origem e de destino é que, neste último, não é necessário especificar o tipo da variável. Para isso, o programa foi implementado conforme as seguintes etapas: identificaram-se os elementos (variáveis e constantes) dos padrões de *input* e *output*, foram verificados os nomes de cada sequência para determinar se estavam enquadrados no padrão de *input*, foi capturado o valor das variáveis e realizada alteração no nome para o padrão definido no *output*.

Após realizar o processo de nomeação dos arquivos, eles foram transferidos para uma pasta denominada chromat_dir. Depois, foram submetidas as sequências ao programa PHRED, no qual os sinais de fluorescência são mapeados em sequências de bases e, ao final, são geradas sequências no formato FASTA.

A estrutura de diretórios criada no projeto *M. urundeuva* foi:

/aroeira/chromat_dir

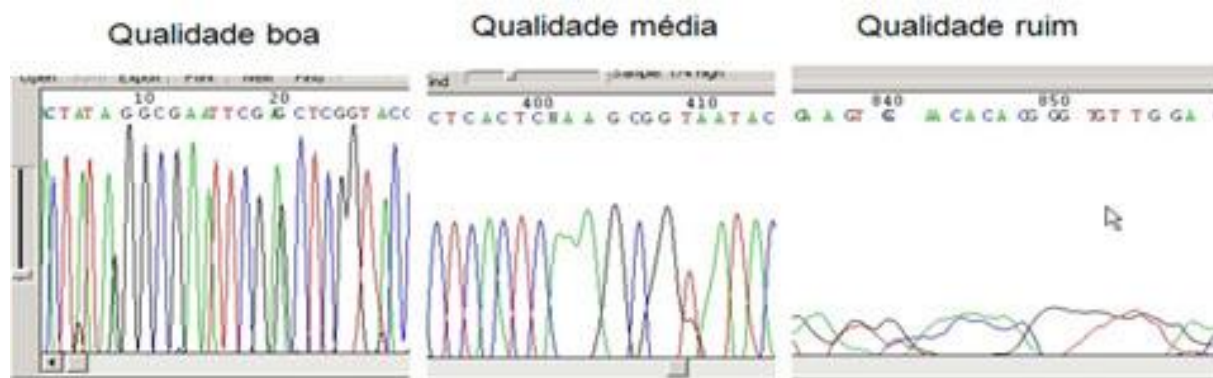
/phd_dir

/edit_dir

Os arquivos no formato cromatograma foram transferidos para o subdiretório chromat_dir. A seguir, executou-se o programa PHRED a partir do subdiretório edit_dir. Durante a execução, os arquivos com extensão.phd foram automaticamente criados e armazenados no subdiretório phd_dir.

A função do programa PHRED é realizar a trimagem que avalia os eletroferogramas gerados nos sequenciamentos dos clones, observando a qualidade dos picos correspondentes a cada base sequenciada (vide a Figura 11) e conferindo um valor de probabilidade de erro a cada uma das amostras. Quanto mais altos e agudos os picos, mais qualidade têm.

Figura 11 – Exemplo de um cromatograma gerado pelo sequenciador.



Fonte: (CARAZZOLLE, 2008)

http://www.lge.ibi.unicamp.br/lgeextensao2008/extsup/anali_mont.pdf.

O valor de qualidade está relacionado com a probabilidade de erro do *base-call* através da fórmula: $q = -10 \times \log^{10}(p)$, onde q é a qualidade e p a probabilidade de encontrar uma base errada (EWING et al., 1998; EWING, GREEN, 1998). Na maioria dos casos, emprega-se como padrão aceitável de qualidade o valor 20, que corresponde a aproximadamente 99% de certeza da base indicada. Aqui, foram consideradas as bases com qualidade igual ou maior que 20.

Segue um exemplo do resultado:

```
BEGIN_SEQUENCE AROFOPL09F10R1.scf
BEGIN_COMMENT
CHROMAT_FILE: AROFOPL09F10R1.scf
ABI_THUMBPRINT: 0
PHRED_VERSION: 0.990722.f
CALL_METHOD: phred
QUALITY_LEVELS: 99
TIME: Tue Apr 2 20:31:19 2013
TRACE_ARRAY_MIN_INDEX: 0
TRACE_ARRAY_MAX_INDEX: 2053
TRIM: 35 89 0.0500
CHEM: term
DYE: ET
END_COMMENT
BEGIN_DNA
t 11 1
c 11 11
a 15 18
a 18 27
c 5 34
a 5 43
c 5 51
```

No cabeçalho, contém o nome do arquivo AROFOPL09F10R1.scf, o método utilizado para o sequenciamento e para o *base-callings*. Os valores que aparecem após a palavra “TRIM” devem ser considerados como região de alta qualidade da base 35 até a base 89. O terceiro valor representa a probabilidade de erro do *base-call* e nos valores que aparecem após “BEGIN_DNA” estão os dados organizados em três colunas, representando a base “chamada”, o valor de qualidade e a posição do pico.

Em seguida, foi executado o programa PHD2FASTA, que tem como entrada os arquivos com extensão .phd e gera um arquivo de sequência no formato FASTA, além de um arquivo com os valores de qualidade determinado pelo Phred para cada base lida (arquivo com extensão .qual). Isto é, o phd2fasta converte os arquivos phd na sua respectiva sequência de nucleotídeos no

formato FASTA, o qual é a forma mais habitual de representar sequências de nucleotídios ou sequências de aminoácidos (PEARSON, LIPMAN, 1988).

A sequência foi identificada como AROFO-PL01G02.scf e, em seguida, o corpo da sequência. Todo cabeçalho de uma sequência FASTA inicia com o sinal de ">" (maior) seguido pelo texto da sequência.

Exemplo formato FASTA:

```
>AROFO-PL01G02.scf
TCCGTAAGCAGGTCATATGTATTTTCACCTTGGCAAACATAAAGTGATGTNC
TTAAGGATAAGTCAGAGTACTATGCGTGTGGNGGTTACGTGTACACCCATAA
AGACACAGCGATAAGTATCGGTGTTAGCGTTAGCACTCAACCTAAGTGAGTA
ATACACGTTACTGTTCGATGACCACACATTCGTGATACATTGAGTCTCGTACT
```

No arquivo com extensão .qual estão os valores de qualidade associados com cada base. O formato é parecido com do arquivo .fasta, porém, no lugar da sequência de bases, ele tem os valores de qualidade na escala phred separados por espaços.

Exemplo formato .qual:

```
>AROFO-PL01G02.ab1 PHD_FILE: AROFO-PL01G02.ab1.phd.1
51 42 42 42 43 42 42 42 42 46 56 56 47 56 56 56 56
56 56 56 56 56 42 42 35 28 28 33 33 48 48 56 48 44
44 44 44 48 48 56 56 48 44 44 44 56 44 44 47 42 33
25 17 19 13 13 13 20 33 35 47 44 44 44 44 56 46 46
```

Após as sequências serem geradas, foram submetidas ao programa CROSSMATCH, que identificou as regiões pertencentes ao vetor, ou seja, fez a identificação da região do vetor através da comparação entre as sequências FASTA da *M. urundeuva* e o arquivo tipo texto contendo o vetor, além de marcar com a letra X a região identificada com o vetor na sequência FASTA.

Exemplo de (o vetor X está na cor verde):

```
>AROFO-PL04H02.ab1 CHROMAT_FILE: AROFO-PL04H02.ab1
PHD_FILE: AROFO-PL04H02.ab1.ab1.phd.1 CHEM: term DYE: big
TIME: Mon Jun 30 17:49:51 2013
CAGGGAATACTCCTATAGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCTTAAAGGTATCAACGC
AGNANAGGGANAGNAAGGATCCTTTAGAGAGAGCTCTTTCTTGATTC
GGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGCGATTTGTCTG
ATTCCGTTAACGAACGAGACCTCAGCCTGCTAACTAGCTATGCGGAG
TCCCTTCGCGGCCAGCTTCTTAGAGGGACTACGGCCTTTTAGGCCGC
```

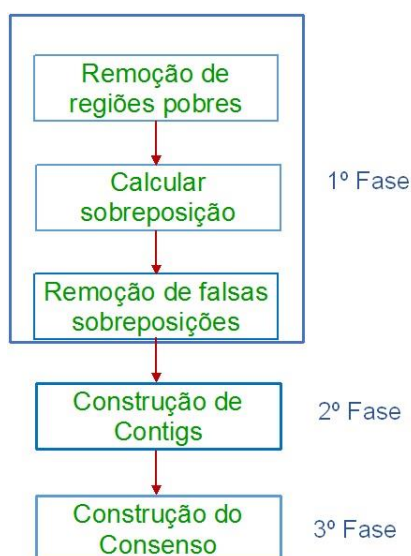
Posteriormente, executou-se o *script* MASK.pl para identificar, filtrar e retirar a sequência correspondente ao vetor. Foram obtidas, assim, as sequências FASTA livres do vetor.

O próximo passo foi processar as sequências no programa CAP3, que utiliza os valores de qualidade de base produzidos pelo programa PHRED e realiza as sobreposições entre as leituras, alinha várias sequências e gera as sequências de consenso, formando os *contigs* e os *singlets*.

O CAP3 faz uso de *forward-reverse constraints* para corrigir erros de montagem da sequência e ligar *contigs* separados por *gaps*. Essa capacidade permite o CAP3 encontrar erros provenientes de repetições. A remoção de regiões pobres é usada para garantir que apenas as regiões boas dos fragmentos sejam montadas. E sobreposições falsas são identificados e removidos.

Em outra fase, as leituras são unidas para formar *contigs* em ordem decrescente de pontuações de sobreposição. Depois, um alinhamento de múltiplas sequências de leituras é construído, além de uma sequência de consenso, juntamente a um valor de qualidade para cada base é calculado para cada *contig* (HUNAG E MADAN, 1999). A Figura 12 representa as três principais fases do algoritmo de montagem utilizado pelo programa CAP3.

Figura 12 - Representação das três principais fases do algoritmo de montagem do CAP3.



Fonte: (HUNAG E MADAN, 1999, Adaptado pela autora).

A partir daí, obtêm-se as sequências de genes expressos ainda sem funções conhecidas. A identificação deles é feita pela comparação das sequências de cada *contig* e de cada *singlet* com genes já conhecidos, que estão disponíveis em bancos de dados *on-line* no portal do NCBI.

Com os arquivos de *singlets* e *contigs* gerados pelo programa CAP3, foram feitos os alinhamentos com o uso da ferramenta de comparação de sequências de nucleotídeos com proteínas BLASTx, disponível *on-line* no portal do NCBI. Se há alto grau de alinhamento entre um *contig* e as proteínas no banco de dados, significa que há propabilidade que o *contig* tenha gene homólogo aos encontrados no alinhamento.

Com o intuito de automatizar e agilizar o alinhamento no BLASTx implementou-se o programa Online Blast XLS Reporter e nessa implementação utilizou-se a linguagem de programação C# (CSharp). O primeiro procedimento foi estabelecer a conexão com o navegador de Internet e o servidor no portal NCBI. Segundo Tanenbaum e Wetherall (2012), para um navegador entrar em contato com um servidor *web* ele utiliza o protocolo HTTP/1.1, estabelecendo uma conexão *Transmission Control Protocol* - TCP (Protocolo de Controle de Transmissão) na porta 80 entre o navegador e o servidor, embora o número da porta possa variar. Após a conexão realizada, é possível enviar uma solicitação, obter uma resposta e, depois, enviar solicitações adicionais, recebendo as respostas adicionais. Essa solicitação ou requisição é composta por vários campos, assim como estão ilustrados na Figura 13 mais adiante.

Para a execução do programa *blast.exe* é necessário enviar o método POST, bem como os campos Host, Content-length, Content-Type e o Referer. O método POST é utilizado na requisição e especifica qual ação deve ser realizada pela URL determinada (ver Figura 13). Nesse caso, o POST especifica o envio dos dados para serem processados. O método Host é o recurso em que a requisição será aplicada e o Content-length é o tamanho da mensagem. O Content-Type especifica a estrutura do corpo da mensagem e, nesse caso, a *multipart/form-data* diz que os parâmetros passados serão separados por um delimitador que é especificado pelo *boundary* e o delimitador dessa requisição é “----WebKitFormBoundarykPsm4kmRUBXAgknf”. E, por fim, o *Referer* é o endereço anterior da URL requisitada

O programa Online Blast XLS Reporter pode ser dividido nas seguintes partes: ExecelHelper.cs, frmMain.cs, BlastParser.cst tal como demonstra o esquema da Figura 14.

Utilizou-se também a classe chamada `HttpRequest` do Microsoft *framework* .NET e, por meio dela é possível especificar todos os campos de uma requisição HTTP. Após todas as especificações serem feitas, a requisição foi enviada para o servidor *web* receber a resposta pelo método `GetResponse`. O *framework* .NET é essencial para executar os programas que foram desenvolvidos com o uso dessa tecnologia, pois oferece segurança, independência de plataforma e executa também no sistema operacional Linux baseado Fedora.

Figura 13 – Requisição HTTP enviada para o servidor do NCBI, objetivando o processamento da sequência.

```
POST http://blast.ncbi.nlm.nih.gov:80/Blast.cgi HTTP/1.1
Host: blast.ncbi.nlm.nih.gov
Proxy-Connection: keep-alive
Content-length: 9639
Cache-Control: max-age=0
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Origin: http://blast.ncbi.nlm.nih.gov
User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/27.0.1453.110 Safari/537.36
Content-Type: multipart/form-data; boundary=----WebKitFormBoundarykPsm4kmRUBXAgknf
Referer: http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&BLAST_PROGRAMS=blastx&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS
Accept-Encoding: gzip, deflate, sdch
Accept-Language: pt-BR,pt;q=0.8,en-US;q=0.6,en;q=0.4
Cookie: MyBlastUser=1FMjvzGmwiwTcblAB574CA824; BlastCubbyImported=passive; ncbi_sid=50C90E8B1B80F981_0000SID

----WebKitFormBoundarykPsm4kmRUBXAgknf
Content-Disposition: form-data; name="QUERY"

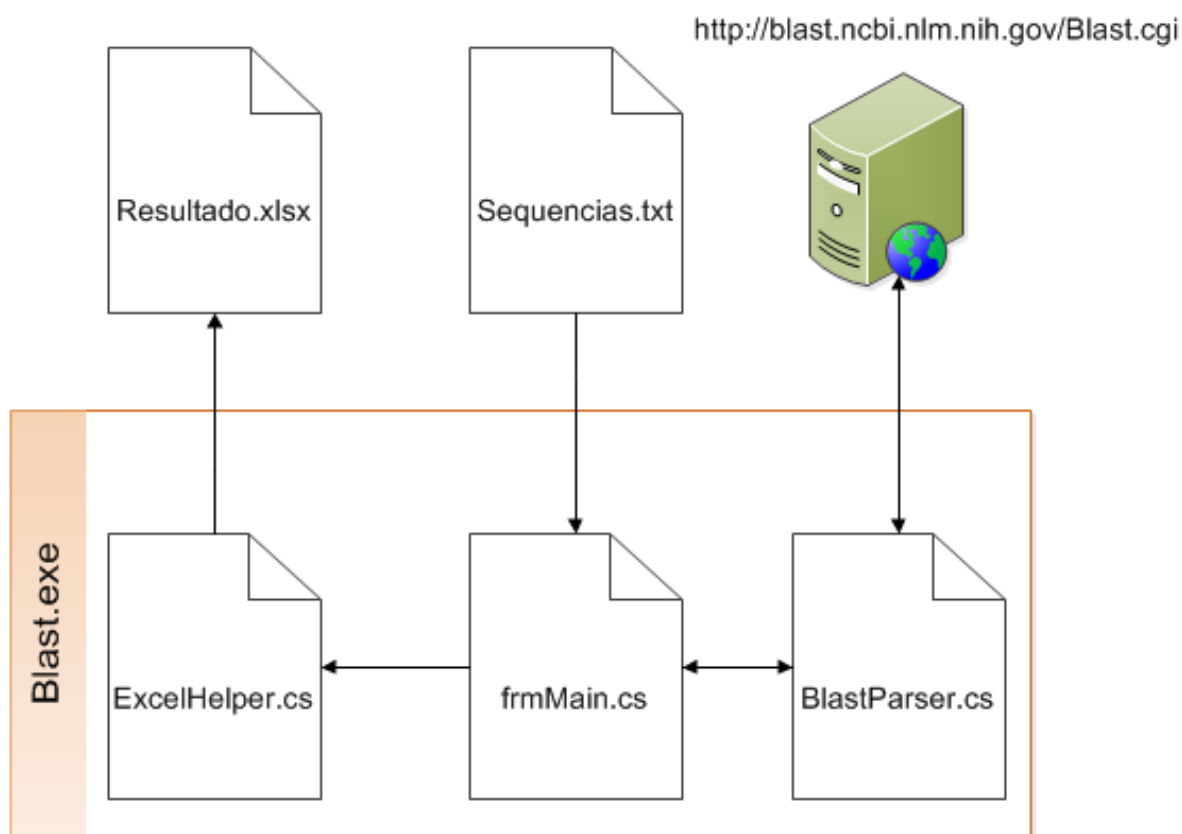
>SEQ491A_A12_B_096.ab1
TACCCTTAAAGGGTCTCGGGNTTATCTTGAAATTTATTTACTGTTCCCTTTAAGATTATA
CATATGATTATAATTTTTGGTTATTGATCTTTAAGGTTCAAGACAGTATCTGTAAGTTT
GCCAAGAAGGGTCTTACACCATCCCAGATTGGTGTCAATTCTTCGTGATTACACGGTATT
GCTCAGGTGAACAGCGTAACAGGGAGCAAGATCCTCAGAATTCTCAAGGCACACGGACTT
GCTCCTGAAATCCCAGAGGATTTATACCACCTCATCAAGAAGGCAGTTGCTATCAGAAAG
CACTTGGAGAGGAACAGGAAGGACAAGGATTCCAAATTCAGGTTGATTCTTGTTGAGAGC
AGGATCCACCGACTTGCCCGCTACTACAAGAAGACAAGAAGCTTCCACCGGTCTGAAAA
TACGAGTCAACCACCACCAGCCTCT
----WebKitFormBoundarykPsm4kmRUBXAgknf
Content-Disposition: form-data; name="db"
```

Fonte: Autoria própria.

A classe `frmMain.cs` (ver Figura 14) é responsável por ler o arquivo tipo texto com extensão `.txt`, contendo todas as sequências que serão analisadas. Uma sequência por vez é enviada para classe `BlastParser`, que utiliza o `HttpRequest` para enviar a requisição para o servidor *web* do NCBI.

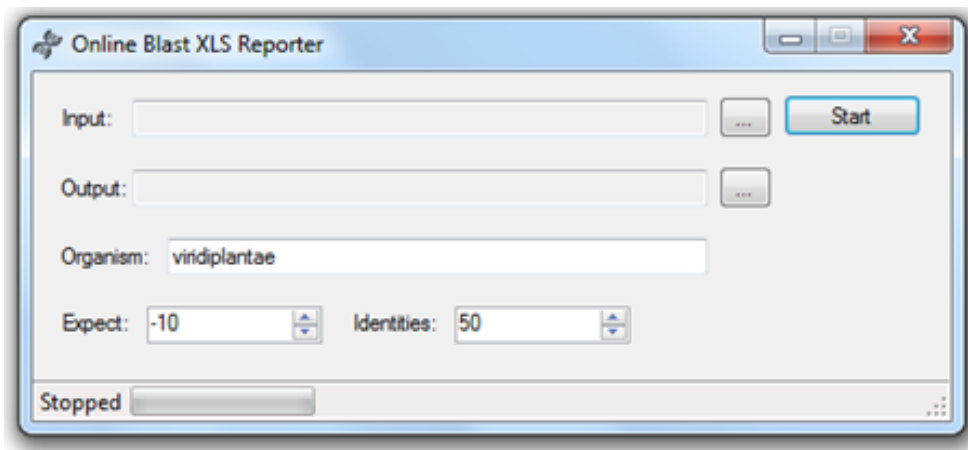
No corpo da mensagem, são passados vários parâmetros para o processamento do servidor do NCBI e entre eles estão a sequência e o tipo de organismo. A resposta obtida é processada ainda pela classe BlastParser e os resultados são transferidos para classe frmMain.cs, que, por sua vez, faz a verificação em cada resultado para ver se atende o critério de filtragem (*expect e identities*). Caso esse critério seja atendido, o resultado é transferido para a classe ExcelHelper.cs, a qual é encarregada de gravar o valor recebido em uma planilha do aplicativo Microsoft Excel. A Figura 15 demonstra a interface do programa Online Blast XLS Reporter.

Figura 14 – Estrutura do programa Online Blast XLS Reporter.



Fonte: Autoria Própria

Figura 15 -Interface do programa Online Blast XLS Reporter.



Fonte: Autoria própria.

Com a execução do BLASTx utilizando-se o programa *Online Blast XLS Reporter* é feita primeiro a tradução, através do código genético, das sequências de nucleotídeos nas possíveis sequências de aminoácidos armazenadas na base de dados. O percentual mínimo de identidade entre as sequências para validar o alinhamento é configurável. O *score* é calculado de acordo com o número de lacunas e divergências entre as bases em um alinhamento. Se o *score* é alto melhor é o alinhamento. O *e-value (expect)* é o número de vezes que um alinhamento pode ocorrer ao acaso em uma busca completa no banco de dados, isto é, parâmetro de confiança do alinhamento. Quanto menor o *e-value*, maior pode ser a similaridade entre as sequências alinhadas e as mesmas serem homólogas (MELO, 2009).

Identities é o valor em porcentagem do número de acertos (*matches* ou letras similares) em relação ao tamanho total do alinhamento, é o grau de similaridade entre as sequências. O *Positives* traz a porcentagem de *matches* de semelhança química (FORMIGHIERI, 2002). *Gaps* são espaços inseridos em um alinhamento para compensar regiões de inserção ou deleção em uma das duas sequências (entrada ou banco) (EMBRAPA, 2007). No quadro 2, um exemplo da informação de *M.urundeuva* que é gerada em planilha do excel após o processamento do programa *Online Blast XLS*. No presente trabalho, foi aplicado o *e-value* E-10 e *Identitie* maior ou igual a 50 para a base de dados de *viridiplantae* (apenas plantas). Vale ressaltar, que todas as planilhas geradas pelo programa *Online Blast XLS Reporter* referente *singlets* e *contigs* da *M.urundeuva* estão inseridas nos anexos do presente trabalho.

Quadro 2 – Dados gerados após a execução do programa Online Blast XLS Reporter.

Contig12						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_003537949.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	204142 bits(518)	1,88E-57	96/131(73%)	111/131(85%)	0/131(0%)
ref XP_002513443.1	lipoxygenase, putative [Ricinus communis]	199519 bits(518)	7,32E-56	93/130(72%)	107/130(82%)	0/130(0%)
gb AAZ57444.1	lipoxygenase LOX1 [Populus deltoides]	199134 bits(518)	1,15E-55	93/130(72%)	108/130(83%)	0/130(0%)
ref XP_002323952.1	predicted protein [Populus trichocarpa]	195667 bits(518)	1,51E-54	91/130(70%)	107/130(82%)	0/130(0%)
gb ABM88259.1	lipoxygenase [Phaseolus vulgaris]	191815 bits(518)	4,26E-53	92/131(70%)	105/131(80%)	0/131(0%)
gb EMJ07637.1	hypothetical protein PRUPE_ppa000968mg [Prunus persica]	196052 bits(518)	2,18E-54	91/131(69%)	106/131(81%)	0/131(0%)
ref XP_004306503.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	194897 bits(518)	4,11E-54	90/131(69%)	108/131(82%)	0/131(0%)
gb ADZ73653.1	putative 13-lipoxygenase [Capsicum annuum]	191045 bits(518)	8,86E-53	89/131(68%)	105/131(80%)	0/131(0%)
dbj BAB84352.1	lipoxygenase [Citrus jambhiri]	189504 bits(518)	2,81E-52	89/130(68%)	104/130(80%)	0/130(0%)
ref XP_002297796.1	predicted protein [Populus trichocarpa]	189119 bits(518)	4,02E-52	89/131(68%)	106/131(81%)	0/131(0%)
emb CAN80268.1	hypothetical protein VITISV_027685 [Vitis vinifera]	193741 bits(518)	1,11E-55	88/131(67%)	109/131(83%)	0/131(0%)
ref XP_002285574.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic [Vitis vinifera]	193356 bits(518)	1,37E-53	88/131(67%)	109/131(83%)	0/131(0%)
gb ACZ17391.1	lipoxygenase [Vitis vinifera]	193356 bits(518)	1,45E-53	88/131(67%)	109/131(83%)	0/131(0%)
emb CAA05278.1	loxC homologue [Solanum pimpinellifolium]	19143 bits(518)	3,85E-53	88/131(67%)	104/131(79%)	0/131(0%)
ref NP_001233812.1	lipoxygenase [Solanum lycopersicum]	189889 bits(518)	2,49E-52	87/130(67%)	104/130(80%)	0/130(0%)
gb AAZ57445.1	lipoxygenase LOX2 [Populus deltoides]	187963 bits(518)	1,05E-51	88/131(67%)	105/131(80%)	0/131(0%)
gb EMJ21452.1	hypothetical protein PRUPE_ppa001634mg [Prunus persica]	182956 bits(518)	3,73E-50	88/131(67%)	103/131(79%)	0/131(0%)
gb ACJ54281.1	lipoxygenase [Camellia sinensis]	179104 bits(518)	1,84E-48	88/131(67%)	103/131(79%)	0/131(0%)
gb AAT77551.1	LoxC-like [Solanum pimpinellifolium]	184882 bits(518)	2,24E-54	86/131(66%)	105/131(80%)	0/131(0%)
emb CAA05270.1	unnamed protein product [Solanum habrochaites]	184496 bits(518)	4,45E-54	86/131(66%)	105/131(80%)	0/131(0%)

Fonte: Autoria própria.

Na Figura 16, é possível observar um desenho esquemático do *pipeline* criado para o presente projeto. Ele representa a ordem sequencial em que os programas foram processados em cada fase da montagem das sequências.

Figura 16 – Esquema do *pipeline*.

Fonte: Autoria própria.

O trabalho de montagem e análise das sequências, que anteriormente era realizado manualmente por dias, pelos biólogos moleculares do departamento de biotecnologia vegetal da UNAERP, com a aplicação do *pipeline*, foi reduzido para poucas horas. Contudo, os resultados gerados por esse *pipeline* precisavam ser armazenados em um banco de dados para consultas futuras. A partir daí, foram necessárias a modelagem e a criação de um banco de dados. No entanto, com relação ao uso do *pipeline*, bem como ao manuseio do banco de dados, é preciso ter conhecimentos específicos em informática para que tais processos sejam realizados. Nesse caso, criou-se uma

interface para viabilizar independência e usabilidade aos pesquisadores. E pensando na referida necessidade, foi criado o sistema denominado WISE.

A infraestrutura (com relação ao *hardware*) utilizada no desenvolvimento dos programas e *scripts* do presente projeto foram: um computador Dell Core I3 com 6 GB de memória RAM e HD de 500 GB 64 *bits*; o *hardware* em nível de servidor para implementação do sistema gerenciador de banco de dados MySQL; um computador Dell Core I5 com 8 GB de memória RAM e HD de 1 TB 64 *bits*, utilizado na instalação e configuração do Apache Tomcat para a publicação do sistema WISE e, também, para publicar a aplicação *mobile* WISE.

2.1 MODELAGEM E IMPLEMENTAÇÃO DO SISTEMA WISE

Na modelagem do banco de dados do sistema WISE, utilizou-se MySQL Workbench, que é uma ferramenta com interface visual para *design*, desenvolvimento e administração de base de dados MySQL. Possui uma interface gráfica, que possibilita maior agilidade no processo de desenvolvimento, além de possuir recursos administrativos que permitem a automatização de tarefas frequentes com extensões e *scripts*; ambos podem ser obtidos no *site* do MySQL (<http://www.mysql.com/downloads>) (LUCKOW; MELO, 2010). A referida ferramenta foi usada para especificar os dados envolvidos, as suas relações e os requisitos de análise. Para a implementação do banco de dados do sistema WISE, utilizou-se o sistema de gerenciamento de banco de dados MySQL. Na Figura 17, colocada mais adiante, é apresentado o Diagrama Entidade Relacionamento do banco de dados do sistema WISE. Tal banco de dados foi criado para armazenar informações referentes aos usuários do sistema, informações relacionadas aos projetos de cada usuário e armazenar as saídas de cada fase do *pipeline*.

No centro da Figura 17, há a representação da entidade denominada projeto, com o intuito de listar os projetos existentes já cadastrados pelos usuários que acessam o sistema WISE. Ela se relaciona com todas as demais entidades do sistema com exceção das entidades usuário tipo e organismo tipo. A entidade usuário tem o propósito de armazenar informações referentes ao usuário do sistema e possui relacionamento com as entidades usuário_tipo e entidade projeto. Já a entidade usuário_tipo é a responsável por armazenar os

tipos de permissões que o usuário terá no sistema e se relaciona apenas com a entidade usuario. É importante destacar que essa entidade possui uma chave primária composta - tática utilizada para definir as permissões que o usuário possui no sistema. Por padrão, quando o usuário for cadastrado no sistema, ele receberá a permissão de usuário comum (role_usuario). Se o administrador atribuir-lhe a opção de administrador, ele irá possuir duas permissões: role_usuario e role_administrador.

Observa-se, ainda, na Figura 17, a entidade organismo, que obtém informações dos organismos aos quais o projeto está relacionado. E essa entidade está relacionada com a entidade projeto. Na mesma figura, é possível observar a entidade organismo_tipo, contendo informações referentes aos tipos de organismos. Essa entidade está relacionada apenas com a entidade organismo.

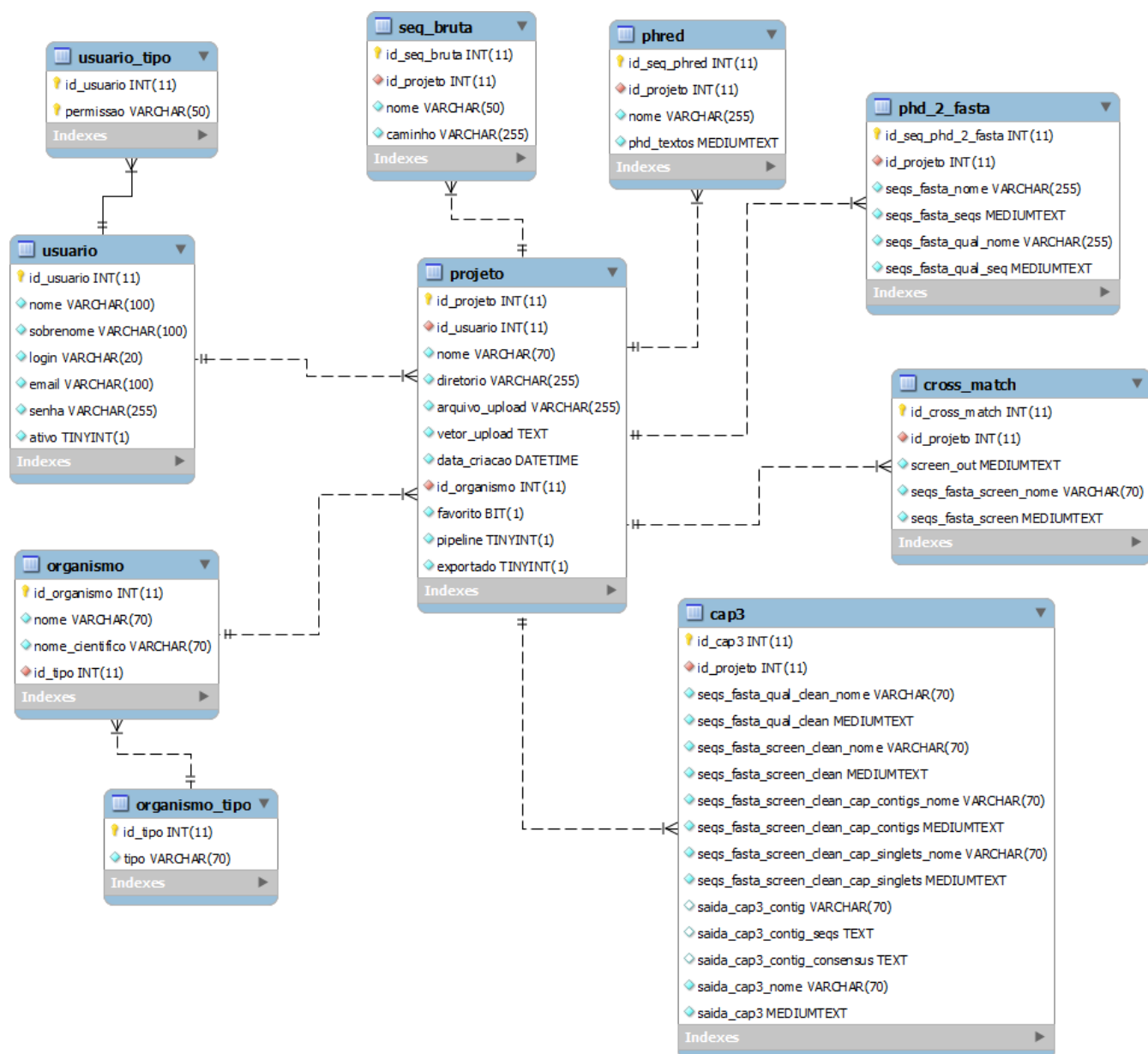
Com relação à entidade denominada seq_bruta, ela é utilizada para armazenar informações relacionadas aos arquivos cromatogramas fornecidos pelo usuário. É utilizada para armazenar informações relacionadas aos arquivos cromatogramas fornecidos pelo usuário e possui relacionamento com a entidade projeto.

No que se refere à entidade denominada phred, são armazenados os arquivos com extensão .phd e é preciso ressaltar que os arquivos armazenados aqui, são oriundos da saída do programa PHRED. Além disso, possuem relacionamento com a entidade projeto. A entidade phd2fasta é semelhante à entidade phred, porém, os dados armazenados naquela são os arquivos FASTA obtidos como saída do *script* phd2fasta. Possui relacionamento com a entidade projeto.

Na entidade cross_match são armazenados os arquivos referentes à saída do programa cross_match.

Por fim, a entidade cap3 armazena os arquivos de saída do programa CAP3, assim como os arquivos .contigs, .singlets, .screen.clean, .qual.clean e um arquivo .txt. No Quadro 3, é possível observar detalhes referentes às entidades, aos atributos e à descrição deles.

Figura 17 - Diagrama Entidade Relacionamento.



Fonte: Autoria própria.

Quadro 3 - Descrições de todas as entidades e atributos criados no banco de dados do sistema WISE.

Entidade	Atributos	Descrição do atributo
projeto	id_projeto	chave primária da entidade projeto
projeto	id_usuario	chave de relacionamento com a entidade usuário
projeto	nome	nome do projeto
projeto	diretorio	diretório que a pasta do projeto é criada no sistema
projeto	arquivos_upload	diretório que os arquivos cromatogramas são armazenados no sistema
projeto	vetor_upload	diretório que o arquivo vetor é armazenado no sistema
projeto	data_criacao	data de criação do projeto
projeto	id_organismo	chave de relacionamento com a entidade organismo
projeto	favorito	indica se o projeto está selecionado no sistema
projeto	pipeline	indica se o projeto já passou pelo processo de <i>pipeline</i>
projeto	exportado	indica se o projeto já foi exportado para o banco de dados.
usuario	id_usuario	chave primária da entidade usuario
usuario	nome	nome do usuário
usuario	sobrenome	sobrenome do usuário
usuario	login	identificação de acesso do usuário no sistema
usuario	email	<i>email</i> do usuário
usuario	senha	senha para usuário acessar o sistema
usuario	ativo	permite visualizar se o usuário está ou não ativo no sistema
usuario_tipo	id_usuario	atributo referente a entidade usuario e chave primária
usuario_tipo	permissao	possui informações referentes as permissões do usuário no sistema e também é a chave primária da entidade

organismo	id_organismo	chave primária da entidade organismo
organismo	nome	nome do organismo
organismo	nome_cientifico	nome científico do organismo
organismo	id_tipo	chave de relacionamento com a entidade organismo_tipo
organismo_tipo	id_tipo	chave primária da entidade organismo
organismo_tipo	tipo	tipo de organismo
seq_bruta	id_seq_bruta	chave primária da entidade organismo
seq_bruta	id_projeto	chave de relacionamento com a entidade projeto
seq_bruta	nome	nome da sequência cromatograma.
seq_bruta	caminho	caminho onde a sequência foi armazenada
phred	id_seq_phred	chave primária da entidade phred.
phred	id_projeto	chave de relacionamento com a entidade projeto
phred	nome	nome do arquivo phd
phred	phd_textos	conteúdo do arquivo com extensão .phd
phd2fasta	id_seq_phd_2_fasta	chave primária da entidade phd_2_fasta.
phd2fasta	id_projeto	chave de relacionamento com a entidade projeto
phd2fasta	seqs_fasta_nome	nome do arquivo seq_fasta
phd2fasta	seqs_fasta_seqs	conteúdo do arquivo seqs_fasta
phd2fasta	seqs_fasta_qual_nome	nome do arquivo de saída do programa com extensão .qual.
phd2fasta	seqs_fasta_qual_seq	conteúdo do arquivo seqs_fasta.qual
cross_match	id_cross_match	chave primária da entidade cross_match
cross_match	id_projeto	chave de relacionamento com a entidade projeto
cross_match	screen_out	conteúdo do arquivo screen_out
cross_match	seqs_fasta_screen_nome	nome do arquivo de saída do programa cross_match
cross_match	seqs_fasta_screen	conteúdo do arquivo seqs_fasta.screen

cap3	id_cap3	chave primária da entidade cap3
cap3	id_projeto	chave de relacionamento com a entidade projeto
cap3	seqs_fasta_qual_clean_nome	nome do arquivo com extensão .qual.clean
cap3	seqs_fasta_qual_clean	conteúdo do arquivo com extensão .qual.clean.
cap3	seqs_fasta_screen_clean_nome	nome do arquivo com extensão .screen.clean
cap3	seqs_fasta_screen_clean	conteúdo do arquivo com extensão .screen.clean
cap3	seqs_fasta_screen_clean_cap_contigs_nome	nome do arquivo <i>contig</i>
cap3	seqs_fasta_screen_clean_cap_contigs	conteúdo do arquivo <i>contigs</i>
cap3	seqs_fasta_screen_clean_cap_singlets_nome	nome do arquivo <i>singlet</i> .
cap3	seqs_fasta_screen_clean_cap_singlets	conteúdo do arquivo <i>singlet</i>
cap3	saída_cap3_nome	nome do arquivo de saída com extensão .txt
cap3	saída_cap3	conteúdo do arquivo de saída com extensão .txt

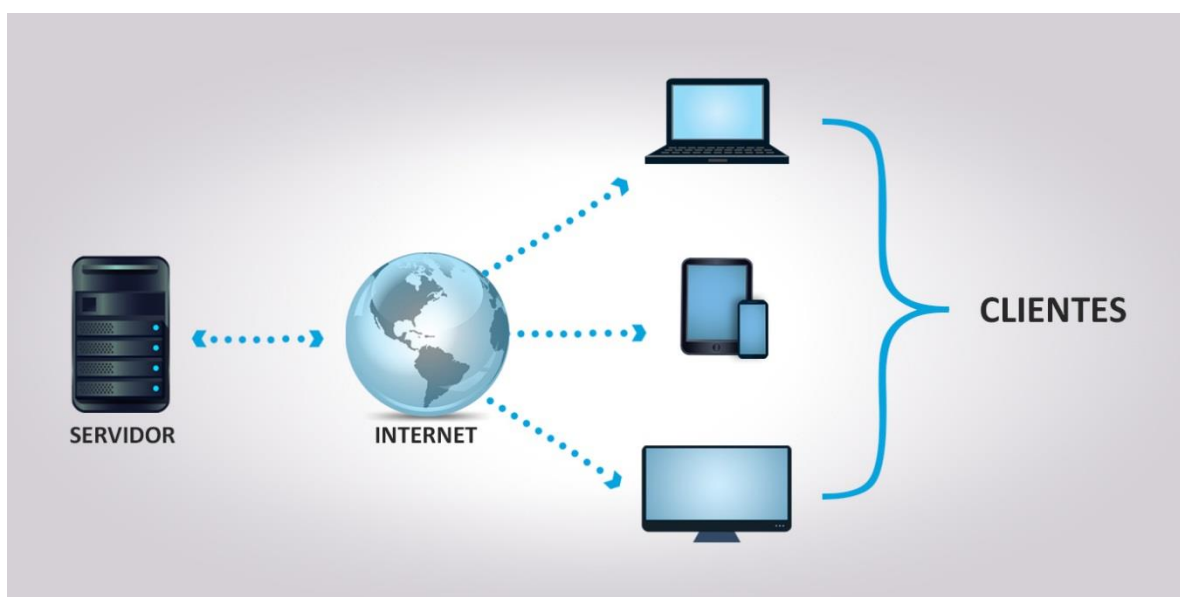
Fonte: Autoria própria.

Na implementação (ou programação) do código do sistema WISE, foi utilizada a linguagem de programação JAVA em razão da independência de plataforma, por ser seguro e confiável. Os *frameworks* utilizados são o JSF, o Hibernate, o Spring Security e o Primefaces. O JSF é um *framework* Java que viabiliza da parte de controle, dentro do padrão MVC (*Model View Controller*) (ANDRADE, 2010); seu uso é vantajoso, pois é baseado em componentes e também por sua facilidade em desenvolver interfaces gráficas. Já o Hibernate é um *framework* de mapeamento objeto relacional, pertencente à camada de modelo do padrão MVC, que foi usado para padronizar os comandos SQLs repetitivos referentes à manipulação dos dados no banco. O Spring Security foi escolhido para implementar a parte de segurança no sistema WISE e o Primefaces é uma extensão da linguagem JSF, que contém alguns componentes extras para o JSF (KING, 2013).

O sistema WISE está baseado na arquitetura cliente-servidor, em que o servidor *web* é responsável por todo o processamento e armazenamento dos dados, compartilhando esses recursos com os clientes que são os usuários, os

quais utilizam seus navegadores de Internet (Figura18). O servidor *web*, usado para publicar o sistema WISE, é o Apache Tomcat que tem as funções de receber requisições *Hypertext Transfer Protocol* (HTTP), executar os componentes Java e devolver para o *browser* o código HTML (*Hypertext Markup Language* ou em português Linguagem de Marcação de Hipertexto), imagens, estilos, etc.

Figura 18 - Arquitetura cliente/servidor – Sistema WISE



Fonte: Autoria própria

Com relação à criação dos ícones no *design* do sistema WISE, foi necessário um estudo no que diz respeito ao redimensionamento da imagem, já que para cada dispositivo eletrônico é necessária uma proporção diferente da figura, sendo mapeada para *smartphones*, celulares e *tablets* com diferentes resoluções. Foi criada uma identidade visual para os ícones de modo a unificar e equilibrar cores e formas com a tipografia adotada e *background* (plano de fundo), buscando referências com *design* atual. O Flat Design foi usado para a tradução do recurso em forma de ilustração para os ícones, tornando-os mais intuitivos para o usuário.

O conceito principal do Flat Design é a representação da ilustração com clareza e simplicidade, sendo essa uma nova tendência geral de *design*. O Flat Design traz o desapego ao estilo e enfatiza a funcionalidade, excluindo de seus padrões bordas, sombras, gradientes e reflexos para buscar uma comunicação

plena e clara. Utilizou-se, também, o banco de vetores gratuitos Freepik com o intuito de fazer a composição das imagens figurativas.

No que cerne à criação do logotipo (Figura 19) do sistema WISE, contou-se com o *software* Adobe Photoshop CS6 para criar e editar imagens bidimensionais a partir do *bitmap*, isso quer dizer, baseado em ilustrações criadas a partir de um mapeamento de *bits*, que suporta camadas, máscaras, sistemas de cores como RGB e CYMK, tal *software* dá suporte a múltiplos formatos de imagem. A definição das cores do logotipo está relacionada à representação das áreas cerebrais que são responsáveis por (DAMÁSIO E DAMÁSIO, 2004):

- Azul: Área de Associação Somatossensorial / Área Visual, avaliação de peso, textura temperatura, reconhecimento de objetos;
- Verde: Área de Associação / Área Emocional, memória de curto prazo, equilíbrio e emoção.
- Rosa: Área de Wernicke / Funções Mentais Superiores, conhecimento, interpretação e associação das informações, mais especificamente da linguagem; concentração, planejamento, julgamento, expressões emocionais, criatividade e inibição.
- LARANJADO: Área de Associação Sensorial. Processamento complexo da informação visual.

Figura 19 – Logotipo do sistema WISE.



Fonte: Autoria própria.

Para a referência em cores, foi utilizado o *software* Adobe Kuler - um recurso disponível gratuitamente pelo fabricante. Ele auxilia na criação de paletas de cores baseado na teoria das cores.

Foi desenvolvido, também, um programa em linguagem de programação PERL com o objetivo de realizar o *backup* (cópia de segurança) do diretório corrente do projeto *M. urundeuva*. O *backup* foi criado em um computador servidor de arquivos configurado no sistema operacional Linux baseado a distribuição no Fedora 17. Para isso, foi utilizado o recurso compartilhamento remoto do referido sistema operacional, chamado Network File System (NFS). O NFS permite que um sistema compartilhe seus diretórios e arquivos com outros sistemas por meio da rede de computadores. Usando o NFS, os usuários e programas acessarão os arquivos em sistemas remotos como se fossem arquivos locais (SMITH, 2006).

O programa grava os arquivos no servidor de arquivos remoto no formato de AAAA-MM-DD-HH-MM-SS-MS (Ano-Mês-Dia-Hora-Minuto-Segundo-Milisegundo), com o objetivo de facilitar a consulta e recuperar os dados do *backup*, caso seja necessário.

Além disso, foi desenvolvido um programa para extrair os dados dos arquivos correntes do projeto *M. urundeuva* e armazena-los no banco de dados MySQL, que funciona em alta disponibilidade, em dois servidores com replicação dos dados. Tanto o conteúdo do *backup* quanto o conteúdo do banco de dados estão disponíveis para consulta por meio do acesso ao *link* do sistema WISE, que solicita a autenticação do usuário. O mecanismo de autenticação consiste na verificação da identidade dos usuários que realizam acesso ao sistema WISE, objetivando proteger os dados e recursos contra modificações impróprias, acesso não autorizado e, ao mesmo tempo, garantir a utilização aos usuários cadastrados com permissão no sistema WISE.

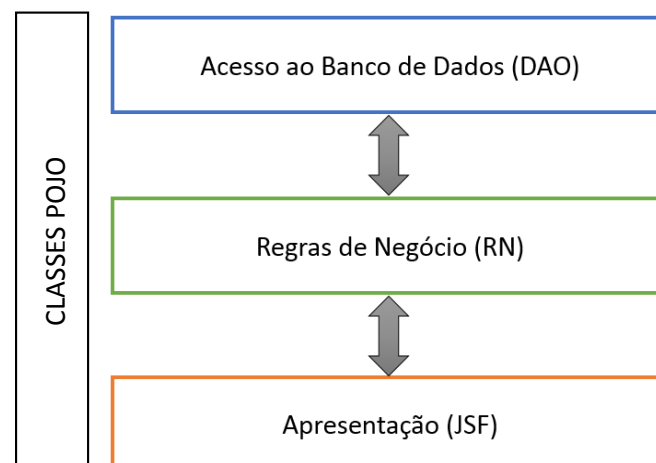
2.2. IMPLEMENTAÇÃO DO SISTEMA WISE.

Tal como descrito anteriormente, para gerenciar as sequências geradas no *pipeline* criado para o presente projeto e automatizar o processo de análise, foi implementado um sistema *web* denominado WISE. Para o desenvolvimento desse sistema, utilizou-se a arquitetura *Model-View-Controller* (MVC) em português Modelo-Visão-Controle. Essa arquitetura costuma ser

muito usada no desenvolvimento de aplicações e sistemas *web*. Ela consiste em separar o sistema em três incumbências: modelo, visualização e controle. A incumbência modelo faz a representação das regras de negócio, lógica e funções; mantém o estado da aplicação e fornece ao controlador o acesso aos dados. A visualização é responsável pela interface do usuário e define a forma como os dados são apresentados, assim como encaminha as ações do usuário para o controlador. O controlador é responsável por ligar o modelo e a visualização; interpreta as entradas do *mouse* e teclado do usuário, informando o modelo e retornando a visualização adequada à solicitação (LOSSIUS *et al*, 2014; MICROSOFT, 2014).

Vale destacar que no modelo MVC, a camada de modelo não acessa a camada de visualização e vice-versa. A única camada que consegue conectar essas duas é a camada de controle, ou seja, a camada de controle busca os dados na camada de modelo e os disponibiliza para serem demonstrados na camada de visualização, tal como se pode observar na Figura 20. Primeiro, foi criada uma classe simples para mapear uma tabela do banco de dados, depois, criou-se uma classe DAO para realização do acesso à base de dados; após essa etapa, foi criada uma classe responsável pelas regras de negócio - a única que consegue acessar a classe de acesso a dados. E, por último, gerou-se a classe Bean, que se comunica apenas com a classe de regras de negócio (ANDRADE, 2010; LUCKOW; MELO, 2010).

Figura 20 - Representação de arquitetura MVC em camadas.



Fonte: LUCKOW; MELO, 2010, Adaptado pela autora).

Seguindo a metodologia descrita anteriormente, no presente projeto foram utilizados alguns *frameworks* da linguagem Java: Java Server Faces (JSF) + Primefaces + Facelets, Hibernate e para a parte de segurança do sistema, foi utilizado o *framework* Spring Security.

O JSF é uma especificação para um *framework* de componentes na implementação de *software web* em Java. Ele utiliza um modelo de interfaces gráficas baseado em evento e simplifica a construção de interfaces de usuário para aplicações JavaServer (JAVASERVERFACES, 2014). Por exemplo, quando o usuário selecionar um ícone ou um botão usando o *mouse* ou o teclado, causará ações particulares. O botão em questão será considerando a fonte desse evento e o evento gerado a partir do clique do *mouse* será associado com o objeto JButton (que representa o botão na tela) no programa. Sempre um evento possuirá um objeto fonte (neste caso o objeto JButton) (INFO, 2014).

A tecnologia JSF foi definida pelo Java Community Process (JCP) e isso significa dizer que grandes empresas como Oracle, IBM e Sun participaram da sua definição e aprovaram sua especificação. Ao utilizar os componentes dessa tecnologia, o desenvolvedor não precisa se preocupar com *Javascript* e HTML, pois todos os componentes serão renderizados e exibidos em formato HTML. Pensando no modelo MVC, o JSF possui uma camada de visualização baseada em *Facelets* e um conjunto de classes, chamadas de Managed Beans, que fazem parte da camada de controle (ANDRADE, 2010; LUCKOW; MELO, 2010; CAELUM, 2013).

O Facelets é um *framework* que auxilia na criação de *templates*, padronizando-os e centralizando as definições visuais e estruturais das telas do sistema. A vantagem da utilização desse *framework* é a diminuição da quantidade de códigos necessária para definição dos estilos das páginas. Com a utilização de Facelets, é possível criar uma estrutura padrão de visualização e, depois, associar essa estrutura às páginas *web*. Essas “páginas modelos” foram salvas dentro da pasta *template*, que, por sua vez, encontra-se dentro da pasta WebContent. Os estilos do *site* foram feitos com o uso do *Cascading Style Sheets* (CSS) e, em uma *uma página web*, o CSS viabiliza a parte visual e o

HTML viabiliza marcações de texto (HOOKOM, 2005). A Figura 21 mostra o *template* padrão adotado na implementação do sistema WISE.

Entre as mais conhecidas implementações do JSF estão a Sun Mojarra e a MyFaces da Apache. Nessa aplicação foi utilizada a implementação Mojarra, por ser mais simples e por oferecer alguns componentes básicos para o modelo de desenvolvimento. Para instalar o JSF, basta adicionar os jars obtidos no *site* oficial do Mojarra (<http://javaserverfaces.java.net/>) no projeto *web* desejado e configurar as *servlets* no arquivo *web.xml* (LUCKOW; MELO, 2010). Arquivos jars permitem agrupar vários arquivos em um arquivo. Normalmente, esses arquivos contém arquivos de classe e recursos auxiliares associados a *applets* e aplicações (ORACLE, 2013). No desenvolvimento do sistema WISE, além do JSF, foi utilizada uma biblioteca de componentes chamada PrimeFaces. Existem várias bibliotecas para o JSF, tais como:

- RichFaces, da JBoss (<http://www.jboss.org/richfaces/>)
- ICEFaces, da ICESoft (<http://www.icefaces.org/>)
- Tomahawk, da Apache MyFaces (<http://myfaces.apache.org/tomahawk>)
- Trinidad, da Apache MyFaces (<http://myfaces.apache.org/trinidad>)
- Tobago, da Apache MyFaces (<http://myfaces.apache.org/tobago>)

Figura 21 - *Template* padrão das páginas do sistema WISE.

Fonte: Autoria própria.

Esses componentes são recursos adicionais aos padrões do JSF. No sistema WISE, foram utilizados vários componentes dos PrimeFaces, como: fileUpload, DataTable, Button, PanelGrid, Panel, Dialog, MenuBar, Messages, fileDownload.

Para utilizar o PrimeFaces aqui, foi preciso obter os .jars no *site* do PrimeFaces (<http://www.primefaces.org/downloads.html>). E para auxiliar na utilização dos componentes, usou-se como referência a documentação dos componentes também disponíveis no *site* do PrimeFaces.

O *framework* Hibernate é o responsável pelo acesso aos dados no sistema WISE e é um *framework* de mapeamento objeto relacional, pertencente à camada de modelo do padrão MVC. Ele é utilizado para padronizar comandos SQLs repetitivos, referentes à manipulação de dados. O mapeamento objeto relacional é uma técnica que converte os dados de um banco de dados relacional em linguagens orientadas a objeto. Ou seja, uma tabela no modelo relacional é representada por uma classe no modelo orientado a objeto. Uma tupla do modelo relacional é representada por um objeto no modelo orientado a objeto e uma coluna no modelo relacional é representado por um atributo no modelo orientado a objeto (ANDRADE, 2010).

Para utilizar o *framework* Hibernate no sistema WISE, foi necessário adquirir os jars no *site* do Hibernate (<http://www.hibernate.org>). Além dessa biblioteca, adquiriu-se outra chamada slf4j. Essa biblioteca é pré-requisito para o *framework* funcionar. Após as bibliotecas serem adicionadas ao sistema, deve-se configurar um arquivo XML para que seja possível conectar o Hibernate ao MySQL. Depois da criação desse arquivo XML, criou-se uma classe para realizar o carregamento desse arquivo XML, fazendo a ponte entre tal arquivo e a conexão com o banco de dados. Após configurado, é necessário realizar o mapeamento das classes, que indica qual atributo representa cada coluna da tabela, qual classe representa a tabela que será mapeada, qual o tipo de relacionamento entre as tabelas e quem será a chave primária, entre outros. Existem duas formas de realizar esse mapeamento, a primeira é por arquivos XML e a segunda, e mais recente, é por meio de anotações. No sistema WISE foi utilizado o mapeamento por anotações. Alguns exemplos de anotações são:

- @Table – possui vários parâmetros e um deles indica qual o nome da tabela que a classe representará;
- @Column – possui vários parâmetros e um deles indica qual o nome da coluna no banco de dados que o atributo representará;
- @Id – utilizada para informar quem o atributo representa na chave primária;
- @GeneratedValue – utilizada para indicar a forma que o banco de dados utilizará para autonumerar a chave primária.

Vale destacar, também, que toda classe mapeada pelo Hibernate deve implementar a interface *Serializable* e os métodos equals e hashCode devem ser reimplementados. Foi utilizada a IDE Eclipse no desenvolvimento do código do sistema WISE, que auxiliou na sobrescrita desses métodos (LUCKOW; MELO, 2010).

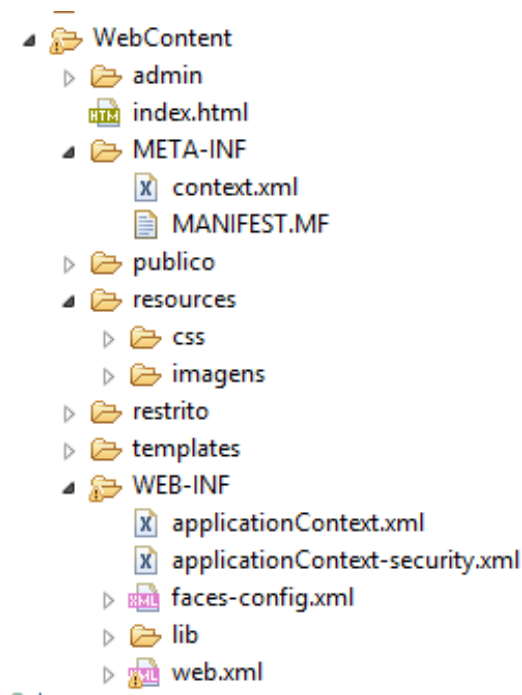
A segurança do sistema foi implementada com a ajuda de um *framework* chamado Spring Security - um dos projetos do Spring *framework* e garante a autenticação e a autorização dos usuários no sistema (ALEX, TAYLOR, WINCH, 2014). Com ele, foi possível criar regras de restrições para usuários de nível básico e usuários de níveis administrativos. Assim, pode-se controlar o que cada usuário vê e acessa no sistema. Para conseguir esse efeito de permissões de usuário, foi necessário separar em pastas as páginas que cada

tipo de usuário iria acessar. Os usuários com privilégio de administrador têm acesso a todas as páginas do sistema; já os usuários com privilégios básicos têm acesso ao conteúdo das páginas *web* que estão nas pastas público e restrito, e os usuários que não estão cadastrados no sistema têm acesso apenas às páginas que estão dentro da pasta público. A Figura 22, colocada mais adiante, mostra a estrutura das pastas do sistema WISE.

No topo da hierarquia das pastas do sistema está a pasta *WebContent* e dentro dela estão todas as páginas do sistema. A pasta *admin* contém as páginas que os usuários com privilégios de administrador podem acessar. Dentro da pasta META-INF são armazenados os arquivos de configuração do Spring Security. Já na pasta público contém as páginas que os usuários que não são cadastrados no sistema podem acessar. A pasta *resource* armazena os arquivos CSS e as imagens utilizadas no *site* e é geralmente utilizada, pois o componente *graphicImage* do JSF identifica essa pasta, não sendo necessário passar por todo o caminho para identificar as imagens e os arquivos CSS. Na pasta restrito, há páginas que os usuários com privilégios mais restritos podem acessar. A pasta *templates* armazena os templates padrões que as páginas utilizam. Dentro da pasta WEB-INF há arquivos XML referentes à configuração do *Spring Security*, JSF e do servidor Tomcat. A pasta *lib* contém todas as bibliotecas (*jars*) utilizadas no desenvolvimento do sistema.

A parte de autenticação também é gerenciada pelo *Spring Security*. Quando o usuário tenta acessar alguma página do sistema sem estar logado, o *framework* viviliza para que isso não aconteça.

Figura 22 – Estrutura de pastas do sistema WISE.



Fonte: *Software Eclipse* configurado para o sistema WISE.

Para instalar o *Spring Security*, é preciso fazer *download* dos jar no *site* <http://www.static.springframework.org/spring-security/site/downloads.html>. Além disso, o *web.xml* deve ser configurado e, ainda, são criados dois arquivos de configuração, o *application Context.xml* e o *application Context-security.xml*. O primeiro possui informações referentes ao *DataSource*, criado no sistema, e o segundo possui informações referentes à configuração do *framework*, como por exemplo, a consulta que ele deve fazer ao banco de dados para obter as informações de *login*, configuração das permissões de cada pasta, configuração da página de *login*, configuração da página padrão, quando o usuário entrar no sistema e várias outras opções puderem ser configuradas (LUCKOW; MELO, 2010).

Tal *DataSource* é utilizado para que não seja preciso executar dois acessos ao banco de dados ao mesmo tempo (um utilizado pelo *Hibernate* e outro utilizado pelo *Spring*). Esse *framework* também oferece dois tipos de criptografias de senha, o MD5 (*Message-Digest algorithm 5*) e o SHA (*Secure Hash Algorithm*). No desenvolvimento do sistema WISE, foi utilizada criptografia MD5 por ela comparar a senha salva no banco com a senha digitada pelo usuário e validá-las.

Para auxiliar a implementação com relação ao código do sistema WISE, focou-se a *Integrated development enviroment* (IDE) Eclipse. Essa IDE foi desenvolvida pela IBM e possui código livre. O servidor *web* utilizado no desenvolvimento foi o Apache Tomcat, o qual é um *container* Java e um servidor *web*. Um Servidor *web* tem como objetivo receber requisições *Hypertext Transfer Protocol* (HTTP), executar os componentes Java e devolver para o *browser* código HTML, imagens e estilos (ANDRADE, 2010).

Para instalar o Apache Tomcat, foi feito o *download* do pacote de instalação no formato zip, pelo *site* <http://tomcat.apache.org/>. Depois disso, o arquivo foi descompactado e, em seguida, criou-se uma variável de ambiente com o nome de CATALINA_HOME, atribuindo-se a essa variável o caminho de onde o Tomcat foi extraído (instalado). Para realizar o teste de funcionamento do servidor, basta entrar na pasta da qual o Tomcat foi extraído, isto é, entrar na pasta *bin* e executa o arquivo *startup.bat*. Depois disso, é só acessar o *browser* de Internet e digitar <http://localhost:8080/SABIO/publico/login.jsf>. Se tudo ocorrer bem, uma página será mostrada. Para desligar o servidor, executa-se o arquivo *shutdown.bat* no mesmo local em que foi executado o arquivo de inicialização.

Para realizar os testes do sistema WISE, o Tomcat foi integrado ao Eclipse para que ele fosse executado a partir da IDE, permitindo que uma classe ou uma página *web* possa ser executada de dentro do Eclipse.

A seguir, serão mostradas e explicadas as funcionalidades e os componentes utilizados no desenvolvimento do sistema WISE.

A Figura 23 ilustra a tela de *login* do sistema WISE. Se for o primeiro acesso do usuário, ele deverá clicar e posicionar a seta do *mouse* no botão “Registre-se” e efetuar o cadastro. Caso ele complete o cadastro, será redirecionado para a tela de *login* para entrar no sistema, podendo observar o uso de um componente Primefaces na barra de menu. O botão de *logout* também foi feito utilizando o mesmo *framework*. Além do Primefaces fornecer os componentes adicionais para o JSF, ele também fornece estilos para seus componentes.

A Figura 24 mostra a tela de cadastro do usuário e assim que ele entrar no sistema, estará na página principal. Se ele for um usuário novo, aparecerá um *link* indicando que ele deve cadastrar um novo projeto. Caso ele

já possua projetos cadastrados, esse *link* não será exibido. A Figura 25 ilustra a tela principal do sistema.

Figura 23 – Tela de *login* do sistema WISE.

Registre-se

Login sandra

Senha *****

Entrar automaticamente

Entrar

© 2015 - WISE - Sistema Automático de Processos Biotecnológicos
Usuario:

BIOTECNOLOGIA
UNAERP

Fonte: Autoria própria.

Figura 24 - Tela de cadastro de usuários.

CADASTRO DE USUÁRIOS

Cadastro de Usuarios

Nome: *

Sobrenome: *

e-Mail: *

Login: *

Senha: *

Confirmar Senha: *

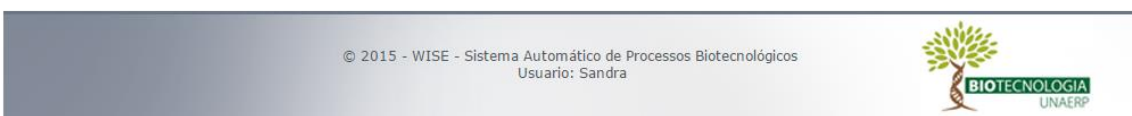


Fonte: Autoria própria.

Figura 25 - Tela principal do sistema.

[Home](#)
[Administração](#)
[Projetos](#)
[Visualizar](#)
[Executar](#)
[Mapa](#)
[Anotação](#)

WISE - Sistema Automático de Processos Biotecnológicos
 Usuário logado: sandra
 Nome do Usuário Logado: Sandra



Fonte: Autoria própria.

A Figura 26 mostra a tela de administração de usuários. O círculo verde representa que o usuário está ativo no sistema. O símbolo de uma coroa amarela indica que o usuário tem a permissão de administrador e, quando a coroa estiver cinza, significará que é um usuário sem privilégios.

A guia projetos contém as informações referentes aos projetos cadastrados no sistema WISE. É nesse botão que o usuário visualizará seus projetos cadastrados ou poderá cadastrar um novo projeto, além de fazer o *upload* dos arquivos cromatogramas e do vetor, bem como executar o *pipeline*. Quando um usuário cadastrar um projeto no sistema, automaticamente será criada uma pasta com seu nome de *login* no servidor. Dentro dessa pasta será criada outra pasta com o nome do projeto mais a data de criação do projeto. Todos os arquivos para o iniciar o *pipeline* e as saídas de cada programa do *pipeline* estarão armazenados ali. Será criada, ainda, outra pasta dentro dessa com o nome do projeto mais a data de criação do projeto. A Figura 27 mostra como ficou a respectiva tela. O sistema trabalha sempre com relação ao projeto selecionado.

Figura 26 - Tela de edição e administração de usuários do sistema.

ÁREA ADMINISTRATIVA



Home Administração Projetos Visualizar Executar Mapa Anotação Logout

A listagem abaixo exibe a relação de todos os usuários do sistema. Você poderá realizar ativação e desativação, edição e exclusão para os usuários.

Ativo	Codigo	Nome	Sobrenome	E-Mail	Login	Permissão	Editar	Excluir
	1	Sandra	Fantinato	srfantinato@gmail.com	sandra			
	4	caio	bugorin	caiobugorin@gmail.com	mbugorin			
	5	Ana	Teste	sandra@odin.unaerp.br	anaana			

© 2015 - WISE - Sistema Automático de Processos Biotecnológicos
Usuário: Sandra



Fonte: Autoria própria.

Figura 27 - Tela de meus projetos.

MEUS PROJETOS

Consulta por qualquer campo:

A listagem abaixo exibe a relação de todos os seus projetos. Você poderá selecionar um projeto como favorito para trabalhar com ele.

(1 of 1) 1 10

Codigo	Projeto	Organismo	Nome Científico	Tipo de Organismo	Data de Criação	Pipeline	Excluir
2	Aroeira	Planta	Myracrodun Urundeuva	Viridiplantae	09/12/2013		

(1 of 1) 1 10

© 2015 - WISE - Sistema Automático de Processos Biotecnológicos
Usuário: Sandra

Fonte: Autoria própria.

A próxima tela que será mostrada (Figura 28) é a de execução do *pipeline*. A aba arquivo mostra dois itens: o primeiro é responsável por redirecionar o usuário para a tela de execução do *pipeline* e o segundo é utilizado caso o usuário necessitar realizar *download* de seu projeto para seu computador pessoal. Nesse caso, será disponibilizado um arquivo .zip com todo o conteúdo (arquivos de cada fase do *pipeline*, o vetor e os arquivos cromatogramas) da pasta de seu projeto.

A Figura 29 mostra a tela de visualização dos arquivos que estão no banco de dados e a Figura 30 a tela de listagem dos arquivos de saída do programa Phred e o conteúdo de um arquivo selecionado.

Quando for selecionada a opção referente à saída de qual programa o usuário deseja ver, serão listados todos os arquivos de saída daquele programa em uma tabela. Para visualizá-lo, basta clicar com a seta do *mouse* sobre o *link* “visualizar” que será redirecionado para a página de listagem do arquivo desejado.

Figura 28 –Tela de execução do *pipeline*.

MONTAGEM DE ESTs

© 2015 - WISE - Sistema Automático de Processos Biotecnológicos
 Usuário: Sandra

BIOTECNOLOGIA
UNAERP

Fonte: Autoria própria.

Figura 29 – Tela com informações de saída do PHRED.

ARQUIVOS PHRED

A listagem abaixo exibe a relação de todos os arquivos '.PHD'.

(1 of 1) 1 10

Codigo	Nome	
3	AROFOPLO2B04.scf.phd.1	Visualizar
4	AROFOPLO4A01.ab1.phd.1	Visualizar
5	AROFOPLO2A06.scf.phd.1	Visualizar
6	AROFOPLO2A09.scf.phd.1	Visualizar

(1 of 1) 1 10

thred_listar.jsf

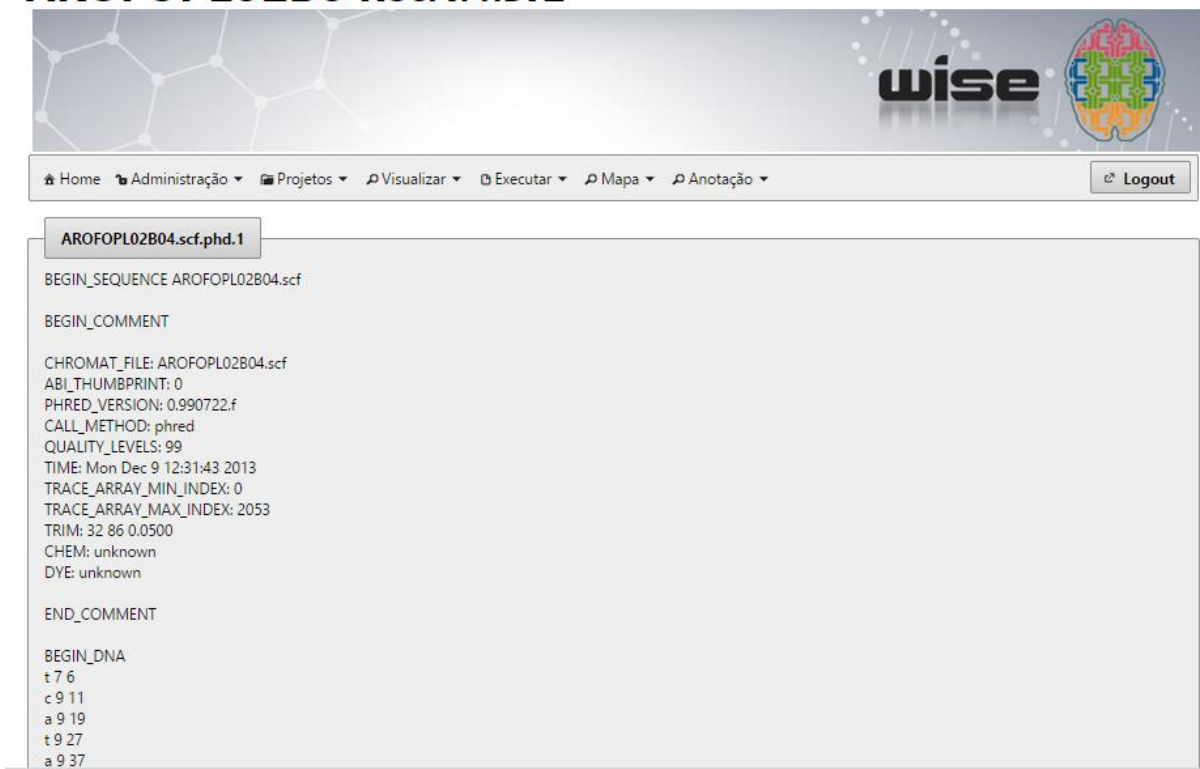
© 2015 - WISE - Sistema Automático de Processos Biotecnológicos
 Usuário: Sandra

BIOTECNOLOGIA
UNAERP

Fonte: Autoria própria.

Figura 30 - Listagem dos arquivos após saída do PHRED.

AROFOPLO2B04.scf.PHD.1



The screenshot shows a web application interface. At the top, there is a header with the 'wise' logo and a navigation menu with items: Home, Administração, Projetos, Visualizar, Executar, Mapa, and Anotação. A 'Logout' button is also present. Below the header, a tab labeled 'AROFOPLO2B04.scf.phd.1' is active. The main content area displays the following text:

```

BEGIN_SEQUENCE AROFOPLO2B04.scf

BEGIN_COMMENT

CHROMAT_FILE: AROFOPLO2B04.scf
ABI_THUMBPRINT: 0
PHRED_VERSION: 0.990722.f
CALL_METHOD: phred
QUALITY_LEVELS: 99
TIME: Mon Dec 9 12:31:43 2013
TRACE_ARRAY_MIN_INDEX: 0
TRACE_ARRAY_MAX_INDEX: 2053
TRIM: 32 86 0.0500
CHEM: unknown
DYE: unknown

END_COMMENT


BEGIN_DNA
t 7 6
c 9 11
a 9 19
t 9 27
a 9 37
  
```

Fonte: Autoria própria.

Para visualizar o mapa metabólico da enzima encontrada, basta clicar na guia Mapa e, depois, na opção KEGG; ocorrerá um redirecionamento para o portal do banco de dados KEGG (<http://www.genome.jp/kegg/pathway.html>).

A Figura 31 mostra um exemplo da página do KEGG. Em seguida, digita-se o nome da enzima e aparecerá o mapa metabólico e as descrições detalhadas dele (Figura 32).

Figura 31 – Portal KEGG.



KEGG PATHWAY Database
Wiring diagrams of molecular interactions, reactions, and relations

KEGG2 PATHWAY BRITE MODULE KO GENOME GENES LIGAND DISEASE DRUG DBGET

Select prefix: Enter keywords: [Help](#)

[[New pathway maps](#) | [Update history](#)]

Pathway Maps

KEGG PATHWAY is a collection of manually drawn [pathway maps](#) representing our knowledge on the molecular interaction and reaction networks for:

- 1. Metabolism**
Global/overview Carbohydrate Energy Lipid Nucleotide Amino acid Other amino Glycan Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Chemical structure
- 2. Genetic Information Processing**
- 3. Environmental Information Processing**
- 4. Cellular Processes**
- 5. Organismal Systems**
- 6. Human Diseases**

and also on the structure relationships (KEGG drug structure maps) in:

- 7. Drug Development**

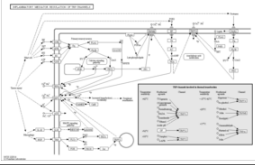
Pathway Mapping

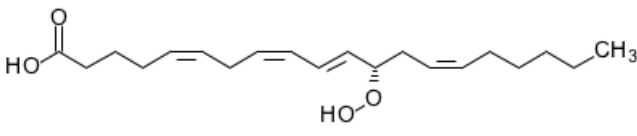
KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretation of higher-level systemic functions.

- [Search Pathway](#) - basic pathway mapping tool
- [Search&Color Pathway](#) - advanced pathway mapping tool
- [Color Pathway](#) - selected pathway map coloring tool

Fonte: <http://www.genome.jp/kegg/> acessado em 22 de julho de 2014.

Figura 32 – Mapa Metabólico da enzima e descrição.

Pathway Text Search					
Number of entries in a page 20 <input type="button" value="Hide thumbnail"/>					
Items : 1 - 1 of 1					
Entry	Thumbnail Image	Name	Description	Object	Legend
map04750		Inflammatory mediator regulation of TRP channels	The TRP channels that exhibit a unique response to temperature have been given the name thermo-TRPs...	C00002 (ATP) C00306 (Bradykinin) C00780 (Serotonin) C00388 (Histamine) C00584 (Prostaglandin E2) C00...	...d in thermal transduction Channel Capsaicin Lipoxygenase products Resiniferatoxin NADA Anandamide Et...
Items : 1 - 1 of 1					
[PATHWAY KEGG]					

KEGG		COMPOSTO: C05965	Help
Entrada	C05965 Composto		
Nome	12 (S) -HPETE, (5Z, 8Z, 10E, 14Z) - (12S) -12-Hydroperoxyicosa-5,8,10,14-tetraenóico ácido; ácido 12-Hydroperoxyicosatetraenoic		
Fórmula	C20H32O4		
Massa exacta	336.2301		
Peso Mol	336.4657		
Estrutura	 <p>C05965</p> <p>Mol file KCF file DB search Jmol KegDraw</p>		
Reação	R01596 R07039 R07040		
Caminho	map00590 Metabolismo do ácido araquidônico map01100 Vias metabólicas map04726 Sinapse serotoninérgica map04750 Regulação mediador inflamatório de canais TRP		
Enzima	1.13.11.31 5.4.4.-		
Brite	Lipídios [BR: br08002] FA Fatty acilos FA01 Fatty Acids e conjugados FA0104 Hydroperoxy ácidos graxos C05965 (5Z, 8Z, 10E, 14Z) - ácido (12S) -12-Hydroperoxyicosa-5,8,10, FA03 Eicosanoids FA0206 hidroperóxidos / hidroperóxidos tetraenoicos		

Fonte: <http://www.genome.jp/kegg/> acessado em 22 de julho de 2014.

Clicando no botão Anotação, é possível fazer um *link* para o portal UniProt, que é um banco de dados de proteínas com a função de proporcionar à

comunidade científica acesso às informações funcionais. A anotação é a etapa que o pesquisador procura identificar os genes presentes nas sequências *contigs e singlets*, as funções e relações filogenéticas. O UniProtKB é uma Base de Conhecimento de informações funcionais em proteínas com anotação precisa e consistente. Após clicar no botão Anotação, digitar o nome da categoria desejada e clicar no botão *search* (Figura 33).

Figura 33: Exemplo de predição de função de uma proteína.

The screenshot shows the UniProtKB entry for P08170 - LOX1_SOYBN. The protein is identified as Seed linoleate 13S-lipoxygenase-1 from Glycine max (Soybean). The function section describes its role in plant physiology, including growth and development, pest resistance, and senescence. It details the catalytic activity, cofactor (Fe cation), pH dependence (inactive below pH 6.0), and pathway (Lipid metabolism; oxylipin biosynthesis). A table at the bottom lists feature keys, positions, lengths, and descriptions.

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions

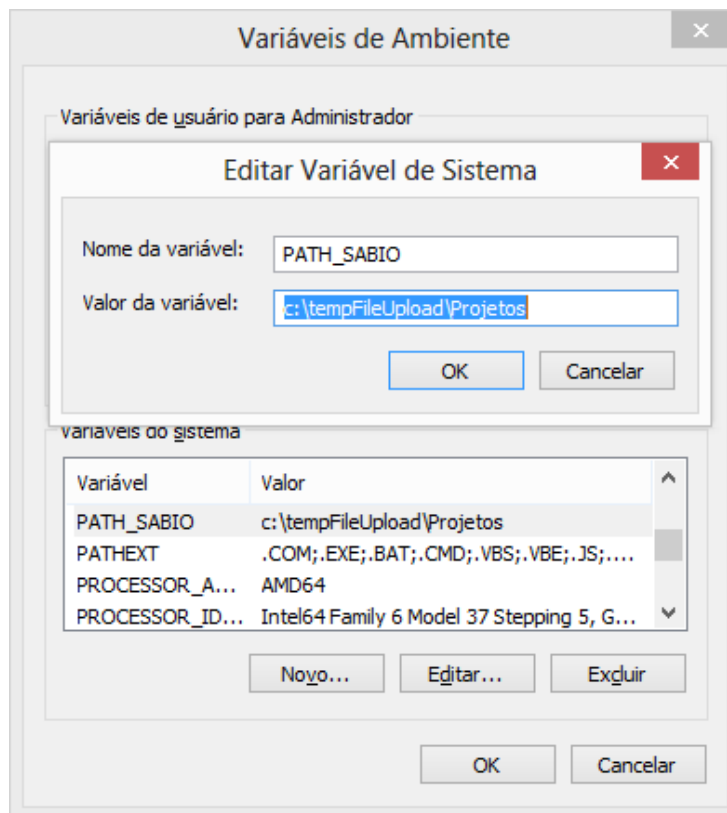
Fonte: <http://www.uniprot.org/> acessado em 22 julho de 2014.

Vale salientar que para o sistema WISE identificar onde deverá criar as pastas no servidor, é importante que seja configurada uma variável de ambiente no sistema operacional. Nessa variável deve possuir o nome de PATH_WISE e em seu valor deve-se informar o caminho em que as pastas serão salvas. A Figura 34 mostra um exemplo de criação da variável de ambiente.

No processo de automatização do sistema WISE, foi necessário implementar um *script* para executar a sequência de programas utilizados no *pipeline* de montagem. Esse *script* foi desenvolvido como uso do *Shell script* no sistema operacional Linux. Assim, quando o usuário clicar na guia para a

execução do *pipeline* no sistema WISE, invocará um método de uma classe Java, que é responsável por iniciar o *script*. Essa fase do desenvolvimento foi implementada no sistema operacional Linux devido ao fato de os *softwares* utilizados no *pipeline* só funcionarem no mesmo.

Figura 34 - Exemplo de criação de uma variável de ambiente.



Fonte: Sistema Operacional Linux baseado na distribuição Fedora 17.

2.3 TESTE DE EXECUÇÃO DO SISTEMA WISE.

Na fase de testes, é importante executar o sistema sob ponto de vista de seu usuário. O sistema é testado completamente, isto é, todas as funcionalidades do mesmo e a integração com outros sistemas são testadas. Com a execução dos testes em todas as fases do desenvolvimento do sistema é possível detectar falhas com antecedência e homologar o sistema com mais qualidade. Foram testados os seguintes atributos no sistema WISE: Funcionalidade, interface, performance, usabilidade, aceitação, *stress*,

configuração, instalação e segurança. No quadro 4 é possível observar a descrição dos atributos que foram devidamente testados no sistema WISE.

Quadro 4 : Tipos de teste.

Tipo de Teste	Descrição
Teste Funcionalidade	Testar as funcionalidades, requerimentos, regras de negócio presentes na documentação.
Teste de Interface	Verifica se a navegabilidade e os objetivos da tela funcionam como especificados e se atendem da melhor forma ao usuário.
Teste de Performance	Verifica se o tempo de resposta é o desejado para o momento de utilização da aplicação.
Teste de Usabilidade	Teste focado na experiência do usuário, consistência da interface, <i>layout</i> , acesso às funcionalidades.
Teste de aceitação do usuário	Testa se a solução será bem vista pelo usuário.
Testes de <i>stress</i>	Testar a aplicação em situações inesperadas.
Testes de Configuração	Testar se a aplicação funciona corretamente em diferentes ambientes de <i>hardware</i> ou de <i>software</i> .
Testes de Instalação	Testar se a instalação da aplicação foi OK.
Testes de Segurança	Testar a segurança da aplicação das mais diversas formas. Utilizar os diversos papéis, perfis, permissões, para navegar no sistema.

Fonte: (MOLINARI, 2015) acessado em <http://testesdesoftware.blogspot.com.br/2009/09/tipos-de-teste.html> - adaptado pela autora.

Os testes de execução do sistema WISE foram realizados nos sistemas operacionais Linux baseado nas distribuições Fedora 17 e Oracle Linux. Para a sua realização, os seguintes passos foram realizados:

- a) Instalação e configuração do JDK;
- b) instalação e configuração do banco de dados MySQL, do servidor Apache Tomcat e da IDE Eclipse no sistema operacional Linux;
- c) criação da variável de ambiente `PATH_WISE`;
- d) criação de um novo usuário no sistema;
- e) criação de um novo projeto no sistema e *upload* dos arquivos;
- f) verificação das pastas criadas no servidor;
- g) execução do *pipeline*;
- h) verificação dos arquivos criados no servidor;
- i) visualização dos arquivos no sistema;
- j) *download* do projeto para o computador pessoal.

O primeiro destaque sobre essa fase de testes é referente à característica das ferramentas (Java, MySQL, Apache Tomcat, etc) utilizadas no desenvolvimento. Todas são multiplataforma, não dificultando a migração entre o sistema operacional utilizado para o desenvolvimento, no caso o Microsoft Windows 7.0 e o Linux. Entretanto, foi necessária a realização de um ajuste no MySQL, pelo fato de o Linux ser *case sensitive*. Para realiza-lo, foi preciso entrar no arquivo de configuração do MySQL no Linux (`/etc/mysql/my.cnf`), procurar a sessão dentro do arquivo chamado `[mysqld]` e adicionar o comando `"lower_case_table_names = 1"`. Ainda nesse arquivo, inseriu-se outra linha de configuração: `"max_allowed_packet=500M"`.

Vale destacar, também, a importância da criação da variável de ambiente `PATH_WISE` contendo a localização de onde serão salvas as pastas criadas pelo sistema WISE. Se essa variável não for criada, o sistema WISE não funcionará. Para cria-la no Linux, basta executar o comando `"export PATH_WISE=/caminho/"`.

Durante a realização dos testes, constatou-se um *"bug"* (erro) em uma classe Java utilizada no desenvolvimento do projeto. A classe `java.util.zip.ZipEntry` possui um *"bug"* que, ao encontrar caracteres especiais

como "~" e "ç", ela lança uma exceção. Essa classe foi utilizada para descompactar arquivos .zip quando o usuário fizer *upload* das sequências cromatogramas no sistema e para compactar o projeto quando o usuário solicitar o *download* do mesmo. Porém, isso não deve atrapalhar o funcionamento do sistema, uma vez que as sequências não utilizam caracteres especiais em seus nomes. Outra maneira de solucionar o referido *bug* é fazer um *upload* das sequências cromatogramas sem elas estarem no formato zip (compactadas), uma vez que o componente utilizado para *upload* suporta o *upload* de vários arquivos de uma vez.

2.4 IMPLEMENTAÇÃO DE CONSULTAS DO SISTEMA WISE VIA APLICAÇÃO *MOBILE*.

Após os testes e homologação do sistema WISE, o propósito seguinte foi implementar uma aplicação *mobile* com o intuito de consultar os projetos, bem como os resultados gerados no sistema WISE.

Atualmente, sistemas móveis são desenvolvidos para processar em *smartphones*, *palmtops*, *tablet*, *notebooks*, televisores, PCs e similares.

Reza B'Far (2005), define que sistemas computacionais móveis são aqueles que podem ser movidos fisicamente e facilmente ou cujas capacidades podem ser utilizadas enquanto eles estão sendo movidos. Destaca-se como vantagens dos sistemas *mobile* a possibilidade de: acessar dados em qualquer lugar e a qualquer momento; reduzir custos de comunicação, pois não é necessário telefonar para outras pessoas para saber informações que já estão disponíveis, acessando-as do próprio dispositivo; otimizar o tempo, já que o sistema poderá enviar e receber informações remotamente, dispensando o deslocamento para outros locais para recebê-las.

A intenção de desenvolver um sistema para celular surgiu quando alguns pesquisadores da Unidade de Biotecnologia se depararam com situações em que precisavam obter informações de seus projetos e pesquisas mesmo quando não estavam presentes fisicamente na unidade. O usuário que possuir o aplicativo instalado em seu celular poderá, por exemplo, pesquisar uma determinada sequência ou projeto em andamento.

O desenvolvimento do WISE *mobile* iniciou com a implementação do WebService, responsável pela consulta no banco de dados oriundos do banco

de dados do sistema WISE. O mesmo foi implementado na linguagem de programação C# (CSharp), utilizando o *software* Visual Studio 2012. O WebService tem por sua característica responder às solicitações do aplicativo em XML - *eXtensible Markup Language* (linguagem de marcação), possibilitando a comunicação entre eles. É uma característica cliente/servidor, sendo que o cliente é o aplicativo que estará rodando no celular e a comunicação com o servidor deve ser estabelecida para que os dados sejam transmitidos. Ao receber os dados, o servidor deve processá-los e, então, encaminhá-los de volta como resposta à requisição enviada pelo cliente.

A aplicação do WebService foi instalada no servidor localizado na rede de computadores do LIAPE – Laboratório de Informática para Atividades de Pesquisa e Ensino da UNAERP (Universidade de Ribeirão Preto). São usados o sistema operacional Windows Server 2012 e a ferramenta do próprio sistema operacional o (IIS - *Internet Information Services*), que é responsável pela locação da aplicação, sendo apto em responder solicitações.

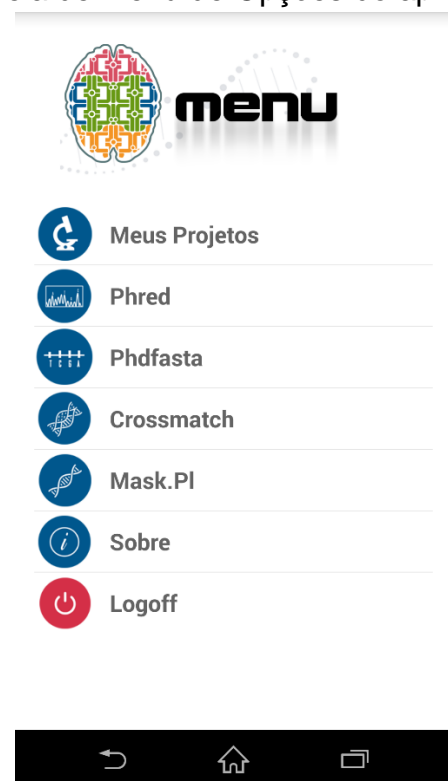
O aplicativo foi desenvolvido na IDE Eclipse, com o uso da linguagem de programação Java com o *plugin* SDK, e realiza solicitações ao WebService, além de mostrar a resposta no *design mobile*.

As funções do WISE *mobile* são:

- tela de autenticação do usuário – o aplicativo requisita o usuário e a senha com o intuito de validar se o usuário está cadastrado para acesso ao sistema (Figura 35);
- menu com opções de listar Projetos, consulta phred, phfasta, crossmatch e mask.pl. - cada opção exibe o resultado gerado pelo projeto escolhido (Figura 36).

Figura 35– Tela de autenticação do usuário para acesso ao aplicativo *mobile*.

Fonte: Autoria própria.

Figura 36– Tela de Menu de Opções do aplicativo *mobile*.

Fonte: Autoria própria.

3.0 RESULTADOS E DISCUSSÃO

Como resultado, o programa SequenceName Converter, implementado na linguagem de programação C# (CSharp), executou com sucesso a função de renomear os nomes das sequências. A interface do programa é simples, pois, após o usuário definir o formato desejado para o novo nome da sequência, o tempo de processamento foi de apenas 3 minutos e meio para um total de 1.050 sequências.

O *pipeline* definido para as análises da *M.urundeuva* demonstrou-se eficiente ao executar cada um dos programas propostos. Os resultados comprovaram que: houve a identificação dos picos de qualidade das sequências; converteu de formato cromatograma para sequência FASTA; identificou e retirou o vetor, obtendo-se as sequências FASTA livres do vetor; gerou as sequências consenso, formando os *contigs* e *singlets*.

No que concerne ao alinhamento entre as sequências, o programa Online Blast XLS Reporter automatizou o processo manual de fazer BLASTx. Com os arquivos de *singlets* e *contigs* gerados pelo sistema WISE, foram feitos os alinhamentos, utilizando-se o algoritmo BLASTx de comparação de sequências de nucleotídeos com base de dados não redundante de proteínas. O programa Online Blast XLS Reporter realizou com sucesso o BLASTx em batelada (grande quantidade) e o tempo de execução também é um ponto forte do programa, pois processou o BLASTx das sequências *singlets* e *contigs* do projeto *M.urundeuva* em 26 horas; já o processo manual que era realizado pelo biólogo molecular demorava dias ou até mesmo semanas. Gerou-se uma planilha no Microsoft Excel com as informações de *id* da sequência, nome, *score*, *e-value (expect)*, *identities*, *positities* e *gaps*. Nos anexos deste projeto, estão disponíveis as planilhas geradas pelo programa Online Blast XLS Reporter.

Como resultado, também pode ser destacado a implantação do sistema WISE que, uma vez homologado, proporcionou aos pesquisadores interoperabilidade, usabilidade e agilidade na execução das análises, ao ser comparado com o processo de técnicas de manipulação antes utilizado. Os pesquisadores cadastrados no sistema WISE podem visualizar seus projetos de pesquisa, as sequências específicas de cada projeto de forma organizada e registrar as suas anotações. Destaca-se, ainda, a disponibilidade do sistema

WISE 24 horas por dia, 7 dias na semana, sendo possível acessá-lo de qualquer lugar que tenha Internet. Não é preciso instalar nada no computador do usuário, somente ter o acesso a um *browser*.

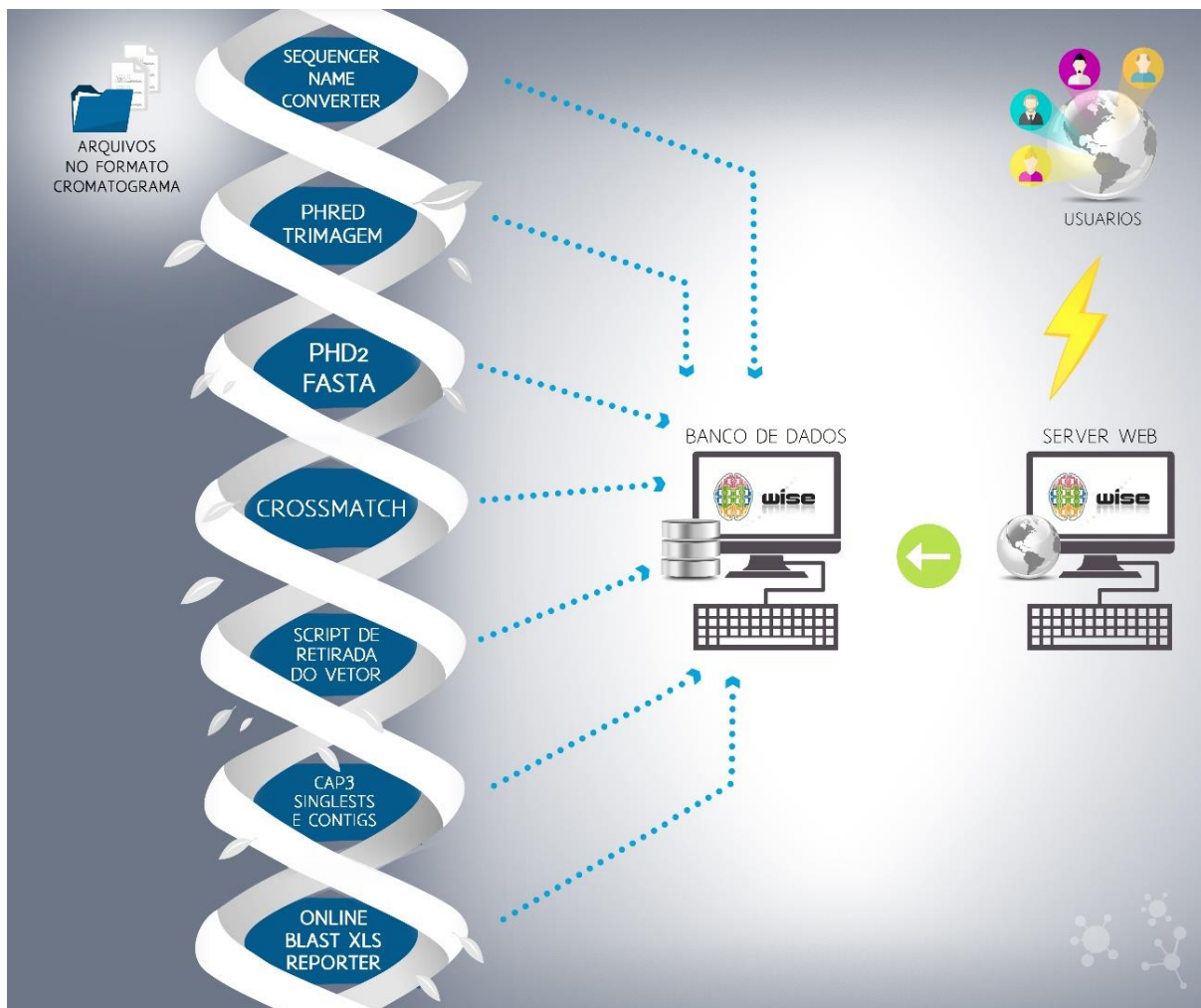
Sobretudo, o sistema WISE proporciona aos pesquisadores informações necessárias para que eles atuem diretamente em suas pesquisas. Promove suporte ao controle, à integração das pesquisas realizadas e em andamento, e suporte ao processo decisório nos diversos níveis da pesquisa. Partindo-se dessa afirmação, é possível atingir eficiência e agilidade, desenvolver e integrar novas pesquisas, estreitar o relacionamento com outros pesquisadores da Unidade de Biotecnologia ou, até mesmo, com outros pesquisadores da Universidade, além de melhorar a tomada de decisão em termos de precisão e velocidade.

Nos testes de utilização desse sistema, o processamento das sequências de *M.urundeuva* no formato cromatograma via WISE foi rápido e eficiente. E é importante destacar que ele poderá ser utilizado para qualquer outro projeto composto por sequências brutas, oriundas de sequenciamento no formato cromatograma. Ainda na fase de testes, o *bug* encontrado na classe `java.util.zip.ZipEntry` foi resolvido realizando-se o *upload* das sequências cromatogramas descompactadas em seu formato original, uma vez que o componente utilizado para *upload* suporta vários arquivos de uma vez.

Na Figura 37, pode-se observar uma representação gráfica da realização do *pipeline* via sistema WISE, definido para o projeto *M.urundeuva*. Demonstra cada fase do *pipeline* sendo gravada automaticamente no banco de dados. Observa-se, também, a interação do usuário com o sistema WISE, utilizando-se da comunicação via Internet. Há, ainda, representação da arquitetura com relação ao servidor onde está publicado o sistema WISE.

No que se refere à utilização da aplicação WISE *mobile*, é necessária apenas a comunicação com a Internet. A escolha do sistema operacional foi muito importante, pois precisava atender a maioria dos usuários dentro do contexto do projeto que o sistema está inserido. No caso do projeto WISE *mobile*, a opção pelo sistema operacional Android foi feita em razão de que se destaca com mais frequência na dinâmica do mercado atual e é usado pela maioria dos pesquisadores da Unidade de Biotecnologia Vegetal da UNAERP. O Wise *mobile* mostrou-se eficiente em termos de usabilidade e segurança.

Figura 37– Representação esquemática da execução do *pipeline* de programas no sistema WISE.



Fonte: Autoria Própria.

A análise quantitativa desta pesquisa foi feita a partir dos dados obtidos após o processamento das sequências da *M. urundeuva* via sistema WISE. No processo de geração dos *singlets* e *contigs*, foram geradas 1.050 sequências *singlets* e 36 *contigs* (composto pelo total de 101 sequências); sendo 91,22% de sequências *singlets* e 8,78% de sequências *contigs*.

Por meio da Figura 38, é possível observar que dos 91,22% de *singlets*, 85,51% são sequências que não encontraram similaridades com outras

sequências de plantas submetidas ao NCBI e 14,49% encontraram funcionalidade putativa dos transcritos de *M. urundeuva*.

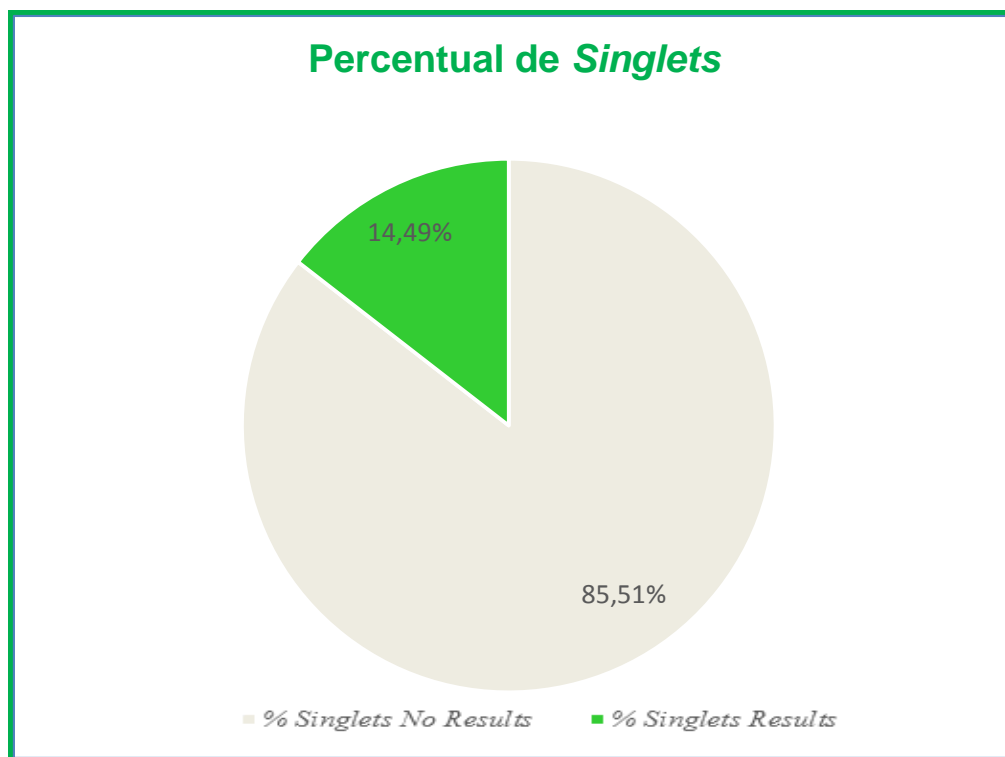
No processamento do Online Blast XLS Reporter, identificou-se no Blastx das sequências *singlets* uma seleção de genes alvos relacionados às proteínas das categorias citadas a seguir:

- Adenosine
- Aquoporin
- Aminotransferase
- Calmodulin
- Cysteine
- Cystatin
- Cyslathionine
- Cyclophlin
- Flavanol
- Flavanol 3 – sulfotransferase
- Flavanol 3
- Flavanol 4
- Isoprene
- Lipoxygenase
- Phosphatase
- Serine
- Sulfotransferase 2A
- Tyrosine
- Ubiquitin
- Metallothionein
- Metallothionein 1
- Metallothionein tipo 2
- Xyloglucan

Para anotações das funcionalidades das categorias acima citadas, o processo deu-se via guia Anotação no sistema WISE, que faz *link* na base de dados do UniProtKB. O UniProtKB é uma Base de Conhecimento de informações

funcionais em proteínas, com anotação precisa e consistente. Em seguida, digitou-se o nome da categoria desejada e clicou-se no botão *Search*.

Figura 38 – Percentual de *singlets no results* e *singlets* que apresentam resultados com plantas.



Fonte: Autoria Própria.

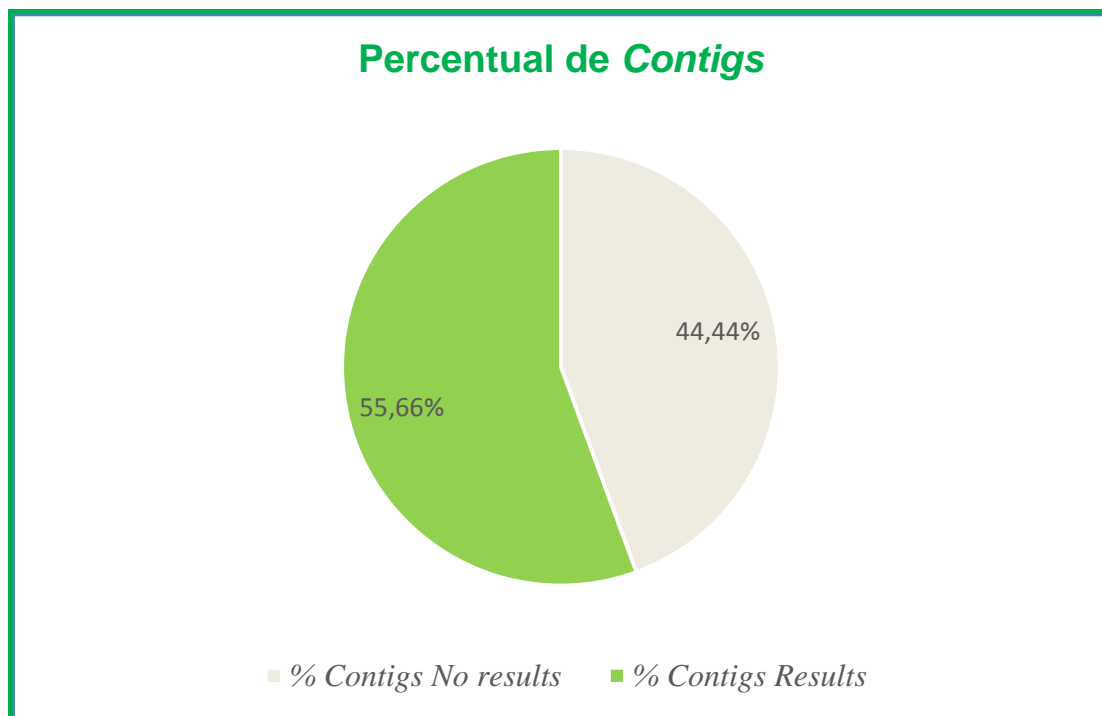
A Figura 39 demonstra que dos 8,78 % de *contigs*, 44,44% não encontraram similaridades com outras sequências de plantas e 55,56% encontraram funcionalidade putativa dos transcritos de *M. urundeuva*.

No processamento do Online Blast XLS Reporter, foi encontrado no Blastx das sequências *contigs* uma seleção de genes alvos com as respectivas funcionalidades:

- Flavonol 3-sulfotransferase
- Flavonol 4-sulfotransferase
- Lipoxygenase
- Miraculin
- Metallothionein
- Metallothionein 1A

- Metallothionein 2B
- Sulfotransferase 2 A

Figura 39 – Percentual de *contigs no results* e *contigs* que apresentam resultados com plantas.



Fonte: Autoria Própria.

No processamento do Online Blast XLS Reporter, identificou-se no Blastx das sequências *singlets* uma seleção de genes alvos relacionados às proteínas das categorias citadas a seguir:

✓ Miraculin

É uma proteína com propriedade incomum de modificar o sabor amargo para sabor doce. A sequência primária da proteína purificada isolada da *Synsepalum dulcificum* "miracle fruit" foi determinada por Theerasilp, Kurihara, (1988), como uma cadeia polipeptídica contendo 191 aminoácidos. Resíduos de asparagina dessa proteína estão ligados a N-glicosídeos. E proteínas dessa classe são membros da família de inibidores de proteases, com papel importante na defesa de plantas contra insetos. Miraculina tem alta homologia com a proteína inibidora de tripsina da soja (THEERASILP *et al.*, 1989).

Proteínas doces com potencial uso como alternativa para substituir adoçantes de baixa caloria, como ciclamatos e sacarina, têm despertado contínuo interesse quando o assunto é adoçante de fontes naturais com melhor valor nutricional. Tais proteínas não funcionam como gatilho de demanda de insulina em pacientes diabéticos assim como faz o açúcar sacarose. Em humanos, a percepção do sabor doce está relacionada a receptores heterodiméricos acoplados a uma proteína G com sabor específico. Esses receptores reconhecem diversos adoçantes naturais e sintéticos, como miraculina, monelina, brazzeína, thaumatina, curculina, mabinlina, pentadina (GNANAVEL, SERVA PEDDHA, 2011).

Hirai *et al.* (2010) demonstraram que uma proteína modificadora de sabor do tipo miraculina é acumulada em alto nível (10% da proteína solúvel total dos frutos) em frutos maduros de *Richadella dulcifica*, também conhecida como “*miracle fruit*”. Eles mostraram que a proteína nativa é secretada nos espaços intracelulares do fruto da *R. dulcifica* assim como a proteína recombinante produzida em tomate é secretada no meio de cultivo de raízes de tomates transgênicos expressando miraculina.

Mondego *et al.* (2011) isolaram e caracterizaram um gene codificador da proteína CoMir similar à proteína miraculina-like de *R. dulcifica*, e também à proteína do tomate LeMir.

Estudos sobre a funcionalidade de miraculina recombinante de tomate e alface transgênicos foram realizados por Sun *et al.*, 2006; 2007.

Como parte deste trabalho, um gene putativo da miraculina SuMir foi isolado de frutos de *Miracrodruon urundeuva* e descrito pela primeira vez nessa espécie.

✓ Metalotioneínas (MTs)

São proteínas de baixa massa molecular, ricas em resíduos de enxofre e capazes de sequestrar quantidades excessivas de íons metálicos por meio da formação de complexos do tipo tiolatos. Assim, essas proteínas desempenham importante papel na detoxificação de metais e na homeostasis. Membros típicos dessa família de proteínas apresentam duas regiões ricas em cisteínas contendo de seis a oito resíduos Cys, separados por uma região constituída por 41 aminoácidos. A expressão de genes relacionados a metaloproteínas de plantas sujeitas a estresses ambientais demonstra que

essas proteínas (com seu amplo número de resíduos de cisteína) podem funcionar segundo dois tipos de mecanismos: na segregação de metais pesados e no sequestro de moléculas de oxigênio reativas, protegendo as plantas de danos oxidativos.

✓ Peptídeos bioativos

Peptídeos biossintetizados em plantas, assim como aqueles de células animais, desempenham importantes papéis na resposta a agentes causadores de estresse biológico, químico ou físico-químico. Em geral, esses peptídeos são codificados por famílias de genes cujos membros apresentam diversificados padrões temporais de expressão, de modo a realizarem a regulação de processos metabólicos ligados ao desenvolvimento e defesa das plantas.

✓ Aquaporinas

Pertencem à grande família das proteínas intrínsecas de membrana (MIP), as quais formam poros na membrana de células biológicas. A membrana plasmática de uma variedade de células animais e também de células vegetais contém aquaporinas, por meio das quais o fluxo de água para dentro da célula pode ocorrer mais rapidamente do que por difusão pela bicamada fosfolipídica, a qual forma uma contínua barreira ao redor de todas as células. Defeitos genéticos envolvendo genes codificadores de aquaporinas têm sido associados a várias doenças humanas. Aquaporinas do hospedeiro funcionam como peptídeos de defesa ativos contra bactérias, fungos e muitos vírus envelopados (fáceis de matar) e não envelopados (difíceis de matar).

✓ Calmodulin

É um peptídeo que atua como parte da via de sinalização de cálcio e se liga a proteínas alvo para alterar suas funções. Recentemente, Jia *et al.* (2014) investigaram o papel de Calmodulin na tolerância térmica e demonstraram que existe uma interação entre óxido nítrico (NO) e peróxido de hidrogênio (H₂O₂) – elementos-chaves na via de sinalização de choque térmico (*Heat Shock* – HS) – e canais de cálcio Ca²⁺, de modo que a ativação de Calmodulin estimula a atividade DNA-binding de fatores de transcrição HS, bem como o acúmulo de proteínas HS e a decorrente tolerância térmica.

Na área de saúde, uma abordagem que vem ganhando destaque é a busca de novos ativos anti arritmia cardíaca e entre os alvos para controle da

movilidade de cálcio intracelular está a proteína quinase II (CaMKII), Ca(2+) / Calmodulin-dependente (MATSUOKA, 2014).

✓ Laccases

São multicobreoxidoreduções consideradas por muitos da área de biotecnologia como o mais promissor “catalisador verde” da atualidade, devido ao amplo rol de substratos por elas degradados, conferindo a essas enzimas possibilidades diversas de aplicações: como agentes de pré-tratamento na produção de biocombustíveis; como enzima deslignificante; em biorremediação; síntese orgânica; branqueamento de polpa; nas indústrias de alimentos, bebidas e têxtil; na construção de biossensores e células biocombustíveis (MATE, ALCALDE; 2015).

4. CONCLUSÃO

É fato que cada vez mais os pesquisadores precisam da ajuda de sistemas computacionais que os auxiliem nas tarefas de identificação de genes expressos e predição de suas respectivas funcionalidades. Partindo desse princípio, este trabalho justificou a importância da eliciação dos requisitos funcionais e apresentou a estruturação de um *pipeline* de programas específicos para elucidar a regulação no âmbito transcricional do metabolismo de plantas medicinais, com foco na biossíntese de metabólitos secundários da classe dos fenilpropanoides. Nesse estudo, os dados levantados para validar a eficiência do *pipeline* foram obtidos de uma biblioteca de cDNA da *M. urundeuva*, sendo alvos da investigação os genes expressos em folhas dessa espécie.

A partir desse *pipeline*, foi possível decodificar com agilidade e eficiência a análise de eletroferogramas, convertê-los em sequências FASTA livres do vetor, realizar o agrupamento de sequências similares e gerar as sequências consenso, formando os *contigs* e *singlets*. Portanto, o *pipeline* criado é eficiente e apresentou bons resultados.

A criação do aplicativo Sequence Name Converter proporcionou aos pesquisadores uma interface simples de usar e que atende plenamente a demanda para executar a função de renomear sequências.

A realização do BLASTx em batelada, com resultados sistematizados pelo programa Online Blast XLS Reporter, reproduziu de forma sintética os parâmetros de similaridades com genes conhecidos depositados na base de dados de sequências não redundantes do NCBI. Além das vantagens oferecidas pelo sistema WISE, referentes à otimização e integridade dos dados, assim como à interoperabilidade, no geral, os pesquisadores participantes da avaliação do WISE consideram que o sistema atende suas necessidades e expectativas, pois facilitou de maneira significativa o manuseio dos projetos cadastrados, possibilitando a identificação e localização fácil dos organismos, perfil de genes expressos em órgãos específicos, categorização de genes expressos e anotações de suas respectivas funcionalidades putativas, além de proporcionar exatidão e agilidade nas consultas.

Desse modo, os resultados obtidos permitem afirmar que o desempenho do sistema foi avaliado como satisfatório; é eficiente no tempo de

resposta e de processamento em relação ao volume de informações produzidas na execução de suas funções. Sobretudo, o WISE viabilizou aos pesquisadores a integração dos resultados de seus projetos e a independência para realizar análises de sequências em razão da facilidade na usabilidade e navegabilidade.

No quesito segurança do sistema, o WISE controla o acesso por meio de senhas e é eficiente na capacidade de evitar o acesso não autorizado, acidental ou intencional aos dados. No que se refere à aplicação WISE *mobile*, a consulta dos resultados de projetos gerados a partir do sistema WISE via *smartphones* e *tablets* é eficiente, pois a aplicação apresenta interface simples e intuitiva aos usuários.

5. REFERÊNCIAS

ANDRADE, T. F. **DWJSF – Desenvolvimento Web com JavaServer Faces**. Apostila do curso de JavaServer Faces da Algaworks Softwares e Treinamentos. 2ª ed. Uberlandia. setembro/2010.

ALBUQUERQUE, R.J.M. **Análise Clínica, Morfológica e Morfométrica da Contadajutivite Alérgica Induzida por Ovalbumina, em Cobaias, e Tratada com Chalconas (Muracradruon urundeuva Fr. All.)**. (Dissertação). Universidade Federal do Ceará – UFC, 86 f. Fortaleza, 2003.

ALEX,B., TAYLOR,L., WINCH, R. **Spring Security is a powerful and highly customizable authentication and access-control framework. It is the de-facto standard for securing Spring-based applications**. 2014 - Disponível em <http://docs.spring.io/autorepo/docs/spring-security/4.0.0.CI-SNAPSHOT/reference/htmlsingle/> acessado em 14 janeiro 2014.

AMERSHAM BIOSCIENCES. **MegaBACE DNA Analysis System: Enabling technology for DNA sequencing and genotyping laboratories**. 2003a, 4 p.

Aroeira do Sertão - Fotografia

http://www.cnip.org.br/banco_img/Aroeira/myracrodruponurundeuvaallemo3.html
Acessado março 2015 às 20:00

BANDEIRA, M.A.M. **Aroeira-do-sertão (Myracrodrupon urundeuva Allemão): cosntituíntes químicos ativos da planta em desenvolvimento e adulta**. Tese (Doutorado em Química Orgânica) – Departamento de química Orgânica e Inorgânica. Universidade Federal do Ceará Fortaleza, 2002.

BAIROCH, A.; APWEILER, R. **The SWISS-PROT Protein Sequence Data Bank and Its New Supplement TREMBL**. Oxford Journals Science & Mathematics Nucleic Acids Research Volume 24, Issue 1 Pp. 21-25. 2010.

BASSI, S.; GONZALEZ V. **DNA Linux Virtual Desktop Edition**. Disponível em <<http://dx.doi.org/10.1038/npre.2007.670.1>>. Acesso em 14/05/2013.

BENNETT, R.N.; WALLSGROVE, R.M **Secondary metabolites in plant defence mechanisms**. New Phytologist. Volume 127, Issue 4, pages 617-633, (1994).

BERNARDES, NR., GLÓRIA L.L., NUNES, C.R., PESSANHA, F.F., MUZITANO M.F. **Quantificação dos teores de taninos e fenóis totais e avaliação da atividade antioxidante dos frutos de Aroeira**. *Vértices*. 2011;13(3):117-28.

BLAST <http://blast.ncbi.nlm.nih.gov/Blast.cgi> acessado em 20/01/2013 às 3:01.

BORATYN, G.; CAMACHO, C.; FEDERHEN, S.; MERZHUH, Y.; MADDEN, T.; SCHOCH, C.; ZARETSKAYA, I. **MOLE-BLAST a new tool to search and classify multiple sequences**. National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health - <ftp://ftp.ncbi.nlm.nih.gov/blast/documents/> acessado em 17/01/2015

CAELUM. Apostila do curso FJ-21: Java para Desenvolvimento Web. Disponível em: < <http://www.caelum.com.br/apostila-java-web/>>. Acesso em 04/04/2013.

CAMPOS, A.C.S. **Estudo do uso do creme vaginal de aroeira do sertão (Myracrodruon urundeuva-Allemano) em pacientes atendidas no ambulatório de ginecologia de uma unidade básica de saúde em Fortaleza**. Dissertação (Mestrado em Ciências Farmacêuticas) Universidade Federal do Ceará. Fortaleza, 2008.

CARAZZOLLE, M.F. **Tratamento e montagem de sequências**, 2008
http://www.lge.ibi.unicamp.br/lgeextensao2008/extsup/anali_mont.pdf

DAMÁSIO, A. DAMÁSIO, H. **O Cérebro e a Linguagem**. Revista: Viver Mente & Cérebro Scientific American. Ano XIII Nº143 - Dezembro 2004

DEITEL, H. *et al.* **Perl: Como Programar**. 1ª ed. Porto Alegre: Bookman, 2002. 952 p.

DENNIS A. BENSON, D.A.; CAVANAUGH, M.; CLARK, K.; KARSCH-MIZRACHI, L.; LIPMAN, D.J.; OSTELL, J.; SAYERS, E.W. **GenBank**. Oxford Journals Science & Mathematics Nucleic Acids Research Volume 41, Issue D1 Pp. D36-D42.

DNAI **Shotgun sequencing and dealing with repeat sections, 3D animation with basic narration** - acessado em <http://dnalc02.cshl.edu/mydnalc/?a=dnai:614> em 12/2014 as 21:00

ELMASRI, R.; NAVATHE, S. B. **Sistemas de banco de dados**. 6ª ed. São Paulo: PearsonAddison Wesley, 2011. 788 p.

EMBRAPA.Recursos Genéticos e Biotecnologia. **O programa BLAST: Guia prático de utilização**. ISSN 0102-110, Brasília, dezembro de 2007.

EWING, B. et al. **Base-calling of automated sequencer traces using phred. accuracy assessment**. Genome Research, v.8, p. 175--185, 1998.

EWING, B.; Green, P. **Base-calling of automated sequencer traces using phred error probabilities**. Genome Research, v.8, p. 186--194, 1998.

FERREIRA, P.M.; FARIAS,D.F.; VIANA, M.P; SOUZA T.M., VASCONCELOS, I.M.; SOARES, B.M.; COSTA-LOTUFO, L.V; MORAES, M.O.; CARVALHO, A.F. **Study of the antiproliferative potential of seed extracts from Northeastern Brazilian plants**. Na Acad. Bras. Cienc 2011 83(3):1045-1058. Epub 2011 Jul 15.

FORMENT, J.; GILABERT F.; ROBLES A.; CONEJERO ;NUEZ F; E BLANCA J. M. **EST2uni: uma ferramenta paralela aberta para análise EST automatizada e criação de banco de dados, com uma interface web mining e microarray expressão de integração de dados**. BMC Bioinformatics 2008, 9 : 5

FORMIGHIERI, E. F. **Bioinformática e anotação de genes em *Xanthomonas axonopodis* pv. *citri* e *Xylella fastidiosa*: metabolismo de ferro e biossíntese de pequenas moléculas**. 177 p. Dissertação (Mestrado em Ciências, Área de Concentração: Energia Nuclear na Agricultura) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2002.

FRANÇA, E.L.T., QUEIROZ, T. M. , ARAÚJO, A. R.S., MACÊDO, A. A. M., **Atividade antioxidante pelo método DPPH de extrato vegetal da casca da aroeira (*Myracrodruon urundeuva* Fr. Ail.) VII CONNEPI 2012**.

FREEMAN, R. G. **Oracle, referência para o DBA: Técnicas essenciais para o dia-a-dia do DBA**. 1ª ed. Rio de Janeiro: Elsevier, 2005. 598 p.

GENEBIO – Genética e bioinformática Universidade Federal Da Bahia – IMS/CAT http://www.genebio.ufba.br/?page_id=303 acessado em 12 de dez 2014 as 21:00.

GIADA, M.L.R - **Food Phenolic Compounds: Main Classes, Sources and Their Antioxidant Power.** 2013 - Agricultural and Biological Sciences » "Oxidative Stress and Chronic Degenerative Diseases - A Role for Antioxidants", Department of Basic and Experimental Nutrition, Institute of Nutrition, Health Sciences Center, Federal University of Rio de Janeiro, Brazil

GNANAVEL, M.; SERVA, P.M. **Identification of novel sweet protein for nutritional applications.** Bioinformation. 2011;7(3):112-4. Epub 2011 Sep 28.

GOBBO-NETO, L.; LOPES, N.P. **Medicinal plants: factors of influence on the content of secondary metabolites.** Química Nova vol.30 no.2 São Paulo Mar./Apr. 2007.

GORDON, D.; ABAJIAN, C.; GREEN, P. **Consed: A Graphical Tool for Sequence Finishing.** Disponível em: <<http://bozeman.mbt.washington.edu/consed/distributions/README.24.0.txt>>. Acesso em 20/05/2013

GREEN, P., EWING, B. **PHRED Documentation, 2002**, disponível em <http://www.phrap.org/phredphrap/phred.html>, acessado em 13 de dezembro 2012.

GRUBER, A. Phred/Phrap/Consed Analysis A User's View. **International Training Course on Bioinformatics Applied to Genomic Studies.** Rio de Janeiro, 2001. Disponível em: < <http://goo.gl/O9MPm> >. Acesso em 16/06/2013.

HIRAI T.; SATO, M.; TOYOOKA, K.; SUN, H.J.; YANO, M.; EZURA, H. **Miraculin, a taste-modifying protein is secreted into intercellular spaces in plant cells.** J Plant Physiol. (2010). 167(3):209-15. doi:10.1016/j.jplph.2009.08.001. Epub 2009 Aug 26.

HOHPE, G; WOOLF, B. **Enterprise Integration Patterns: concepção, construção e implantação de Soluções de Mensagens.** Addison-Wesley Longman Publishing Co., Inc., Boston, MA , EUA . 2003.

HOOKOM, J. **Inside Facelets Part 1: An Introduction**. 17 Aug 2005 Disponível em http://www.jsfcentral.com/articles/facelets_1.html acessado em 11 fevereiro 2013.

HUANG, X.; MADAN, A.. **CAP3: A DNA Sequence Assembly Program**. [Journal List Genome Research v.9\(9\); 1999 Sep](#)

HUANG, W.; CAI, Y.; ZHANG, Y. **Natural phenolic compounds from medicinal herbs and dietary plants: potential use for cancer prevention**. Nutrition and Cancer. Taylor & Francis Group, LLC. 2010; 62(1):1-20

ILLUMINA. **History of Illumina Sequencing**. Disponível em: <https://www.illumina.com/technology/solexa_technology.ilmn>. Acesso em 27/05/2013.

JACOBSON, I; SPENCE, I; BITTNER, K. **USE-CASE 2.0 The Guide to Succeeding with Use Cases**. Ivar Jacobson International SA. 2011

JAVASERVERFACES <http://javaserverfaces.java.net/> acessado em 06 de junho 2014 às 23:02

JIA.Z; ZHAO, Q ; ZHANG, C.; HUANG, Y.; LI H.; SONG, S. **Involvement of calmodulin in regulation of primary root elongation by N-3-oxo-hexanoyl homoserine lactone in Arabidopsis thaliana**. Frontiers in Plant Science 01/2014; 5:807. DOI: 10.3389/fpls.2014.00807.

KEGG: Kyoto Encyclopedia of Genes e Genomas. <http://www.genome.jp/kegg/> acessado em 22 de julho de 2014.

KING,G.; BAUER C.; ANDERSEN, M.R; BERNARD, E.; EBERSOLE, S. **HIBERNATE - Relational Persistence for Idiomatic Java 2010**, Disponível em <http://docs.jboss.org/hibernate/orm/3.5/reference/en/html/> Acessado em 12 de fevereiro de 2013.

KURKIN, V.A. **Phenylpropanoids from medicinal plants: distribution, classification, structural análise, and biological activity**. Chemistry of Natural Compounds 39 (2): 123-153. 2003

LESK, A. M. **Introdução à Bioinformática**. 2ª ed. Porto Alegre: Artmed, 2008. 384 p.

LEMOS, M.; SEIBEL, L. F. B.; CASANOVA, M. A. **Sistemas de Anotações em Biossequências**. 66 p. Monografia. Pontifícia Universidade do Rio de Janeiro, Rio de Janeiro, 2003.

LIANG C, F SUN, WANG H, QU J, FREEMAN RM, PRATT LH, **MAGIC-SPP: um pacote de processamento de sequência de DNA para banco de dados com ferramentas de gerenciamento de associados**. Cordonnier-Pratt MM - BMC Bioinformatics (2006)

LIFE TECHNOLOGIES. 3730xl DNA Analyzer. Disponível em: <<https://products.invitrogen.com/ivgn/product/3730XL>>. Acesso em 24/05/2013.

LIFSCHITZ, S. **Algumas pesquisas em Banco de Dados e Bioinformática. Anais do XXVI Congresso da SBC**. Campo Grande, julho/2006.

LORENZI, H. **Árvores Brasileiras**. São Paulo: Instituto Plantarum, 1992. 352p.

LORENZI, H.; MATOS, F.J.A. **Plantas medicinais no Brasil: nativas e exóticas**. São Paulo: Instituto Plantarum, 2002.

LOSSIUS, T.; HOGUE, T.; BALTAZAR, P.; PLACE, T.; WOLEK, N.; RABIN, J. **Model-View-Controller separation in Max using Jamona**. A. Georgaki and G. Kouroupetroglou (Eds.), Proceedings ICMC|SMC|2014, 14-20 September 2014, Athens, Greece.

LUCKOW, D. H.; MELO, A. A. **Programação Java para Web**. 1ª ed. São Paulo: Novatec Editora, 2010. 640 p.

MACHADO, A.C.; OLIVEIRA, R.C. **Phytotherapy medicines in dentistry: evidence and perspectives on the use of "Aroeira-do-sertão" (*Myracrodruon Urundeuva* Allemão)** Rev. bras. plantas med. vol.16 no.2 Botucatu June 2014

MADEIRA, H. M. F. Tutorial Phred/Phrap/Consed. **Anais do I JOBIP – Jornada de Bioinformática do Paraná**. São José dos Pinhais, janeiro/2004.

MAHER, P.; HANNEKEN, A. **Flavonoids Protect Retinal Ganglion Cells from Oxidative Stress-Induced**. Death IOVS, December 2005, Vol. 46, No. 12

MANZANO, J. A. N. G. **MySQL 5.5 interativo: guia essencial de orientação e desenvolvimento**. 1ª ed. São Paulo: Erica, 2011. 240 p.

MATSUOKA, M. **Protective Effects of Humanin and Calmodulin-Like Skin Protein in Alzheimer's Disease and Broad Range of Abnormalities**. *Mol Neurobiol*. Jun 27. [Epub ahead of print]. (2014).

MATE, D.M.; ALCALDE, M. **Laccase engineering: From rational design to directed evolution**. *Biotechnol Adv*. 33(1):25-40.
Doi:10.1016/j.biotechadv.2014.12.007. Epub 2014 Dec 27. (2015)

MATOS, F.J.A. **Plantas medicinais – Guia de seleção e emprego de plantas medicinais no Nordeste do Brasil**. Fortaleza: IOCE, 1989. 164 p.

MELO, H.V.F. **Desenvolvimento de um pipeline para análise genômica e transcriptômica com base em web services**. Dissertação de Mestrado – Pós graduação em Biotecnologia - Universidade Federal de São Carlos. 2009.

MOLINARI, L. **Tipos de Teste**

<http://testesdesoftware.blogspot.com.br/2009/09/tipos-de-teste.html> acessado em 02 fevereiro 2015.

MONDEGO, J.M.; DUARTE, M.P.; KIYOTA, E.; MARTÍNEZ, L.; CAMARGO, S.R.; CAROLI, F.P.; ALVES, B.S.; GUERREIRO, S.M.; OLIVA, M.L.; GUERREIRO-FILHO, O.; MENOSSI, M. **Molecular characterization of a miraculin-like gene differentially expressed during coffee development and coffee leaf miner infestation**. *Planta*, (2011). 233(1):123-37. doi: 10.1007/s00425-010-1284-9. Epub 2010 Oct 8.

MONTEIRO, J.M.; ARAÚJO, E.L.; AMORIM, E.L. 2005. **Taninos: uma abordagem da química à ecologia**. *Quim Nova* 28: 892-896.

MONTEIRO, J.M.; ARAÚJO, E.L.; AMORIM, E.L.C.; ALBUQUERQUE, U.P. **Valuation of the Aroeira (*Myracrodruon urundeuva* Allemão): perspectives on conservation**. *Acta Botanica Brasilica* 26(1): 125-132. 2012

Microsoft <https://msdn.microsoft.com/en-us/library/ff649643.aspx> acessado 09 janeiro 2014 [às 13:10](#).

Multimudas <http://multimudas.com.br> acessado em 16/05/2013 às 01:30

NCBI <http://www.ncbi.nlm.nih.gov/> acessado em 10/02/2013 às 02:12

NILGES, M.; LINGE, J.P. **Unité de Bio-Informatique Structurale**, Bioinformática Institut Pasteur, 25-28 rue du Docteur Roux, F-75015 Paris, França. Disponível em: http://www.pasteur.fr/recherche/unites/Binfs/definition/bioinformatics_definition.html

Acessado em 02 de fevereiro 2014 às 00:32.

[NOBRE, H.V.J.](#); [OLIVEIRA, R.A.](#); [MAIA, F.D.](#); [NOGUEIRA, M.A.](#); [MORAES, M.O.](#); [BANDEIRA, M.A.](#); [ANDRADE, G.M.](#); [VIANA, G.S.](#) **Neuroprotective effects of chalcones from Myracrodruon urundeuva on 6-hydroxydopamine-induced cytotoxicity in rat mesencephalic cells.** [Neurochem Res.](#) 2009 Jun;34(6):1066-75. doi: 10.1007/s11064-008-9876-5. Epub 2008 Nov 13.

NORONHA, M. F. **Controle da execução e disponibilização de dados para aplicativos sobre sequências biológicas: o caso BLAST.** 83 p. Dissertação (Mestrado em Informática) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

NUNES, Y.; FERREIRA, R., FAGUNDES, M. , ALMEIDA, H.S. VELOSO, M. D. M., **ASPECTOS ECOLÓGICOS DA AROEIRA (Myracrodruon urundeuva ALLEMÃO - ANACARDIACEAE): FENOLOGIA E GERMINAÇÃO DE SEMENTES,** R. Árvore, Viçosa-MG, v.32, n.2, p.233-243, 2008.

OKURA V.K. **Bioinformática de Projetos Genoma de Bactérias** Dissertação (mestrado em Ciência da Computação) Universidade Estadual de Campinas Campinas Fevereiro 2002.

ORACLE <http://www.oracle.com/index.html> acessado em 21 de abril de 2013 às 20:40

OTTO, T. D.; VASCONCELLOS E.A; GOMES, L.H.F. ; MOREIRA, A. S; Degraive, W.M.; MENDONÇA, L.-L.; FERREIRA M.A. - **ChromaPipe: a pipeline for analysis, quality control and management for a DNA sequencing facility.** Genetics and Molecular Research 7 (3): 861-871 (2008).

PARKINSON, J.; BLAXTER, M. **Expressed sequence tags: an overview.**

Methods Mol Biol. 2009; 533:1-12. doi: 10.1007/978-1-60327-136-3_1. Review.

PHD2FASTA <http://bldg6.arsusda.gov/mtucker/Public/Consed/phd2fasta.html> acessado em 21 de janeiro de 2013.

PHRAP <http://www.phrap.com/consed/>>. Acessado em 27/05/2013 às 20:32.

[PIROOZANIA, M.](#); [PERKINS, E.J.](#); [DENG, Y.](#) **Batch Blast Extractor: an automated blastx parser application.** [Journal List BMC Genomics v.9 \(Suppl 2\); 2008.](#)

[PEARSON, W.R.](#); [LIPMAN, D.J.](#) **IMPROVED TOOLS FOR BIOLOGICAL SEQUENCE COMPARISON.** [PROC NATL ACAD SCI U S A.](#) 1988 APR;85(8):2444-8.

PERES, L.E.P. **Metabolismo secundário das plantas.** <http://www.oleos essenciais.org/metabolismo-secundario-das-plantas/> acessado em 12 de janeiro 2015 às 21:21

PROSDOCIMI, F. *et al.* **Bioinformática: manual do usuário.** Biotecnologia Ciência e Desenvolvimento, n. 29, p. 18-31, jan.2003.

PROSDOCIMI, F. **Curso Online: Introdução a Bioinformática.** ano 2007, Disponível em: <http://www.iq.usp.br/setubal/bmc/2013/FProsdocimi07_CursoBioinfo.pdf>. Acesso em 24/02/2013 ÀS 14:30.

RANGEL, L.T.; NOVAES, J.; DURHAM, A.M.; MADEIRA, A. M. B. N.; GRUBER, A. **The Eimeria Transcript DB: na integrated resource for annotated transcripts of protozoan parasites of the genus Eimeria.** Database Oxford Journals . Vol.2013.

REZA B'FAR **Mobile computing principles designing and developing mobile applications with uml and xml** Cambridge University Press The Edinburgh Building, Cambridge CB2 2RU, UK – 2005.

[ROBERT, J.](#); [NIJVELDT, R.J.](#); [NOOD, E.V.](#); [HOORN, D.EC.](#); [BOELENS, P.G.](#); [NORREN, K.](#); [LEEUVEN, P.AM.](#) **Flavonoids: a review of probable**

mechanisms of action and potential applications – The American Journal of Clinical Nutrition. ©2001, American Society for Clinical Nutrition

ROCHE. 454 Sequencing: GS FLX+ System. Disponível em: <<https://454.com/products/gs-flx-system/index.asp>>. Acesso em 25/05/2013.

SÁ, R.A.; SANTOS, N.D.; DA SILVA, C.S.; NAPOLEÃO, T.H; GOMES, F.S.;CAVADA, B.S., COELHO, L.C.; NAVARRO, D.M.; BIEBER, L.W.; PAIVA, P.M. **Larvicidal activity of lectins from Myracrodruon urundeuva on Aedes aegypti.** Comp Biochem Physiol C Toxicol Phamacol, 2009; 149:300-306.

[Http://science.howstuffworks.com/alkaloid-info.htm](http://science.howstuffworks.com/alkaloid-info.htm) 01 de Julho de 2009.
Acessado em 26 de fevereiro, 2014.

SANTOS, R. S.; ORTEGA, J. M. **Bioinformática aplicada à Genômica. Manuscrito para capítulo do Biowork IV.** Viçosa, março/2003.

SMITH, C. Linux NFS-HOWTO, 2006, disponível em
<http://nfs.sourceforge.net/nfs-howto/> acessado em 02 fevereiro 2013.

STAJICH, J. E. et al. **The Bioperl toolkit: Perl modules for the life sciences.** Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/12368254#>> Acesso em: 14/06/2013.

SUN, H.J.; CUI, M.L.; MA, B.; EZURA, H. (2006). **Functional expression of the taste-modifying protein, miraculin, in transgenic lettuce.**FEBS Lett. Jan 23;580(2):620-6. Epub 2006 Jan 3.

SUN, H.J.; KATAOKA, H.; YANO, M.; EZURA, H. (2007). **Genetically stable expression of functional miraculin, a new type of alternative sweetener, in transgenic tomato plants.** Plant Biotechnol J. 2007 Nov;5(6):768-77. Epub 2007 Aug 13.

TAIZ, I. ZEIGER E. **Plant Physiology.** 5ª Edition, Hardcover. 782 pages
Publisher: Sinauer Associates, Inc – May 31, 2010

TANENBAUM, A.S., WETHERALL, D.J - **Computer Networks** , 5th Edition, Prentice Hall , 17 Jan 2012

THEERASILP S, KURIHARA Y. **Complete purification and characterization of the taste-modifying protein, miraculin, from miracle fruit.** [J Biol Chem.](#) 1988 Aug 15;263(23):11536-9

THEERASILP, S.; HITOTSUYA, H.; NAKAJO, S.; NAKAYA, K.; NAKAMURA, Y.; KURIHARA, Y. J. **Complete amino acid sequence and structure characterization of the taste-modifying protein, miraculin.** *Biol Chem.* 1989 Apr 25; 264(12):6655-9.

TRAZZI, G. **Lignan synthesis from Morita-Baylis-Hillman adducts : a general strategy to biological active lignans.** Tese (Doutorado em Química) Universidade Estadual de Campinas, Campinas, 2008.

Uniprot, acessado em <http://www.uniprot.org/> em 20/01/2013 4:05

Updados <http://updados.com.br/tecnologias/>

VIANA, G.S.B.; MATOS, F.J.A.; BANDEIRA, M.A.M.; RAO, V.S.N. **Aroeira-do-sertão (Myracrodruon urundeuva Fr. All.) – Estudo botânico, farmacognóstico, químico e farmacológico.** Fortaleza: UFC, 1995. 144 p.

[VIANA, G.S.;](#) [BANDEIRA, M.A.;](#) [MATOS, F.J.](#) **Analgesic and antiinflammatory effects of chalcones isolated from Myracrodruon urundeuva allemão.** [Phytomedicine.](#) 2003 Mar;10(2-3):189-95.

VORLEY, G. **Quality Managment & Training Limited.** 2008, Published by Quality Management & Training (Publications) Ltd.

WALL, L.; CHRISTIANSEN, T; ORWANT, J. **Programaçã Perl.** Editora Campus, 3ª edição, Rio de Janeiro, 2001.

WIECZOREK, E. M; LEAL, E. M. Caminhos e tendências do uso de banco de dados em bioinformática. **IV Encontro de Estudantes de Informática do Estado do Tocantins.** Palmas, 2002.

WOLFSBERG, T.G. ; LANDSMAN, D. **A comparison of expressed sequence tags (ESTs) to human genomic sequences.** National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 1626–1632 Nucleic Acids Research, 1997, Vol. 25, No. 8 - 1997 Oxford University Press.

[YANG, N.](#); [PATIL, S.](#); [ZHUGE, J.](#); [WEN, M.C.](#); [BOLLEDDULA, J.](#); [DODDAGA, S.](#); [GOLDFARB, J.](#); [SAMPSON, H.A.](#); [LI, X.M.](#) **Glycyrrhiza uralensis flavonoids present in anti-asthma formula, ASHMI™, inhibit memory Th2 responses in vitro and in vivo.** [Phytother Res.](#) 2013 Sep;27(9):1381-91. doi: 10.1002/ptr.4862.

YOSHIHARA, E.; MINHO, A.P.; YAMAMURA, M.H. **Anthelmintic effect of condensed tannins in gastrointestinal nematodes of sheep (Ovis aries)** Semina: Ciências Agrárias, Londrina, v. 34, n. 6, suplemento 2, p. 3935-3950, 2013

ZHOU, S.; KVIKSTAD, E.; KILE, A.; SEVERIN, J.; FORREST, D.; RUNNHEIM, R.; CHURAS, C.; HICKMAN, J.W.; MACKENZIE, C.; CHOUDHARY, M.; DONOHUE, T.; KAPLAN, S.; SCHWARTZ, D.C. **Whole-Genome Shotgun Optical Mapping of Rhodobacter sphaeroides strain 2.4.1 and Its Use for Whole-Genome Shotgun Sequence Assembly.** Genome Research 13:2142–2151 ©2003 by Cold Spring Harbor Laboratory Press.



Contig2						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_004294730.1	PREDICTED: non-specific lipid-transfer protein 1-like [Fragaria vesca subsp. vesca]	847297 bits(208)	4.88E-20	38/59(64%)	46/59(78%)	0/59(0%)
gb EMJ07354.1	hypothetical protein PRUPE_ppa013519mg [Prunus persica]	789518 bits(208)	2.22E-18	38/60(63%)	45/60(75%)	1/60(2%)
gb AAF61436.1 AF137353_1	lipid transfer protein precursor [Pisum sativum]	797221 bits(208)	1.61E-17	36/59(61%)	44/59(75%)	0/59(0%)
gb AFK48111.1	unknown [Lotus japonicus]	79337 bits(208)	2.25E-18	35/58(60%)	42/58(72%)	0/58(0%)
ref XP_003553202.1	PREDICTED: non-specific lipid-transfer protein-like [Glycine max]	804925 bits(208)	2.53E-17	35/59(59%)	45/59(76%)	0/59(0%)
gb ACJ83942.1	unknown [Medicago truncatula]	79337 bits(208)	1.13E-17	34/59(58%)	45/59(76%)	0/59(0%)
gb EMJ08287.1	hypothetical protein PRUPE_ppa023836mg [Prunus persica]	73559 bits(208)	2.28E-15	33/59(56%)	41/59(69%)	0/59(0%)
gb ADC80503.1	non-specific lipid transfer protein 1b precursor [Phaseolus vulgaris]	693218 bits(208)	2.96E-12	33/59(56%)	40/59(68%)	0/59(0%)
emb CAH03799.1	lipid transfer protein [Citrus sinensis]	685514 bits(208)	3.32E-12	33/60(55%)	40/60(67%)	0/60(0%)
sp P85894.1 LTP1_MORNI	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Allergen=Mor n 3	67781 bits(208)	5.65E-12	32/58(55%)	39/58(67%)	0/58(0%)
gb AEN70935.1	lipid transfer protein [Gossypium aridum]	673958 bits(208)	1.44E-11	33/60(55%)	41/60(68%)	2/60(3%)
gb AAQ74627.1	lipid transfer protein I [Vigna radiata]	673958 bits(208)	1.45E-11	32/58(55%)	41/58(71%)	0/58(0%)
gb AEN70917.1	lipid transfer protein [Gossypium schwendimanii]	673958 bits(208)	1.60E-11	33/60(55%)	41/60(68%)	2/60(3%)
gb AAT80649.1	lipid transfer protein precursor [Malus x domestica]	662402 bits(208)	3.13E-11	32/58(55%)	38/58(66%)	0/58(0%)
emb CAB96874.1	mal d 3 [Malus x domestica]	65855 bits(208)	3.39E-11	32/58(55%)	37/58(64%)	0/58(0%)
sp Q9M5X7.1 NLTP_MALDO	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Allergen=Mal d 3; Flags: Precursor	662402 bits(208)	3.52E-11	32/58(55%)	37/58(64%)	0/58(0%)
gb ABB96481.1	non-specific lipid transfer protein precursor [Malus x domestica]	662402 bits(208)	3.80E-11	32/58(55%)	37/58(64%)	0/58(0%)
gb AAR22488.1	allergen Mal d 3 [Malus x domestica]	65855 bits(208)	4.35E-11	32/58(55%)	37/58(64%)	0/58(0%)
gb AEN70919.1	lipid transfer protein [Gossypium mustelinum]	646994 bits(208)	1.17E-10	33/60(55%)	40/60(67%)	2/60(3%)
gb AET05733.1	lipid transfer protein isoform d, partial [Pyrus communis]	635438 bits(208)	1.92E-10	32/58(55%)	37/58(64%)	0/58(0%)
gb AAQ74628.1	lipid transfer protein II [Vigna radiata]	635438 bits(208)	2.80E-10	29/53(55%)	36/53(68%)	0/53(0%)

emb CCD33109.1	lipid transfer protein [Vigna radiata]	635438 bits(208)	2,82E-10	29/53(55%)	36/53(68%)	0/53(0%)
sp Q39794.1 NLTP_GERHY	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; Flags: Precursor	623882 bits(208)	3,28E-12	28/52(54%)	34/52(65%)	0/52(0%)
gb AAF35184.1 AF195863_1	lipid transfer protein precursor [Gossypium hirsutum]	670106 bits(208)	5,92E-12	33/61(54%)	41/61(67%)	2/61(3%)
sp P82534.1 NLTP1_PRUDO	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Full=Major allergen Pru d 3; AltName: Allergen=Pur d 3	65855 bits(208)	3,43E-11	32/59(54%)	39/59(66%)	0/59(0%)
gb AAC49860.1	non-specific lipid transfer protein PvLTP-24 [Phaseolus vulgaris]	643142 bits(208)	1,45E-10	32/59(54%)	39/59(66%)	1/59(2%)
ref XP_002276372.1	PREDICTED: non-specific lipid-transfer protein [Vitis vinifera]	747146 bits(208)	8,26E-15	31/59(53%)	42/59(71%)	0/59(0%)
gb AAM21292.1 AF369931_1	lipid-transfer protein [Citrus sinensis]	708626 bits(208)	9,26E-13	32/60(53%)	41/60(68%)	0/60(0%)
dbj BAH03575.1	nonspecific lipid transfer protein [Citrus jambhiri]	69707 bits(208)	1,93E-12	32/60(53%)	40/60(67%)	0/60(0%)
gb AAL27855.1	lipid transfer protein precursor [Davidia involucrata]	685514 bits(208)	4,59E-12	31/58(53%)	40/58(69%)	0/58(0%)
gb AAM22768.1	lipid transfer protein [Prunus persica]	666254 bits(208)	1,49E-11	31/59(53%)	41/59(69%)	0/59(0%)
gb AAK28533.1 AF329829_1	lipid transfer protein precursor [Corylus avellana]	673958 bits(208)	1,51E-11	31/58(53%)	42/58(72%)	0/58(0%)
ref XP_004302608.1	PREDICTED: non-specific lipid-transfer protein-like isoform 1 [Fragaria vesca subsp. vesca]	670106 bits(208)	1,65E-11	31/58(53%)	38/58(66%)	0/58(0%)
gb ABK41612.1	lipid transfer protein [Lilium longiflorum]	666254 bits(208)	2,48E-11	28/53(53%)	38/53(72%)	0/53(0%)
ref XP_002305878.1	predicted protein [Populus trichocarpa]	666254 bits(208)	2,91E-11	31/58(53%)	40/58(69%)	0/58(0%)
gb AGH27866.1	lipid transfer protein precursor [Prunus kansuensis]	65855 bits(208)	5,57E-11	31/59(53%)	40/59(68%)	0/59(0%)
ref XP_003517010.1	PREDICTED: non-specific lipid-transfer protein 1-like isoform 1 [Glycine max]	654698 bits(208)	6,06E-11	31/58(53%)	39/58(67%)	0/58(0%)
pdb 2ALG A	Chain A, Crystal Structure Of Peach Pru P3, The Prototypic Member Of The Family Of Plant Non- Specific Lipid Transfer Protein Pan-Allergens	646994 bits(208)	8,91E-11	31/59(53%)	40/59(68%)	0/59(0%)
sp P81402.1 NLTP1_PRUPE	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Full=Allergen Pru p 1; AltName: Full=Major allergen Pru p 3; AltName: Allergen=Pru p 3	643142 bits(208)	9,18E-11	31/59(53%)	40/59(68%)	0/59(0%)
emb CAB96876.2	pru p 1 [Prunus persica]	643142 bits(208)	9,45E-11	31/59(53%)	40/59(68%)	0/59(0%)
gb AEI70827.1	non-specific lipid-transfer protein [Helianthus annuus]	643142 bits(208)	1,01E-10	28/53(53%)	36/53(68%)	0/53(0%)
gb AGH27864.1	lipid transfer protein precursor [Prunus davidiana]	646994 bits(208)	1,06E-10	31/59(53%)	40/59(68%)	0/59(0%)
gb AAV64877.1	non-specific lipid transfer protein [Prunus persica]	646994 bits(208)	1,08E-10	31/59(53%)	40/59(68%)	0/59(0%)
gb ADR66953.1	non-specific lipid transfer protein [Prunus dulcis]	646994 bits(208)	1,20E-10	31/59(53%)	40/59(68%)	0/59(0%)

gb AEN70933.1	lipid transfer protein [Gossypium davidsonii]	646994 bits(208)	1,38E-10	32/60(53%)	40/60(67%)	2/60(3%)
sp Q43017.1 NLTP1_PRUDU	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; Flags: Precursor	643142 bits(208)	1,50E-10	31/59(53%)	40/59(68%)	0/59(0%)
gb ADR66949.1	non-specific lipid transfer protein [Prunus sargentii]	643142 bits(208)	1,50E-10	31/59(53%)	40/59(68%)	0/59(0%)
sp P81651.2 NLTP1_PRUAR	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Full=Major allergen Pru ar 3; AltName: Allergen=Pur ar 3	63929 bits(208)	1,54E-10	31/58(53%)	38/58(66%)	0/58(0%)
gb ADR66950.1	non-specific lipid transfer protein [Prunus sargentii]	643142 bits(208)	1,61E-10	31/59(53%)	40/59(68%)	0/59(0%)
gb AGH27863.1	lipid transfer protein precursor [Prunus mira]	643142 bits(208)	1,62E-10	31/59(53%)	40/59(68%)	0/59(0%)
gb AA43800.1	FSLTP2 [Gossypium hirsutum]	643142 bits(208)	1,76E-10	32/60(53%)	39/60(65%)	2/60(3%)
gb AAG29777.1	lipid transfer protein 3 precursor [Gossypium hirsutum]	643142 bits(208)	1,88E-10	32/60(53%)	39/60(65%)	2/60(3%)
gb ADY68820.1	lipid transfer protein 3 [Gossypium herbaceum subsp. africanum]	643142 bits(208)	1,95E-10	32/60(53%)	39/60(65%)	2/60(3%)
gb ADR66948.1	non-specific lipid transfer protein [Prunus armeniaca]	63929 bits(208)	1,97E-10	31/58(53%)	38/58(66%)	0/58(0%)
gb ADR66947.1	non-specific lipid transfer protein [Prunus armeniaca]	63929 bits(208)	1,97E-10	31/58(53%)	38/58(66%)	0/58(0%)
gb AEN70918.1	lipid transfer protein [Gossypium turneri]	643142 bits(208)	1,97E-10	32/60(53%)	39/60(65%)	2/60(3%)
gb AA43801.1	FSLTP3 [Gossypium hirsutum]	63929 bits(208)	2,35E-10	32/60(53%)	39/60(65%)	2/60(3%)
gb ACP30545.1	lipid transfer protein precursor [Lilium formosanum]	612326 bits(208)	2,41E-10	28/53(53%)	34/53(64%)	0/53(0%)
gb ABR08655.1	lipid transfer protein [Populus alba x Populus glandulosa]	635438 bits(208)	2,99E-10	31/58(53%)	39/58(67%)	0/58(0%)
gb EMJ07873.1	hypothetical protein PRUPE_ppa023110mg [Prunus persica]	631586 bits(208)	3,97E-10	31/59(53%)	39/59(66%)	0/59(0%)
ref XP_004302624.1	PREDICTED: non-specific lipid-transfer protein 1-like [Fragaria vesca subsp. vesca]	689366 bits(208)	3,60E-12	30/58(52%)	41/58(71%)	0/58(0%)
gb AAF35186.1 AF195865_1	lipid transfer protein precursor [Gossypium hirsutum]	673958 bits(208)	1,32E-11	32/61(52%)	40/61(66%)	2/61(3%)
gb AET05730.1	lipid transfer protein isoform a, partial [Pyrus communis]	65855 bits(208)	3,43E-11	30/58(52%)	38/58(66%)	0/58(0%)
gb AET05732.1	lipid transfer protein isoform c, partial [Pyrus communis]	654698 bits(208)	4,50E-11	30/58(52%)	38/58(66%)	0/58(0%)
gb ACQ42221.1	lipid transfer protein [Chimonanthus praecox]	65855 bits(208)	4,57E-11	30/58(52%)	40/58(69%)	0/58(0%)
gb ABK96515.1	unknown [Populus trichocarpa x Populus deltoides]	646994 bits(208)	1,36E-10	30/58(52%)	39/58(67%)	0/58(0%)
emb CAC86258.1	lipid transfer protein [Fragaria x ananassa]	646994 bits(208)	1,39E-10	30/58(52%)	38/58(66%)	0/58(0%)
gb AA438342.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	63929 bits(208)	2,28E-10	30/58(52%)	37/58(64%)	0/58(0%)

gb AFM35695.1	nonspecific lipid transfer protein [Vitis pseudoreticulata]	635438 bits(208)	1,10E-11	30/59(51%)	38/59(64%)	0/59(0%)
gb ABG54494.1	putative allergen Rub i 3 [Rubus idaeus]	646994 bits(208)	1,29E-10	30/59(51%)	37/59(63%)	0/59(0%)
emb CBW38499.1	lipid transfer protein [Helianthus annuus]	627734 bits(208)	1,71E-10	27/53(51%)	36/53(68%)	0/53(0%)
emb CBW38504.1	lipid transfer protein [Helianthus annuus]	627734 bits(208)	1,72E-10	27/53(51%)	36/53(68%)	0/53(0%)
emb CBW38503.1	lipid transfer protein [Helianthus annuus]	627734 bits(208)	1,80E-10	27/53(51%)	36/53(68%)	0/53(0%)
gb AEI70829.1	non-specific lipid-transfer protein [Helianthus annuus]	631586 bits(208)	1,83E-10	27/53(51%)	36/53(68%)	0/53(0%)
gb AAO33357.1 AF465408_1	nonspecific lipid transfer protein 1 [Vitis cinerea var. helleri x Vitis vinifera]	635438 bits(208)	2,26E-10	30/59(51%)	39/59(66%)	0/59(0%)
emb CBW38497.1	lipid transfer protein [Helianthus annuus]	623882 bits(208)	2,53E-10	27/53(51%)	36/53(68%)	0/53(0%)
gb AEI70826.1	non-specific lipid-transfer protein [Helianthus annuus]	681662 bits(208)	8,73E-12	29/58(50%)	41/58(71%)	0/58(0%)
gb AEI70828.1	non-specific lipid-transfer protein [Helianthus annuus]	670106 bits(208)	1,65E-11	29/58(50%)	40/58(69%)	0/58(0%)
gb AAV83345.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	670106 bits(208)	1,76E-11	29/58(50%)	40/58(69%)	0/58(0%)
gb ACP43542.1	lipid transfer protein [Chimonanthus praecox]	646994 bits(208)	1,85E-11	29/58(50%)	38/58(66%)	0/58(0%)
gb AAT68262.1	lipid transfer protein [Nicotiana glauca]	650846 bits(208)	2,12E-11	29/58(50%)	36/58(62%)	0/58(0%)
gb AAV83344.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	65855 bits(208)	5,20E-11	29/58(50%)	38/58(66%)	0/58(0%)
gb AET05731.1	lipid transfer protein isoform b, partial [Pyrus communis]	635438 bits(208)	2,16E-10	29/58(50%)	37/58(64%)	0/58(0%)
sp Q9M5X6.1 NLTP_PYRCO	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Allergen=Pyr c 3; Flags: Precursor	63929 bits(208)	2,25E-10	29/58(50%)	37/58(64%)	0/58(0%)
gb AAO33394.1 AF467946_1	lipid transfer protein isoform 4 [Vitis vinifera]	62003 bits(208)	6,00E-11	28/57(49%)	38/57(67%)	0/57(0%)
emb CBW38502.1	lipid transfer protein [Helianthus annuus]	63929 bits(208)	7,61E-11	26/53(49%)	37/53(70%)	0/53(0%)
gb AGH13363.1	lipid transfer protein, partial [Helianthus annuus]	627734 bits(208)	2,05E-10	26/53(49%)	37/53(70%)	0/53(0%)
emb CBW38501.1	lipid transfer protein [Helianthus annuus]	627734 bits(208)	2,35E-10	26/53(49%)	37/53(70%)	0/53(0%)
ref XP_002271116.1	PREDICTED: non-specific lipid-transfer protein [Vitis vinifera]	608474 bits(208)	2,64E-10	29/59(49%)	37/59(63%)	0/59(0%)
gb AAS13435.1	lipid-transfer protein [Nicotiana attenuata]	635438 bits(208)	2,11E-11	28/58(48%)	37/58(64%)	0/58(0%)
gb AAT68263.1	lipid transfer protein [Nicotiana glauca]	635438 bits(208)	2,32E-11	28/58(48%)	37/58(64%)	0/58(0%)
gb AAV83343.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	65855 bits(208)	4,96E-11	28/58(48%)	39/58(67%)	0/58(0%)

sp P23802.1 NLTP_ELECO	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Full=Alpha-amylase inhibitor I-2	577658 bits(208)	1,43E-10	28/58(48%)	37/58(64%)	1/58(2%)
gb AAQ96338.1	lipid transfer protein [Vitis aestivalis]	63929 bits(208)	2,61E-10	28/58(48%)	40/58(69%)	0/58(0%)
emb CBW38498.1	lipid transfer protein [Helianthus annuus]	627734 bits(208)	2,75E-10	26/54(48%)	37/54(69%)	0/54(0%)
gb AAT68265.1	lipid transfer protein precursor [Nicotiana glauca]	635438 bits(208)	3,02E-10	28/58(48%)	37/58(64%)	0/58(0%)
gb AAT68264.1	lipid transfer protein [Nicotiana glauca]	62003 bits(208)	9,34E-11	27/58(47%)	36/58(62%)	0/58(0%)
gb AAT45202.1	lipid transfer protein 1 precursor [Nicotiana tabacum]	623882 bits(208)	1,38E-10	28/59(47%)	36/59(61%)	0/59(0%)
gb AAF28385.1 AF151214_1	lipid-transfer protein [Nicotiana glauca]	623882 bits(208)	1,49E-10	27/58(47%)	35/58(60%)	0/58(0%)
dbj BAK19152.1	lipid transfer protein [Nicotiana tabacum]	623882 bits(208)	1,59E-10	28/59(47%)	36/59(61%)	0/59(0%)
Contig3						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_002334686.1	predicted protein [Populus trichocarpa]	808777 bits(194)	8,60E-23	40/71(56%)	48/71(68%)	0/71(0%)
ref XP_002303966.1	predicted protein [Populus trichocarpa]	79337 bits(194)	5,89E-23	39/71(55%)	48/71(68%)	0/71(0%)
emb CAN62733.1	hypothetical protein VITISV_015318 [Vitis vinifera]	835741 bits(194)	8,04E-22	41/78(53%)	56/78(72%)	4/78(5%)
ref XP_002297700.1	predicted protein [Populus trichocarpa]	724034 bits(194)	1,71E-19	41/77(53%)	52/77(68%)	0/77(0%)
emb CAN77893.1	hypothetical protein VITISV_030448 [Vitis vinifera]	812629 bits(194)	4,46E-21	40/78(51%)	55/78(71%)	4/78(5%)
ref XP_002265863.1	PREDICTED: flavonol sulfotransferase-like [Vitis vinifera]	700922 bits(194)	2,16E-19	36/71(51%)	48/71(68%)	3/71(4%)
ref XP_002334685.1	predicted protein [Populus trichocarpa]	654698 bits(194)	1,84E-18	39/77(51%)	53/77(69%)	1/77(1%)
ref XP_002263832.2	PREDICTED: LOW QUALITY PROTEIN: flavonol sulfotransferase-like [Vitis vinifera]	812629 bits(194)	1,17E-21	38/77(49%)	54/77(70%)	1/77(1%)
emb CAN74617.1	hypothetical protein VITISV_002219 [Vitis vinifera]	808777 bits(194)	6,39E-21	38/77(49%)	54/77(70%)	1/77(1%)
ref XP_002264574.1	PREDICTED: flavonol sulfotransferase-like [Vitis vinifera]	804925 bits(194)	7,87E-21	38/77(49%)	54/77(70%)	1/77(1%)
emb CBI35080.3	unnamed protein product [Vitis vinifera]	804925 bits(194)	8,06E-21	38/77(49%)	54/77(70%)	1/77(1%)
ref XP_002337045.1	predicted protein [Populus trichocarpa]	777962 bits(194)	1,26E-20	39/79(49%)	52/79(66%)	4/79(5%)
ref XP_002303947.1	predicted protein [Populus trichocarpa]	777962 bits(194)	2,39E-20	39/79(49%)	52/79(66%)	4/79(5%)

dbj BAB11158.1	steroid sulfotransferase-like protein [Arabidopsis thaliana]	758702 bits(194)	4,35E-20	39/80(49%)	56/80(70%)	4/80(5%)
ref NP_196317.2	sulfotransferase 2B [Arabidopsis thaliana]	758702 bits(194)	4,38E-20	39/80(49%)	56/80(70%)	4/80(5%)
ref XP_004308399.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	75485 bits(194)	1,41E-18	35/71(49%)	50/71(70%)	1/71(1%)
ref XP_003628737.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	654698 bits(194)	7,67E-17	35/71(49%)	42/71(59%)	1/71(1%)
ref XP_002265783.1	PREDICTED: flavonol 4'-sulfotransferase [Vitis vinifera]	770258 bits(194)	8,53E-20	37/77(48%)	50/77(65%)	1/77(1%)
ref XP_003542828.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	743294 bits(194)	3,43E-19	38/79(48%)	48/79(61%)	3/79(4%)
ref XP_002538120.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	673958 bits(194)	1,14E-17	37/77(48%)	53/77(69%)	1/77(1%)
ref XP_002334144.1	predicted protein [Populus trichocarpa]	635438 bits(194)	1,25E-17	37/77(48%)	51/77(66%)	1/77(1%)
ref XP_003542799.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	693218 bits(194)	2,46E-15	39/82(48%)	53/82(65%)	6/82(7%)
ref XP_002525538.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	758702 bits(194)	6,46E-22	36/77(47%)	52/77(68%)	1/77(1%)
ref XP_002264151.2	PREDICTED: sulfotransferase 16-like [Vitis vinifera]	747146 bits(194)	4,68E-20	36/77(47%)	51/77(66%)	1/77(1%)
ref XP_002873289.1	sulfotransferase family protein [Arabidopsis lyrata subsp. lyrata]	727886 bits(194)	1,47E-19	40/86(47%)	58/86(67%)	5/86(6%)
ref XP_003541597.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	724034 bits(194)	1,10E-18	37/79(47%)	48/79(61%)	3/79(4%)
ref XP_003599784.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	62003 bits(194)	5,12E-16	34/72(47%)	47/72(65%)	3/72(4%)
gb AFK33848.1	unknown [Lotus japonicus]	700922 bits(194)	4,22E-19	37/80(46%)	49/80(61%)	4/80(5%)
ref XP_004291155.1	PREDICTED: cytosolic sulfotransferase 15-like isoform 1 [Fragaria vesca subsp. vesca]	731738 bits(194)	4,55E-19	37/80(46%)	51/80(64%)	4/80(5%)
ref XP_004291156.1	PREDICTED: cytosolic sulfotransferase 15-like isoform 2 [Fragaria vesca subsp. vesca]	73559 bits(194)	5,10E-19	37/80(46%)	51/80(64%)	4/80(5%)
ref XP_003539305.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	689366 bits(194)	3,64E-18	37/80(46%)	51/80(64%)	4/80(5%)
gb ABK94985.1	unknown [Populus trichocarpa]	731738 bits(194)	9,56E-18	37/80(46%)	50/80(62%)	4/80(5%)
ref XP_002303948.1	predicted protein [Populus trichocarpa]	731738 bits(194)	9,64E-18	37/80(46%)	50/80(62%)	4/80(5%)
gb EMJ19923.1	hypothetical protein PRUPE_ppa009048mg [Prunus persica]	704774 bits(194)	1,38E-17	38/82(46%)	52/82(63%)	4/82(5%)
ref XP_002524373.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	662402 bits(194)	2,37E-17	38/83(46%)	52/83(63%)	6/83(7%)
ref XP_003545813.1	PREDICTED: LOW QUALITY PROTEIN: flavonol sulfotransferase-like [Glycine max]	69707 bits(194)	1,90E-15	38/82(46%)	52/82(63%)	6/82(7%)

ref XP_002512030.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	700922 bits(194)	2,52E-19	35/77(45%)	51/77(66%)	3/77(4%)
ref NP_568177.1	sulfotransferase 2A [Arabidopsis thaliana]	731738 bits(194)	4,95E-19	36/80(45%)	56/80(70%)	4/80(5%)
dbj BAB11159.1	steroid sulfotransferase-like protein [Arabidopsis thaliana]	731738 bits(194)	5,33E-19	36/80(45%)	56/80(70%)	4/80(5%)
gb AAL06879.1	AT5g07010/MOJ9_18 [Arabidopsis thaliana]	731738 bits(194)	1,75E-18	36/80(45%)	56/80(70%)	4/80(5%)
ref XP_004301099.1	PREDICTED: flavonol 3-sulfotransferase-like [Fragaria vesca subsp. vesca]	700922 bits(194)	3,59E-18	34/76(45%)	47/76(62%)	2/76(3%)
ref XP_002533884.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	643142 bits(194)	3,66E-17	37/83(45%)	48/83(58%)	6/83(7%)
ref XP_003572830.1	PREDICTED: sulfotransferase 16-like [Brachypodium distachyon]	670106 bits(194)	1,77E-16	34/76(45%)	45/76(59%)	2/76(3%)
ref XP_003627950.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	693218 bits(194)	1,32E-19	35/80(44%)	52/80(65%)	4/80(5%)
ref XP_004239455.1	PREDICTED: cytosolic sulfotransferase 5-like [Solanum lycopersicum]	693218 bits(194)	5,75E-19	36/81(44%)	52/81(64%)	4/81(5%)
ref XP_004306829.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	704774 bits(194)	6,44E-19	35/79(44%)	48/79(61%)	4/79(5%)
ref XP_002514612.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	681662 bits(194)	3,89E-18	35/79(44%)	50/79(63%)	2/79(3%)
ref XP_004292840.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	646994 bits(194)	9,11E-18	35/79(44%)	48/79(61%)	4/79(5%)
gb ADX97329.1	sulfotransferase-like protein [Mangifera indica]	724034 bits(194)	3,69E-17	36/81(44%)	53/81(65%)	4/81(5%)
ref XP_002304810.1	predicted protein [Populus trichocarpa]	593066 bits(194)	1,15E-16	35/80(44%)	53/80(66%)	4/80(5%)
ref XP_003628744.1	Flavonol 3-sulfotransferase [Medicago truncatula]	62003 bits(194)	5,31E-16	31/71(44%)	44/71(62%)	1/71(1%)
ref XP_002335751.1	predicted protein [Populus trichocarpa]	881965 bits(194)	4,71E-18	45/104(43%)	63/104(61%)	1/104(1%)
ref XP_004239020.1	PREDICTED: cytosolic sulfotransferase 12-like [Solanum lycopersicum]	627734 bits(194)	5,13E-17	35/81(43%)	48/81(59%)	4/81(5%)
ref XP_004240283.1	PREDICTED: cytosolic sulfotransferase 5-like [Solanum lycopersicum]	63929 bits(194)	2,12E-16	34/79(43%)	48/79(61%)	4/79(5%)
ref XP_002318367.1	predicted protein [Populus trichocarpa]	673958 bits(194)	8,76E-18	33/79(42%)	50/79(63%)	3/79(4%)
gb ABK95273.1	unknown [Populus trichocarpa]	673958 bits(194)	8,76E-18	33/79(42%)	50/79(63%)	3/79(4%)
ref XP_002314916.1	predicted protein [Populus trichocarpa]	623882 bits(194)	9,18E-17	30/71(42%)	45/71(63%)	1/71(1%)
ref XP_004239453.1	PREDICTED: cytosolic sulfotransferase 12-like [Solanum lycopersicum]	654698 bits(194)	2,78E-16	34/80(42%)	49/80(61%)	4/80(5%)
ref XP_004247253.1	PREDICTED: cytosolic sulfotransferase 5-like [Solanum lycopersicum]	673958 bits(194)	2,92E-16	34/81(42%)	51/81(63%)	4/81(5%)
ref XP_003580959.1	PREDICTED: sulfotransferase 17-like [Brachypodium distachyon]	689366 bits(194)	6,25E-16	32/77(42%)	43/77(56%)	2/77(3%)

ref XP_002524377.1	Flavonol 3-sulfotransferase, putative [<i>Ricinus communis</i>]	623882 bits(194)	1,47E-15	35/83(42%)	48/83(58%)	6/83(7%)
dbj BAJ95002.1	predicted protein [<i>Hordeum vulgare</i> subsp. <i>vulgare</i>]	646994 bits(194)	4,99E-15	31/73(42%)	44/73(60%)	3/73(4%)
ref XP_004240284.1	PREDICTED: cytosolic sulfotransferase 5-like [<i>Solanum lycopersicum</i>]	650846 bits(194)	5,77E-15	34/80(42%)	50/80(62%)	4/80(5%)
ref XP_002276423.2	PREDICTED: sulfotransferase 17-like [<i>Vitis vinifera</i>]	643142 bits(194)	4,06E-19	32/79(41%)	47/79(59%)	2/79(3%)
ref XP_003627666.1	Steroid sulfotransferase-like protein [<i>Medicago truncatula</i>]	67781 bits(194)	4,48E-19	33/80(41%)	53/80(66%)	4/80(5%)
ref XP_002529671.1	Flavonol 3-sulfotransferase, putative [<i>Ricinus communis</i>]	708626 bits(194)	1,33E-18	33/80(41%)	52/80(65%)	3/80(4%)
ref XP_004250849.1	PREDICTED: cytosolic sulfotransferase 12-like [<i>Solanum lycopersicum</i>]	608474 bits(194)	2,95E-17	32/79(41%)	47/79(59%)	4/79(5%)
gb EAZ01746.1	hypothetical protein Osl_23778 [<i>Oryza sativa</i> Indica Group]	63929 bits(194)	5,02E-17	32/79(41%)	42/79(53%)	2/79(3%)
emb CBI15346.3	unnamed protein product [<i>Vitis vinifera</i>]	67781 bits(194)	1,33E-16	33/80(41%)	51/80(64%)	4/80(5%)
ref XP_002336441.1	predicted protein [<i>Populus trichocarpa</i>]	612326 bits(194)	1,50E-16	29/71(41%)	45/71(63%)	1/71(1%)
ref XP_002331091.1	predicted protein [<i>Populus trichocarpa</i>]	593066 bits(194)	2,17E-16	34/82(41%)	46/82(56%)	8/82(10%)
gb EMJ10097.1	hypothetical protein PRUPE_ppa026895mg, partial [<i>Prunus persica</i>]	631586 bits(194)	4,40E-16	33/80(41%)	47/80(59%)	6/80(8%)
ref XP_002329759.1	predicted protein [<i>Populus trichocarpa</i>]	654698 bits(194)	1,57E-15	32/79(41%)	46/79(58%)	3/79(4%)
ref XP_003627955.1	Flavonol sulfotransferase-like protein [<i>Medicago truncatula</i>]	627734 bits(194)	3,63E-15	33/80(41%)	49/80(61%)	4/80(5%)
ref XP_002452746.1	hypothetical protein SORBIDRAFT_04g031660 [<i>Sorghum bicolor</i>]	673958 bits(194)	3,70E-15	32/79(41%)	46/79(58%)	2/79(3%)
dbj BAK02574.1	predicted protein [<i>Hordeum vulgare</i> subsp. <i>vulgare</i>]	662402 bits(194)	3,93E-15	33/80(41%)	46/80(58%)	4/80(5%)
ref XP_002524376.1	Flavonol 3-sulfotransferase, putative [<i>Ricinus communis</i>]	569954 bits(194)	4,49E-15	36/87(41%)	49/87(56%)	14/87(16%)
ref XP_002268667.1	PREDICTED: sulfotransferase 16 [<i>Vitis vinifera</i>]	673958 bits(194)	7,83E-19	32/80(40%)	49/80(61%)	3/80(4%)
emb CAN75260.1	hypothetical protein VITISV_041143 [<i>Vitis vinifera</i>]	666254 bits(194)	1,34E-18	32/80(40%)	49/80(61%)	3/80(4%)
ref XP_002264929.1	PREDICTED: sulfotransferase 17-like [<i>Vitis vinifera</i>]	627734 bits(194)	3,37E-17	33/83(40%)	49/83(59%)	9/83(11%)
ref XP_003627983.1	Flavonol sulfotransferase-like protein [<i>Medicago truncatula</i>]	646994 bits(194)	8,46E-16	31/78(40%)	51/78(65%)	3/78(4%)
ref XP_002317222.1	predicted protein [<i>Populus trichocarpa</i>]	654698 bits(194)	9,05E-16	32/80(40%)	46/80(58%)	3/80(4%)
ref XP_002329757.1	predicted protein [<i>Populus trichocarpa</i>]	670106 bits(194)	9,60E-16	32/80(40%)	47/80(59%)	3/80(4%)
ref XP_002317223.1	predicted protein [<i>Populus trichocarpa</i>]	662402 bits(194)	1,01E-15	32/80(40%)	46/80(58%)	3/80(4%)
ref XP_002329758.1	predicted protein [<i>Populus trichocarpa</i>]	650846 bits(194)	2,18E-15	32/80(40%)	46/80(58%)	3/80(4%)

gb EMJ11972.1	hypothetical protein PRUPE_ppa019757mg, partial [Prunus persica]	631586 bits(194)	4,25E-15	35/87(40%)	48/87(55%)	10/87(11%)
ref XP_002276618.2	PREDICTED: sulfotransferase 17-like [Vitis vinifera]	627734 bits(194)	1,89E-19	31/79(39%)	47/79(59%)	2/79(3%)
emb CAN83405.1	hypothetical protein VITISV_022677 [Vitis vinifera]	623882 bits(194)	1,01E-17	31/79(39%)	47/79(59%)	2/79(3%)
ref XP_002529982.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	650846 bits(194)	1,35E-17	34/87(39%)	54/87(62%)	2/87(2%)
emb CBI15343.3	unnamed protein product [Vitis vinifera]	623882 bits(194)	3,37E-17	31/79(39%)	47/79(59%)	3/79(4%)
ref XP_002264512.1	PREDICTED: sulfotransferase 17-like [Vitis vinifera]	627734 bits(194)	4,04E-17	31/80(39%)	47/80(59%)	3/80(4%)
ref XP_002262899.2	PREDICTED: flavonol 4'-sulfotransferase-like [Vitis vinifera]	631586 bits(194)	2,55E-16	31/79(39%)	47/79(59%)	2/79(3%)
ref XP_002524372.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	593066 bits(194)	1,07E-15	34/88(39%)	49/88(56%)	16/88(18%)
sp P52838.1 FSTL_FLABI	RecName: Full=Flavonol sulfotransferase-like	558398 bits(194)	5,42E-15	31/79(39%)	44/79(56%)	3/79(4%)
ref XP_004298346.1	PREDICTED: cytosolic sulfotransferase 5-like [Fragaria vesca subsp. vesca]	662402 bits(194)	7,95E-17	29/76(38%)	47/76(62%)	1/76(1%)
ref XP_002272968.2	PREDICTED: sulfotransferase 16-like [Vitis vinifera]	596918 bits(194)	9,57E-16	30/79(38%)	47/79(59%)	2/79(3%)
ref XP_002267209.1	PREDICTED: flavonol sulfotransferase-like [Vitis vinifera]	604622 bits(194)	4,82E-15	31/81(38%)	46/81(57%)	4/81(5%)
gb AAM62638.1	steroid sulfotransferase, putative [Arabidopsis thaliana]	616178 bits(194)	5,63E-15	29/76(38%)	48/76(63%)	1/76(1%)
ref NP_172799.1	sulfotransferase 4B [Arabidopsis thaliana]	616178 bits(194)	5,67E-15	29/76(38%)	48/76(63%)	1/76(1%)
ref XP_002334799.1	predicted protein [Populus trichocarpa]	60077 bits(194)	9,07E-15	32/86(37%)	49/86(57%)	4/86(5%)
Contig4						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb ACU14699.1	unknown [Glycine max]	101293 bits(251)	8,19E-24	49/55(89%)	53/55(96%)	0/55(0%)
ref XP_003554596.1	PREDICTED: hydrophobic protein LTI6A-like [Glycine max]	100908 bits(251)	1,31E-23	49/55(89%)	53/55(96%)	0/55(0%)
ref XP_002512586.1	Hydrophobic protein LTI6A, putative [Ricinus communis]	100523 bits(251)	1,57E-23	49/56(88%)	53/56(95%)	0/56(0%)
ref XP_003555775.1	PREDICTED: uncharacterized protein LOC100306281 [Glycine max]	985969 bits(251)	7,46E-23	48/55(87%)	51/55(93%)	0/55(0%)

ref XP_003536025.1	PREDICTED: hydrophobic protein LTI6A-like [Glycine max]	978265 bits(251)	1,81E-22	47/55(85%)	51/55(93%)	0/55(0%)
gb ACU14681.1	unknown [Glycine max]	970561 bits(251)	3,32E-22	47/55(85%)	50/55(91%)	0/55(0%)
ref NP_187240.1	Hydrophobic protein RC12B [Arabidopsis thaliana]	920485 bits(251)	1,93E-20	45/53(85%)	51/53(96%)	0/53(0%)
ref XP_002884539.1	low temperature and salt responsive protein LTI6B [Arabidopsis lyrata subsp. lyrata]	920485 bits(251)	2,72E-20	45/53(85%)	51/53(96%)	0/53(0%)
gb AFC41201.1	PM-YC3.6-Lti6b [Binary expression vector PM-YC3.6-LTI6b]	951301 bits(251)	1,33E-18	45/53(85%)	51/53(96%)	0/53(0%)
ref NP_001232820.1	hydrophobic protein LTI6A [Zea mays]	974413 bits(251)	2,34E-22	47/56(84%)	51/56(91%)	0/56(0%)
ref NP_001147403.1	hydrophobic protein LTI6A [Zea mays]	966709 bits(251)	3,62E-22	47/56(84%)	51/56(91%)	0/56(0%)
ref XP_003626132.1	Hydrophobic protein LTI6A [Medicago truncatula]	908929 bits(251)	4,47E-20	44/53(83%)	49/53(92%)	0/53(0%)
gb EAY73580.1	hypothetical protein Osl_01464 [Oryza sativa Indica Group]	974413 bits(251)	1,90E-22	46/56(82%)	52/56(93%)	0/56(0%)
ref NP_001152633.1	hydrophobic protein LTI6A [Zea mays]	943597 bits(251)	2,60E-21	46/56(82%)	50/56(89%)	0/56(0%)
gb AAF26091.1 AC012393_17	low temperature and salt responsive protein LTI6B [Arabidopsis thaliana]	935893 bits(251)	6,31E-21	46/56(82%)	53/56(95%)	0/56(0%)
ref XP_002329990.1	stress-induced hydrophobic peptide [Populus trichocarpa]	932041 bits(251)	6,62E-21	45/55(82%)	50/55(91%)	0/55(0%)
tpg DAA41563.1	TPA: hypothetical protein ZEAMMB73_862570 [Zea mays]	920485 bits(251)	2,13E-20	45/55(82%)	48/55(87%)	0/55(0%)
tpg DAA41564.1	TPA: hypothetical protein ZEAMMB73_862570 [Zea mays]	920485 bits(251)	2,16E-20	45/55(82%)	48/55(87%)	0/55(0%)
gb ABQ41924.1	low temperature and salt responsive protein [Sonneratia alba]	766406 bits(251)	6,05E-15	37/45(82%)	42/45(93%)	0/45(0%)
gb ACA66247.1	cold-induced plasma membrane protein [Musa ABB Group]	928189 bits(251)	8,73E-21	46/57(81%)	53/57(93%)	1/57(2%)
gb AFI47457.1	low temperature and salt responsive protein [Medicago sativa]	889669 bits(251)	2,54E-19	43/53(81%)	48/53(91%)	0/53(0%)
ref XP_002329989.1	stress-induced hydrophobic peptide [Populus trichocarpa]	824185 bits(251)	4,56E-17	39/48(81%)	46/48(96%)	0/48(0%)
gb AAV88601.1	low temperature and salt responsive protein [Cenchrus americanus]	951301 bits(251)	1,56E-21	45/56(80%)	50/56(89%)	0/56(0%)
ref XP_004230367.1	PREDICTED: hydrophobic protein LTI6A-like [Solanum lycopersicum]	939745 bits(251)	3,61E-21	44/56(79%)	52/56(93%)	0/56(0%)
ref NP_001060390.1	Os07g0635900 [Oryza sativa Japonica Group]	939745 bits(251)	3,72E-21	44/56(79%)	51/56(91%)	0/56(0%)
ref NP_001107634.1	LOC100135422 [Zea mays]	920485 bits(251)	1,89E-20	44/56(79%)	49/56(88%)	0/56(0%)
gb ACN26849.1	unknown [Zea mays]	920485 bits(251)	2,15E-20	44/56(79%)	49/56(88%)	0/56(0%)
ref XP_002440557.1	hypothetical protein SORBIDRAFT_09g003060 [Sorghum bicolor]	908929 bits(251)	5,26E-20	44/56(79%)	49/56(88%)	0/56(0%)

gb ABD33207.2	Protein of unknown function UPF0057 [Medicago truncatula]	905077 bits(251)	6,31E-20	42/53(79%)	50/53(94%)	0/53(0%)
ref XP_003626135.1	Hydrophobic protein LTI6B [Medicago truncatula]	912781 bits(251)	6,82E-20	42/53(79%)	50/53(94%)	0/53(0%)
gb AFK38849.1	unknown [Lotus japonicus]	889669 bits(251)	2,52E-19	42/53(79%)	50/53(94%)	0/53(0%)
ref XP_003626131.1	Hydrophobic protein RCI2B [Medicago truncatula]	878113 bits(251)	6,44E-19	42/53(79%)	49/53(92%)	0/53(0%)
gb ADE77881.1	unknown [Picea sitchensis]	855001 bits(251)	4,27E-18	41/52(79%)	46/52(88%)	0/52(0%)
ref XP_002440556.1	hypothetical protein SORBIDRAFT_09g003050 [Sorghum bicolor]	839593 bits(251)	1,51E-17	41/52(79%)	46/52(88%)	0/52(0%)
gb EMJ08608.1	hypothetical protein PRUPE_ppa014548mg [Prunus persica]	935893 bits(251)	6,11E-21	43/55(78%)	51/55(93%)	0/55(0%)
gb ACV50425.1	cold induced plasma membrane protein [Jatropha curcas]	924337 bits(251)	1,47E-20	43/55(78%)	51/55(93%)	0/55(0%)
ref XP_003567502.1	PREDICTED: hydrophobic protein LTI6A-like [Brachypodium distachyon]	912781 bits(251)	4,06E-20	43/56(77%)	49/56(88%)	0/56(0%)
gb ADK27677.1	plasma membrane protein 3-2 [Salvia miltiorrhiza]	866557 bits(251)	1,38E-18	41/53(77%)	48/53(91%)	0/53(0%)
gb AAQ84111.1	Clf1 [Citrus trifoliata]	866557 bits(251)	1,48E-18	41/53(77%)	48/53(91%)	0/53(0%)
ref XP_004302326.1	PREDICTED: hydrophobic protein RCI2A-like [Fragaria vesca subsp. vesca]	843445 bits(251)	1,13E-17	41/53(77%)	48/53(91%)	0/53(0%)
gb ADC45381.1	stress-induced hydrophobic peptide [Cleistogenes songorica]	874261 bits(251)	8,53E-19	42/55(76%)	48/55(87%)	0/55(0%)
ref XP_003557909.1	PREDICTED: hydrophobic protein LTI6A-like [Brachypodium distachyon]	943597 bits(251)	2,79E-21	42/56(75%)	52/56(93%)	0/56(0%)
ref XP_004230366.1	PREDICTED: hydrophobic protein LTI6A-like [Solanum lycopersicum]	920485 bits(251)	1,69E-20	42/56(75%)	51/56(91%)	0/56(0%)
gb ACH60980.1	rare cold inducible protein [Pseudotsuga menziesii]	916633 bits(251)	2,89E-20	42/56(75%)	49/56(88%)	0/56(0%)
gb ACH60977.1	rare cold inducible protein [Pseudotsuga menziesii]	912781 bits(251)	4,56E-20	42/56(75%)	49/56(88%)	0/56(0%)
ref NP_001151840.1	hydrophobic protein LTI6B [Zea mays]	905077 bits(251)	7,51E-20	42/56(75%)	49/56(88%)	0/56(0%)
ref NP_187239.1	Hydrophobic protein RCI2A [Arabidopsis thaliana]	866557 bits(251)	1,35E-18	40/53(75%)	49/53(92%)	0/53(0%)
ref XP_002993149.1	hypothetical protein SELMODRAFT_136629 [Selaginella moellendorffii]	847297 bits(251)	7,07E-18	41/55(75%)	50/55(91%)	1/55(2%)
ref XP_002318921.1	stress-induced hydrophobic peptide [Populus trichocarpa]	839593 bits(251)	1,19E-17	40/53(75%)	46/53(87%)	0/53(0%)
gb EAY96485.1	hypothetical protein Osl_18385 [Oryza sativa Indica Group]	851149 bits(251)	1,19E-17	40/53(75%)	46/53(87%)	0/53(0%)
dbj BAD34658.1	plasma membrane protein 3 [Leymus chinensis]	839593 bits(251)	1,36E-17	40/53(75%)	46/53(87%)	0/53(0%)
ref NP_001054591.1	Os05g0138300 [Oryza sativa Japonica Group]	839593 bits(251)	1,50E-17	40/53(75%)	46/53(87%)	0/53(0%)

ref XP_003568974.1	PREDICTED: hydrophobic protein LTI6B-like [Brachypodium distachyon]	839593 bits(251)	1,50E-17	40/53(75%)	46/53(87%)	0/53(0%)
gb ABK96638.1	unknown [Populus trichocarpa x Populus deltoides]	820333 bits(251)	7,38E-17	40/53(75%)	45/53(85%)	0/53(0%)
gb ACJ09630.1	putative early drought-induced protein [Cupressus sempervirens]	828037 bits(251)	9,85E-17	42/56(75%)	47/56(84%)	1/56(2%)
ref XP_001776246.1	predicted protein [Physcomitrella patens subsp. patens]	804925 bits(251)	3,02E-16	40/53(75%)	44/53(83%)	0/53(0%)
gb ADK27676.1	plasma membrane protein 3-1 [Salvia miltiorrhiza]	73559 bits(251)	9,21E-14	40/53(75%)	49/53(92%)	0/53(0%)
gb ABK25741.1	unknown [Picea sitchensis]	720182 bits(251)	4,27E-13	42/56(75%)	49/56(88%)	0/56(0%)
gb AFK81273.1	rare cold-inducible protein 2A [Camelina sativa]	862705 bits(251)	1,88E-18	39/53(74%)	49/53(92%)	0/53(0%)
gb AEJ20974.1	cold-inducible protein [Caragana jubata]	851149 bits(251)	4,96E-18	39/53(74%)	48/53(91%)	0/53(0%)
gb ACJ09619.1	putative low temperature and salt responsive protein [Cupressus sempervirens]	73559 bits(251)	1,31E-13	35/47(74%)	41/47(87%)	0/47(0%)
dbj BAD34659.1	plasma membrane protein 3 [Leymus chinensis]	708626 bits(251)	9,81E-13	39/53(74%)	47/53(89%)	0/53(0%)
gb ACH61000.1	rare cold inducible protein [Pseudotsuga macrocarpa]	905077 bits(251)	7,39E-20	41/56(73%)	49/56(88%)	0/56(0%)
gb ABK22915.1	unknown [Picea sitchensis]	905077 bits(251)	8,01E-20	41/56(73%)	49/56(88%)	0/56(0%)
ref XP_004302327.1	PREDICTED: hydrophobic protein RC12A-like [Fragaria vesca subsp. vesca]	901225 bits(251)	8,62E-20	40/55(73%)	50/55(91%)	0/55(0%)
ref XP_002884538.1	predicted protein [Arabidopsis lyrata subsp. lyrata]	828037 bits(251)	3,48E-17	38/53(72%)	48/53(91%)	0/53(0%)
ref XP_001751863.1	predicted protein [Physcomitrella patens subsp. patens]	739442 bits(251)	8,12E-14	36/50(72%)	41/50(82%)	0/50(0%)
gb AFP99877.1	plasma membrane protein Am244 [Avicennia marina]	881965 bits(251)	4,06E-19	40/56(71%)	52/56(93%)	0/56(0%)
ref NP_001147508.1	LOC100281117 [Zea mays]	878113 bits(251)	6,04E-19	40/56(71%)	48/56(86%)	0/56(0%)
dbj BAG54793.1	plasma membrane protein3 [Puccinellia tenuiflora]	843445 bits(251)	1,08E-17	39/55(71%)	46/55(84%)	0/55(0%)
gb EAY90184.1	hypothetical protein Osl_11748 [Oryza sativa Indica Group]	862705 bits(251)	2,21E-18	39/56(70%)	49/56(88%)	0/56(0%)
ref NP_001050200.1	Os03g0370600 [Oryza sativa Japonica Group]	851149 bits(251)	6,30E-18	39/56(70%)	47/56(84%)	0/56(0%)
ref XP_003562607.1	PREDICTED: hydrophobic protein LTI6A-like [Brachypodium distachyon]	835741 bits(251)	2,36E-17	38/55(69%)	46/55(84%)	0/55(0%)
ref XP_002465219.1	hypothetical protein SORBIDRAFT_01g034350 [Sorghum bicolor]	801073 bits(251)	3,49E-16	38/55(69%)	46/55(84%)	0/55(0%)
ref XP_002964158.1	hypothetical protein SELMODRAFT_227480 [Selaginella moellendorffii]	789518 bits(251)	9,47E-16	36/52(69%)	46/52(88%)	0/52(0%)
ref XP_002316342.1	stress-induced hydrophobic peptide [Populus trichocarpa]	77411 bits(251)	3,58E-15	35/51(69%)	46/51(90%)	0/51(0%)

ref XP_004143609.1	PREDICTED: hydrophobic protein RCI2A-like [Cucumis sativus]	762554 bits(251)	8,56E-15	35/51(69%)	45/51(88%)	0/51(0%)
gb EMJ07382.1	hypothetical protein PRUPE_ppa013857mg [Prunus persica]	901225 bits(251)	2,15E-19	45/67(67%)	54/67(81%)	3/67(4%)
ref XP_002993150.1	hypothetical protein SELMODRAFT_229736 [Selaginella moellendorffii]	785666 bits(251)	1,48E-15	35/52(67%)	46/52(88%)	0/52(0%)
emb CBI39675.3	unnamed protein product [Vitis vinifera]	73559 bits(251)	3,30E-13	36/54(67%)	40/54(74%)	0/54(0%)
sp Q9ARD5.1 LT02_HORVU	RecName: Full=Low temperature-induced protein It101.2	770258 bits(251)	5,02E-15	35/53(66%)	46/53(87%)	0/53(0%)
gb ACG48459.1	hypothetical protein [Zea mays]	73559 bits(251)	8,84E-14	35/53(66%)	44/53(83%)	0/53(0%)
dbj BAK07288.1	predicted protein [Hordeum vulgare subsp. vulgare]	801073 bits(251)	1,10E-15	35/54(65%)	45/54(83%)	0/54(0%)
ref NP_565897.1	Low temperature and salt responsive protein [Arabidopsis thaliana]	739442 bits(251)	6,55E-14	33/51(65%)	45/51(88%)	0/51(0%)
ref XP_002311114.1	stress-induced hydrophobic peptide [Populus trichocarpa]	73559 bits(251)	7,97E-14	33/51(65%)	45/51(88%)	0/51(0%)
ref XP_004307420.1	PREDICTED: low temperature-induced protein It101.2-like [Fragaria vesca subsp. vesca]	700922 bits(251)	1,65E-12	33/51(65%)	43/51(84%)	0/51(0%)
gb AFW88281.1	hypothetical protein ZEAMMB73_701475 [Zea mays]	797221 bits(251)	4,96E-16	36/56(64%)	47/56(84%)	0/56(0%)
ref XP_001765492.1	predicted protein [Physcomitrella patens subsp. patens]	781814 bits(251)	2,06E-15	35/55(64%)	46/55(84%)	0/55(0%)
ref XP_003558174.1	PREDICTED: low temperature-induced protein It101.2-like [Brachypodium distachyon]	747146 bits(251)	3,38E-14	34/53(64%)	46/53(87%)	0/53(0%)
gb ACG26760.1	hypothetical protein [Zea mays]	731738 bits(251)	1,26E-13	34/53(64%)	45/53(85%)	0/53(0%)
ref XP_002440508.1	hypothetical protein SORBIDRAFT_09g002150 [Sorghum bicolor]	724034 bits(251)	2,71E-13	34/53(64%)	44/53(83%)	0/53(0%)
sp P68178.1 ESI3_LOPEL	RecName: Full=Salt stress-induced hydrophobic peptide ESI3	71633 bits(251)	5,49E-13	33/52(63%)	44/52(85%)	0/52(0%)
ref XP_004137807.1	PREDICTED: low temperature-induced protein It101.2-like [Cucumis sativus]	712478 bits(251)	9,31E-13	32/51(63%)	43/51(84%)	0/51(0%)
dbj BAG54794.1	plasma membrane protein3 [Puccinellia tenuiflora]	724034 bits(251)	2,39E-13	33/53(62%)	45/53(85%)	0/53(0%)
ref XP_002277985.1	PREDICTED: low temperature-induced protein It101.2 [Vitis vinifera]	720182 bits(251)	3,20E-13	33/53(62%)	45/53(85%)	0/53(0%)
ref XP_002465426.1	hypothetical protein SORBIDRAFT_01g038670 [Sorghum bicolor]	720182 bits(251)	3,58E-13	33/53(62%)	45/53(85%)	0/53(0%)
gb AER00467.1	BLT101 [Dasypyrum villosum]	712478 bits(251)	6,09E-13	33/53(62%)	44/53(83%)	0/53(0%)
ref XP_004237520.1	PREDICTED: low temperature-induced protein It101.2-like [Solanum lycopersicum]	71633 bits(251)	4,74E-13	31/51(61%)	45/51(88%)	0/51(0%)
ref XP_004306234.1	PREDICTED: hydrophobic protein LT16A-like [Fragaria vesca subsp. vesca]	704774 bits(251)	1,55E-12	33/54(61%)	43/54(80%)	0/54(0%)

ref XP_004230364.1	PREDICTED: hydrophobic protein LTI6A-like [Solanum lycopersicum]	708626 bits(251)	1,55E-12	33/56(59%)	45/56(80%)	0/56(0%)
Contig5						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb AEN70918.1	lipid transfer protein [Gossypium turneri]	158688 bits(402)	4,08E-45	80/102(78%)	90/102(88%)	2/102(2%)
gb AAF35186.1 AF195865_1	lipid transfer protein precursor [Gossypium hirsutum]	157918 bits(402)	9,01E-45	81/105(77%)	94/105(90%)	2/105(2%)
gb AEN70933.1	lipid transfer protein [Gossypium davidsonii]	154066 bits(402)	2,46E-43	79/102(77%)	88/102(86%)	2/102(2%)
gb ABO42261.1	lipid transfer protein 4 precursor [Gossypium hirsutum]	143665 bits(402)	2,07E-39	81/105(77%)	90/105(86%)	2/105(2%)
gb AAC00499.1	lipid transfer protein precursor [Gossypium hirsutum]	14328 bits(402)	3,31E-39	81/105(77%)	90/105(86%)	2/105(2%)
gb ADY68821.1	lipid transfer protein 3 [Gossypium raimondii]	139813 bits(402)	8,04E-38	79/102(77%)	87/102(85%)	2/102(2%)
gb AEN70926.1	lipid transfer protein [Gossypium barbadense var. brasiliense]	137117 bits(402)	8,05E-37	79/102(77%)	86/102(84%)	2/102(2%)
gb AEN70916.1	lipid transfer protein [Gossypium laxum]	154836 bits(402)	1,03E-43	78/102(76%)	88/102(86%)	2/102(2%)
gb AEN70936.1	lipid transfer protein [Gossypium gossypoides]	154451 bits(402)	1,68E-43	77/101(76%)	88/101(87%)	1/101(1%)
gb AEN70915.1	lipid transfer protein [Gossypium thurberi]	153295 bits(402)	5,14E-43	78/102(76%)	87/102(85%)	2/102(2%)
gb AEN70919.1	lipid transfer protein [Gossypium mustelinum]	151754 bits(402)	1,65E-42	78/102(76%)	87/102(85%)	2/102(2%)
gb AAN77147.1	fiber lipid transfer protein [Gossypium barbadense]	14251 bits(402)	7,11E-39	80/105(76%)	89/105(85%)	2/105(2%)
gb ABA29446.1	non-specific lipid transfer protein [Vitis vinifera]	141739 bits(402)	1,43E-38	68/90(76%)	79/90(88%)	0/90(0%)
gb AAR90329.1	lipid transfer protein precursor [Gossypium barbadense]	140198 bits(402)	5,97E-38	80/105(76%)	89/105(85%)	2/105(2%)
gb ADY68820.1	lipid transfer protein 3 [Gossypium herbaceum subsp. africanum]	155221 bits(402)	1,01E-43	76/102(75%)	90/102(88%)	2/102(2%)
gb AAY43801.1	FSLTP3 [Gossypium hirsutum]	154836 bits(402)	1,27E-43	76/102(75%)	90/102(88%)	2/102(2%)
gb AEN70917.1	lipid transfer protein [Gossypium schwendimanii]	150984 bits(402)	3,59E-42	77/102(75%)	86/102(84%)	2/102(2%)
gb AEN70935.1	lipid transfer protein [Gossypium aridum]	150214 bits(402)	7,43E-42	76/102(75%)	86/102(84%)	2/102(2%)
gb AAG29777.1	lipid transfer protein 3 precursor [Gossypium hirsutum]	159458 bits(402)	2,24E-45	78/105(74%)	93/105(89%)	2/105(2%)
gb AAF35184.1 AF195863_1	lipid transfer protein precursor [Gossypium hirsutum]	152525 bits(402)	1,04E-42	78/105(74%)	88/105(84%)	2/105(2%)

gb AEI70830.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	14405 bits(402)	1,77E-39	76/103(74%)	85/103(83%)	1/103(1%)
gb AGH13362.1	lipid transfer protein, partial [<i>Helianthus annuus</i>]	137887 bits(402)	2,02E-37	73/98(74%)	82/98(84%)	1/98(1%)
gb AAAY43800.1	FSLTP2 [<i>Gossypium hirsutum</i>]	151369 bits(402)	2,42E-42	74/101(73%)	89/101(88%)	2/101(2%)
emb CBW38501.1	lipid transfer protein [<i>Helianthus annuus</i>]	140969 bits(402)	1,77E-38	75/103(73%)	85/103(83%)	1/103(1%)
gb AAO33394.1 AF467946_1	lipid transfer protein isoform 4 [<i>Vitis vinifera</i>]	135191 bits(402)	4,80E-36	65/89(73%)	78/89(88%)	0/89(0%)
sp Q39950.1 NLTP_HELAN	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; Short=NslTP; AltName: Full=SDI-9; Flags: Precursor	143665 bits(402)	1,86E-39	74/103(72%)	85/103(83%)	1/103(1%)
emb CBW38498.1	lipid transfer protein [<i>Helianthus annuus</i>]	142124 bits(402)	5,77E-39	74/103(72%)	85/103(83%)	1/103(1%)
gb AEI70828.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	141354 bits(402)	1,80E-38	74/103(72%)	85/103(83%)	1/103(1%)
gb AGH13363.1	lipid transfer protein, partial [<i>Helianthus annuus</i>]	139813 bits(402)	4,90E-38	74/103(72%)	85/103(83%)	1/103(1%)
emb CBW38499.1	lipid transfer protein [<i>Helianthus annuus</i>]	139043 bits(402)	8,53E-38	74/103(72%)	83/103(81%)	1/103(1%)
gb AEN70924.1	lipid transfer protein [<i>Gossypium tomentosum</i>]	136732 bits(402)	1,02E-36	74/103(72%)	83/103(81%)	4/103(4%)
gb AAO33357.1 AF465408_1	nonspecific lipid transfer protein 1 [<i>Vitis cinerea</i> var. helleri x <i>Vitis vinifera</i>]	134035 bits(402)	1,34E-35	65/90(72%)	77/90(86%)	0/90(0%)
sp Q43129.1 NLTP2_GOSHI	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Full=GH3; Flags: Precursor	151754 bits(402)	2,02E-42	75/105(71%)	91/105(87%)	2/105(2%)
sp Q42762.1 NLTP1_GOSHI	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; Flags: Precursor	150984 bits(402)	3,23E-42	75/105(71%)	91/105(87%)	2/105(2%)
gb AAF35185.1 AF195864_1	lipid transfer protein precursor [<i>Gossypium hirsutum</i>]	150599 bits(402)	4,84E-42	75/105(71%)	89/105(85%)	2/105(2%)
gb AAL27855.1	lipid transfer protein precursor [<i>Davidia involucrata</i>]	145976 bits(402)	3,01E-40	73/103(71%)	85/103(83%)	0/103(0%)
gb AEI70827.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	139043 bits(402)	1,37E-37	73/103(71%)	82/103(80%)	1/103(1%)
emb CBW38497.1	lipid transfer protein [<i>Helianthus annuus</i>]	137887 bits(402)	2,52E-37	73/103(71%)	83/103(81%)	1/103(1%)
gb AEI70829.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	136732 bits(402)	8,61E-37	73/103(71%)	82/103(80%)	1/103(1%)
emb CBW38504.1	lipid transfer protein [<i>Helianthus annuus</i>]	136346 bits(402)	8,74E-37	73/103(71%)	82/103(80%)	1/103(1%)
emb CBW38503.1	lipid transfer protein [<i>Helianthus annuus</i>]	136346 bits(402)	1,04E-36	73/103(71%)	82/103(80%)	1/103(1%)
ref XP_002271116.1	PREDICTED: non-specific lipid-transfer protein [<i>Vitis vinifera</i>]	133265 bits(402)	2,64E-35	64/90(71%)	78/90(87%)	0/90(0%)
emb CBW38502.1	lipid transfer protein [<i>Helianthus annuus</i>]	132494 bits(402)	2,84E-35	70/99(71%)	79/99(80%)	0/99(0%)
ref XP_002305878.1	predicted protein [<i>Populus trichocarpa</i>]	147902 bits(402)	5,26E-41	71/102(70%)	84/102(82%)	0/102(0%)
gb AFM35695.1	nonspecific lipid transfer protein [<i>Vitis pseudoreticulata</i>]	145591 bits(402)	4,24E-40	72/103(70%)	87/103(84%)	1/103(1%)

ref XP_002305877.1	predicted protein [<i>Populus trichocarpa</i>]	145206 bits(402)	5,86E-40	71/102(70%)	81/102(79%)	0/102(0%)
gb ABK96515.1	unknown [<i>Populus trichocarpa</i> x <i>Populus deltoides</i>]	144436 bits(402)	9,88E-40	71/102(70%)	81/102(79%)	0/102(0%)
gb ADK60918.1	lipid transfer protein [<i>Castanea sativa</i>]	14405 bits(402)	1,70E-39	72/103(70%)	87/103(84%)	0/103(0%)
gb AEI70826.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	137117 bits(402)	7,77E-37	74/105(70%)	85/105(81%)	3/105(3%)
gb AAQ96338.1	lipid transfer protein [<i>Vitis aestivalis</i>]	132494 bits(402)	3,87E-35	63/90(70%)	77/90(86%)	0/90(0%)
sp P85894.1 LTP1_MORNI	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Allergen=Mor n 3	131339 bits(402)	6,35E-35	64/91(70%)	74/91(81%)	0/91(0%)
gb ACL01093.1	lipid transfer protein [<i>Castanea mollissima</i>]	14251 bits(402)	5,49E-39	71/103(69%)	86/103(83%)	0/103(0%)
gb AEC04836.1	lipid transfer protein [<i>Dimocarpus longan</i>]	142895 bits(402)	4,11E-39	69/103(67%)	87/103(84%)	0/103(0%)
ref XP_003517010.1	PREDICTED: non-specific lipid-transfer protein 1-like isoform 1 [<i>Glycine max</i>]	135576 bits(402)	2,93E-36	70/104(67%)	81/104(78%)	1/104(1%)
gb AAT68263.1	lipid transfer protein [<i>Nicotiana glauca</i>]	134806 bits(402)	5,43E-36	65/97(67%)	80/97(82%)	0/97(0%)
gb AGH27866.1	lipid transfer protein precursor [<i>Prunus kansuensis</i>]	132494 bits(402)	3,74E-35	69/103(67%)	81/103(79%)	1/103(1%)
gb AAV64877.1	non-specific lipid transfer protein [<i>Prunus persica</i>]	132109 bits(402)	5,53E-35	69/103(67%)	81/103(79%)	1/103(1%)
gb AAT68265.1	lipid transfer protein precursor [<i>Nicotiana glauca</i>]	129798 bits(402)	3,86E-34	62/92(67%)	76/92(83%)	0/92(0%)
gb ABK96813.1	lipid transfer protein isoform 1.1 precursor [<i>Lactuca sativa</i>]	13442 bits(402)	7,87E-36	68/103(66%)	80/103(78%)	3/103(3%)
gb ADR66950.1	non-specific lipid transfer protein [<i>Prunus sargentii</i>]	132109 bits(402)	5,89E-35	68/103(66%)	82/103(80%)	1/103(1%)
gb ADR66949.1	non-specific lipid transfer protein [<i>Prunus sargentii</i>]	131724 bits(402)	7,20E-35	68/103(66%)	81/103(79%)	1/103(1%)
gb AAV40850.1	lipid transfer protein 1 precursor [<i>Prunus persica</i>]	130183 bits(402)	2,82E-34	68/103(66%)	80/103(78%)	1/103(1%)
gb EMJ07873.1	hypothetical protein PRUPE_ppa023110mg [<i>Prunus persica</i>]	129798 bits(402)	4,66E-34	68/103(66%)	80/103(78%)	1/103(1%)
gb AAF28385.1 AF151214_1	lipid-transfer protein [<i>Nicotiana glauca</i>]	128257 bits(402)	1,99E-33	64/97(66%)	76/97(78%)	0/97(0%)
gb AGH27864.1	lipid transfer protein precursor [<i>Prunus davidiana</i>]	132494 bits(402)	4,20E-35	67/103(65%)	81/103(79%)	1/103(1%)
db BAI94496.1	lipid transfer protein [<i>Dianthus caryophyllus</i>]	132109 bits(402)	6,78E-35	66/102(65%)	80/102(78%)	0/102(0%)
gb ADR66953.1	non-specific lipid transfer protein [<i>Prunus dulcis</i>]	130954 bits(402)	1,71E-34	67/103(65%)	81/103(79%)	1/103(1%)
gb ADR66947.1	non-specific lipid transfer protein [<i>Prunus armeniaca</i>]	129798 bits(402)	4,68E-34	67/103(65%)	80/103(78%)	1/103(1%)
gb AGH27863.1	lipid transfer protein precursor [<i>Prunus mira</i>]	129413 bits(402)	5,84E-34	67/103(65%)	80/103(78%)	1/103(1%)
ref XP_003605269.1	Non-specific lipid-transfer protein [<i>Medicago truncatula</i>]	128642 bits(402)	4,65E-33	66/101(65%)	78/101(77%)	1/101(1%)
gb ACP43542.1	lipid transfer protein [<i>Chimonanthus praecox</i>]	125946 bits(402)	1,59E-32	61/94(65%)	73/94(78%)	0/94(0%)

sp Q43017.1 NLTP1_PRUDU	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; Flags: Precursor	130183 bits(402)	2,85E-34	66/103(64%)	81/103(79%)	1/103(1%)
gb AFK44274.1	unknown [Medicago truncatula]	129798 bits(402)	4,14E-34	66/103(64%)	80/103(78%)	1/103(1%)
ref XP_003605266.1	Non-specific lipid-transfer protein [Medicago truncatula]	129413 bits(402)	5,28E-34	66/103(64%)	80/103(78%)	1/103(1%)
gb ADR66948.1	non-specific lipid transfer protein [Prunus armeniaca]	127102 bits(402)	4,30E-33	66/103(64%)	79/103(77%)	1/103(1%)
gb ADR66945.1	non-specific lipid transfer protein [Prunus armeniaca]	126331 bits(402)	9,30E-33	67/104(64%)	79/104(76%)	1/104(1%)
gb ABY21307.1	pollen allergen Pla o 3 [Platanus orientalis]	130568 bits(402)	2,30E-34	65/103(63%)	79/103(77%)	0/103(0%)
gb AAY83345.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	130568 bits(402)	2,70E-34	65/103(63%)	82/103(80%)	1/103(1%)
sp Q9M5X8.1 NLTP_PRUAV	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Allergen=Pru av 3; Flags: Precursor	129028 bits(402)	8,73E-34	65/103(63%)	80/103(78%)	1/103(1%)
gb ADR66942.1	non-specific lipid transfer protein [Prunus avium]	129028 bits(402)	8,91E-34	65/103(63%)	80/103(78%)	1/103(1%)
sp A0AT29.1 NLTP2_LENCU	RecName: Full=Non-specific lipid-transfer protein 2; Short=LTP2; Contains: RecName: Full=Non-specific lipid-transfer protein 7; Short=LTP7; Flags: Precursor	129028 bits(402)	1,03E-33	65/103(63%)	78/103(76%)	0/103(0%)
dbj BAI94497.1	lipid transfer protein [Dianthus caryophyllus]	128642 bits(402)	1,36E-33	65/103(63%)	78/103(76%)	0/103(0%)
gb ADR66943.1	non-specific lipid transfer protein [Prunus avium]	128642 bits(402)	1,48E-33	65/103(63%)	80/103(78%)	1/103(1%)
ref XP_004302624.1	PREDICTED: non-specific lipid-transfer protein 1-like [Fragaria vesca subsp. vesca]	133265 bits(402)	2,00E-35	73/118(62%)	94/118(80%)	1/118(1%)
ref XP_004302608.1	PREDICTED: non-specific lipid-transfer protein-like isoform 1 [Fragaria vesca subsp. vesca]	132494 bits(402)	3,56E-35	64/103(62%)	82/103(80%)	1/103(1%)
gb AAM21292.1 AF369931_1	lipid-transfer protein [Citrus sinensis]	130183 bits(402)	3,62E-34	64/103(62%)	78/103(76%)	0/103(0%)
gb AAY83343.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	130183 bits(402)	3,83E-34	64/103(62%)	82/103(80%)	1/103(1%)
emb CAL07989.1	lipid transfer protein [Platanus x acerifolia]	129798 bits(402)	4,76E-34	64/103(62%)	79/103(77%)	0/103(0%)
gb ABX56711.1	LTP isoallergen 1 precursor [Arachis hypogaea]	127102 bits(402)	4,79E-33	65/104(62%)	80/104(77%)	2/104(2%)
gb AAT68262.1	lipid transfer protein [Nicotiana glauca]	125946 bits(402)	1,52E-32	63/101(62%)	77/101(76%)	0/101(0%)
gb ABG54494.1	putative allergen Rub i 3 [Rubus idaeus]	127487 bits(402)	3,91E-33	63/103(61%)	79/103(77%)	1/103(1%)
gb AFR31532.1	lipid transfer protein [Betula platyphylla]	126716 bits(402)	6,83E-33	64/105(61%)	80/105(76%)	2/105(2%)
gb AFI61838.1	harmalin [Peganum harmala]	132494 bits(402)	3,46E-35	62/103(60%)	81/103(79%)	0/103(0%)
emb CAC86258.1	lipid transfer protein [Fragaria x ananassa]	128257 bits(402)	1,77E-33	71/118(60%)	91/118(77%)	1/118(1%)
dbj BAI94498.1	lipid transfer protein [Dianthus caryophyllus]	126716 bits(402)	7,48E-33	61/102(60%)	76/102(75%)	0/102(0%)

emb CBI18043.3	unnamed protein product [Vitis vinifera]	13442 bits(402)	1,63E-35	60/102(59%)	83/102(81%)	0/102(0%)
ref XP_002270970.1	PREDICTED: non-specific lipid-transfer protein P5-like [Vitis vinifera]	132494 bits(402)	4,18E-35	60/102(59%)	83/102(81%)	0/102(0%)
emb CAN70512.1	hypothetical protein VITISV_005630 [Vitis vinifera]	130954 bits(402)	1,78E-34	59/100(59%)	82/100(82%)	0/100(0%)
ref XP_002528683.1	Nonspecific lipid-transfer protein precursor, putative [Ricinus communis]	127487 bits(402)	3,04E-33	61/103(59%)	79/103(77%)	3/103(3%)
gb AAK28533.1 AF329829_1	lipid transfer protein precursor [Corylus avellana]	129028 bits(402)	9,08E-34	59/103(57%)	81/103(79%)	0/103(0%)

Contig7

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb ACD69680.1	metallothionein [Mangifera indica]	835741 bits(205)	5,74E-18	39/51(76%)	43/51(84%)	0/51(0%)
gb ADH04476.1	metallothionein [Mangifera indica]	835741 bits(205)	5,61E-18	40/53(75%)	44/53(83%)	0/53(0%)
gb ABD97258.1	metallothionin 2 [Camellia sinensis]	650846 bits(205)	4,19E-11	35/50(70%)	38/50(76%)	1/50(2%)
dbj BAA96449.1	metallothionein-like protein [Pyrus pyrifolia]	666254 bits(205)	1,13E-11	34/52(65%)	40/52(77%)	1/52(2%)
gb AFP93964.1	metallothionein type 2 [Ilex paraguariensis]	631586 bits(205)	1,88E-10	32/50(64%)	38/50(76%)	1/50(2%)
gb EMJ06266.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	654698 bits(205)	2,20E-11	33/52(63%)	41/52(79%)	1/52(2%)
gb EMJ06265.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	646994 bits(205)	5,57E-11	33/52(63%)	41/52(79%)	1/52(2%)
gb ABN46988.1	metallothionein-like protein 2b [Nelumbo nucifera]	63929 bits(205)	8,51E-11	33/52(63%)	40/52(77%)	3/52(6%)
gb ABD97257.1	metallothionin 1 [Camellia sinensis]	627734 bits(205)	2,25E-10	31/50(62%)	37/50(74%)	1/50(2%)
gb ACS14665.1	metallothionein [Camellia oleifera]	62003 bits(205)	4,60E-10	31/50(62%)	37/50(74%)	1/50(2%)
ref XP_002314182.1	predicted protein [Populus trichocarpa]	635438 bits(205)	7,76E-11	30/50(60%)	38/50(76%)	1/50(2%)
gb ABK94874.1	unknown [Populus trichocarpa]	635438 bits(205)	1,34E-10	30/50(60%)	38/50(76%)	1/50(2%)
gb AAT02525.1	metallothionein 2b [Populus trichocarpa X Populus deltoides]	62003 bits(205)	4,09E-10	30/50(60%)	37/50(74%)	1/50(2%)
gb ABM21762.1	metallothionein-like protein MT2B [Salix matsudana]	62003 bits(205)	4,97E-10	30/50(60%)	37/50(74%)	1/50(2%)
ref XP_002329424.1	predicted protein [Populus trichocarpa]	612326 bits(205)	5,74E-10	30/50(60%)	36/50(72%)	1/50(2%)

ref XP_002299873.1	predicted protein [Populus trichocarpa]	616178 bits(205)	5,93E-10	30/50(60%)	38/50(76%)	1/50(2%)
dbj BAJ39944.1	metallothionein [Ziziphus jujuba]	627734 bits(205)	2,61E-10	29/51(57%)	38/51(75%)	2/51(4%)
Contig9						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_002336827.1	predicted protein [Populus trichocarpa]	105916 bits(419)	3,60E-26	51/53(96%)	52/53(98%)	0/53(0%)
gb ABA27086.1	TO121-23 [Taraxacum officinale]	135576 bits(419)	2,39E-37	67/74(91%)	69/74(93%)	0/74(0%)
tpg DAA61430.1	TPA: hypothetical protein ZEAMMB73_025341 [Zea mays]	101293 bits(419)	2,98E-24	49/54(91%)	51/54(94%)	0/54(0%)
gb AFW61204.1	hypothetical protein ZEAMMB73_915933 [Zea mays]	101293 bits(419)	3,86E-24	49/54(91%)	51/54(94%)	0/54(0%)
gb AFW57302.1	hypothetical protein ZEAMMB73_857117 [Zea mays]	152525 bits(419)	7,54E-44	76/88(86%)	81/88(92%)	0/88(0%)
gb AFW61205.1	hypothetical protein ZEAMMB73_915933 [Zea mays]	15214 bits(419)	1,25E-43	76/88(86%)	81/88(92%)	0/88(0%)
ref NP_001077600.1	60S ribosomal protein L34-1 [Arabidopsis thaliana]	156377 bits(419)	3,04E-45	75/88(85%)	83/88(94%)	0/88(0%)
tpg DAA42039.1	TPA: hypothetical protein ZEAMMB73_205664 [Zea mays]	955153 bits(419)	5,54E-22	46/54(85%)	50/54(93%)	0/54(0%)
ref XP_002322305.1	predicted protein [Populus trichocarpa]	159458 bits(419)	2,56E-46	77/92(84%)	84/92(91%)	0/92(0%)
gb EAZ05646.1	hypothetical protein Osl_27865 [Oryza sativa Indica Group]	997525 bits(419)	8,63E-24	49/59(83%)	52/59(88%)	0/59(0%)
ref XP_002877140.1	hypothetical protein ARALYDRAFT_484659 [Arabidopsis lyrata subsp. lyrata]	149443 bits(419)	2,97E-42	72/88(82%)	77/88(88%)	0/88(0%)
gb AAT08730.1	ribosomal protein L34 [Hyacinthus orientalis]	146362 bits(419)	2,26E-41	72/88(82%)	78/88(89%)	0/88(0%)
ref NP_001152634.1	60S ribosomal protein L34 [Zea mays]	156762 bits(419)	2,93E-44	77/96(80%)	83/96(86%)	0/96(0%)
ref XP_002329624.1	predicted protein [Populus trichocarpa]	157918 bits(419)	9,39E-46	76/96(79%)	84/96(88%)	0/96(0%)
gb EMJ28713.1	hypothetical protein PRUPE_ppa013472mg [Prunus persica]	166007 bits(419)	1,10E-48	82/108(76%)	91/108(84%)	0/108(0%)
ref XP_004138165.1	PREDICTED: 60S ribosomal protein L34-like [Cucumis sativus]	164466 bits(419)	4,27E-48	82/108(76%)	90/108(83%)	0/108(0%)
gb ADR71241.1	60S ribosomal protein L34A [Hevea brasiliensis]	162925 bits(419)	1,70E-47	82/108(76%)	89/108(82%)	0/108(0%)
ref XP_004138175.1	PREDICTED: 60S ribosomal protein L34-like [Cucumis sativus]	162925 bits(419)	1,75E-47	82/108(76%)	89/108(82%)	0/108(0%)

gb EMJ10969.1	hypothetical protein PRUPE_ppa013459mg [Prunus persica]	163696 bits(419)	8,59E-48	81/108(75%)	90/108(83%)	0/108(0%)
ref XP_002324052.1	predicted protein [Populus trichocarpa]	16331 bits(419)	1,51E-47	81/108(75%)	90/108(83%)	0/108(0%)
ref XP_002318779.1	predicted protein [Populus trichocarpa]	162925 bits(419)	1,76E-47	81/108(75%)	90/108(83%)	0/108(0%)
gb EMJ19008.1	hypothetical protein PRUPE_ppb011184mg [Prunus persica]	162925 bits(419)	1,91E-47	81/108(75%)	90/108(83%)	0/108(0%)
ref XP_004304307.1	PREDICTED: 60S ribosomal protein L34-like [Fragaria vesca subsp. vesca]	16254 bits(419)	2,35E-47	81/108(75%)	90/108(83%)	0/108(0%)
gb ADR71242.1	60S ribosomal protein L34B [Hevea brasiliensis]	16177 bits(419)	5,38E-47	81/108(75%)	89/108(82%)	0/108(0%)
ref XP_003517974.1	PREDICTED: 60S ribosomal protein L34-like [Glycine max]	16177 bits(419)	6,32E-47	81/108(75%)	89/108(82%)	0/108(0%)
ref NP_001236831.1	uncharacterized protein LOC100305798 [Glycine max]	160229 bits(419)	1,92E-46	81/108(75%)	88/108(81%)	0/108(0%)
ref XP_004297304.1	PREDICTED: 60S ribosomal protein L34-1-like [Fragaria vesca subsp. vesca]	16331 bits(419)	1,23E-47	80/108(74%)	90/108(83%)	0/108(0%)
ref XP_004299366.1	PREDICTED: 60S ribosomal protein L34-like [Fragaria vesca subsp. vesca]	162155 bits(419)	3,50E-47	80/108(74%)	90/108(83%)	0/108(0%)
ref XP_002522429.1	60S ribosomal protein L34, putative [Ricinus communis]	161384 bits(419)	7,12E-47	80/108(74%)	89/108(82%)	0/108(0%)
gb AFK36300.1	unknown [Lotus japonicus]	159458 bits(419)	3,66E-46	80/108(74%)	88/108(81%)	0/108(0%)
sp P40590.1 RL34_PEA	RecName: Full=60S ribosomal protein L34	159458 bits(419)	3,82E-46	80/108(74%)	88/108(81%)	0/108(0%)
ref NP_001238172.1	uncharacterized protein LOC100499931 [Glycine max]	159073 bits(419)	5,95E-46	80/108(74%)	88/108(81%)	0/108(0%)
ref NP_001237362.1	uncharacterized protein LOC100499882 [Glycine max]	158303 bits(419)	1,20E-45	80/108(74%)	87/108(81%)	0/108(0%)
gb AFK47541.1	unknown [Lotus japonicus]	158303 bits(419)	1,35E-45	80/108(74%)	87/108(81%)	0/108(0%)
ref XP_003621181.1	60S ribosomal protein L34 [Medicago truncatula]	157918 bits(419)	1,44E-45	80/108(74%)	87/108(81%)	0/108(0%)
ref XP_003613762.1	60S ribosomal protein L34 [Medicago truncatula]	157918 bits(419)	1,62E-45	80/108(74%)	87/108(81%)	0/108(0%)
gb ACJ84129.1	unknown [Medicago truncatula]	157532 bits(419)	2,74E-45	80/108(74%)	87/108(81%)	0/108(0%)
ref XP_003540043.1	PREDICTED: 60S ribosomal protein L34-like [Glycine max]	145206 bits(419)	1,49E-40	71/96(74%)	80/96(83%)	0/96(0%)
ref XP_004238799.1	PREDICTED: 60S ribosomal protein L34-like [Solanum lycopersicum]	161384 bits(419)	8,83E-47	79/108(73%)	90/108(83%)	0/108(0%)
ref XP_004232223.1	PREDICTED: 60S ribosomal protein L34-like isoform 1 [Solanum lycopersicum]	160614 bits(419)	1,34E-46	79/108(73%)	89/108(82%)	0/108(0%)
ref XP_002534058.1	60S ribosomal protein L34, putative [Ricinus communis]	160229 bits(419)	2,41E-46	79/108(73%)	89/108(82%)	0/108(0%)
ref XP_004241579.1	PREDICTED: 60S ribosomal protein L34-like [Solanum lycopersicum]	159844 bits(419)	2,98E-46	79/108(73%)	89/108(82%)	0/108(0%)

gb AAT39969.1	60S ribosomal protein L34, putative [Solanum demissum]	159844 bits(419)	3,40E-46	79/108(73%)	89/108(82%)	0/108(0%)
ref XP_002273676.1	PREDICTED: 60S ribosomal protein L34-like [Vitis vinifera]	157918 bits(419)	1,92E-45	79/108(73%)	88/108(81%)	0/108(0%)
ref XP_002445552.1	hypothetical protein SORBIDRAFT_07g021410 [Sorghum bicolor]	157147 bits(419)	2,90E-45	79/108(73%)	88/108(81%)	0/108(0%)
gb EAZ09088.1	hypothetical protein Osl_31355 [Oryza sativa Indica Group]	157147 bits(419)	3,72E-45	79/108(73%)	88/108(81%)	0/108(0%)
gb ACF86965.1	unknown [Zea mays]	156377 bits(419)	6,87E-45	79/108(73%)	88/108(81%)	0/108(0%)
ref NP_001132454.1	60S ribosomal protein L34 [Zea mays]	156377 bits(419)	7,00E-45	79/108(73%)	88/108(81%)	0/108(0%)
gb ACJ83310.1	unknown [Medicago truncatula]	155606 bits(419)	1,25E-44	79/108(73%)	86/108(80%)	0/108(0%)
gb ABK93580.1	unknown [Populus trichocarpa]	16331 bits(419)	1,43E-47	81/112(72%)	92/112(82%)	0/112(0%)
ref NP_174010.1	60S ribosomal protein L34-1 [Arabidopsis thaliana]	160229 bits(419)	1,94E-46	78/108(72%)	90/108(83%)	0/108(0%)
sp P41098.1 RL34_TOBAC	RecName: Full=60S ribosomal protein L34	159458 bits(419)	3,66E-46	78/108(72%)	89/108(82%)	0/108(0%)
ref NP_001061026.1	Os08g0156800 [Oryza sativa Japonica Group]	155992 bits(419)	9,94E-45	78/108(72%)	88/108(81%)	0/108(0%)
ref NP_001061890.1	Os08g0436800 [Oryza sativa Japonica Group]	155992 bits(419)	1,00E-44	78/108(72%)	88/108(81%)	0/108(0%)
ref XP_003573475.1	PREDICTED: 60S ribosomal protein L34-like [Brachypodium distachyon]	154451 bits(419)	3,14E-44	78/108(72%)	87/108(81%)	0/108(0%)
ref XP_003573474.1	PREDICTED: 60S ribosomal protein L34-like [Brachypodium distachyon]	154451 bits(419)	3,27E-44	78/108(72%)	87/108(81%)	0/108(0%)
ref NP_001152744.1	60S ribosomal protein L34 [Zea mays]	154451 bits(419)	3,34E-44	78/108(72%)	87/108(81%)	0/108(0%)
gb AAM66030.1	60s ribosomal protein L34 [Arabidopsis thaliana]	159073 bits(419)	6,99E-46	77/108(71%)	90/108(83%)	0/108(0%)
ref XP_002887248.1	hypothetical protein ARALYDRAFT_894747 [Arabidopsis lyrata subsp. lyrata]	158303 bits(419)	1,28E-45	77/108(71%)	89/108(82%)	0/108(0%)
ref NP_177120.1	60S ribosomal protein L34-2 [Arabidopsis thaliana]	157532 bits(419)	2,44E-45	77/108(71%)	88/108(81%)	0/108(0%)
ref NP_189532.1	60S ribosomal protein L34-3 [Arabidopsis thaliana]	156377 bits(419)	6,22E-45	77/108(71%)	86/108(80%)	0/108(0%)
dbj BAK01360.1	predicted protein [Hordeum vulgare subsp. vulgare]	15368 bits(419)	7,65E-44	77/108(71%)	87/108(81%)	0/108(0%)
pdb 3IZR II	Chain i, Localization Of The Large Subunit Ribosomal Proteins Into A 5.5 A Cryo-Em Map Of Triticum Aestivum Translating 80s Ribosome	152525 bits(419)	2,20E-43	77/108(71%)	86/108(80%)	0/108(0%)
ref XP_003574550.1	PREDICTED: 60S ribosomal protein L34-like [Brachypodium distachyon]	151369 bits(419)	5,72E-43	77/108(71%)	86/108(80%)	0/108(0%)
ref NP_001152435.1	60S ribosomal protein L34 [Zea mays]	160229 bits(419)	1,82E-46	82/117(70%)	93/117(79%)	0/117(0%)
gb ACG48574.1	60S ribosomal protein L34 [Zea mays]	159844 bits(419)	2,66E-46	82/117(70%)	93/117(79%)	0/117(0%)

ref XP_002443887.1	hypothetical protein SORBIDRAFT_07g003770 [Sorghum bicolor]	159458 bits(419)	4,70E-46	81/117(69%)	93/117(79%)	0/117(0%)
ref NP_001151941.1	60S ribosomal protein L34 [Zea mays]	158303 bits(419)	1,08E-45	81/117(69%)	92/117(79%)	0/117(0%)
gb ACG30531.1	60S ribosomal protein L34 [Zea mays]	157918 bits(419)	1,62E-45	81/117(69%)	92/117(79%)	0/117(0%)
gb ACG33498.1	60S ribosomal protein L34 [Zea mays]	157918 bits(419)	1,83E-45	81/117(69%)	92/117(79%)	0/117(0%)
ref XP_002462376.1	hypothetical protein SORBIDRAFT_02g024610 [Sorghum bicolor]	157532 bits(419)	2,72E-45	81/117(69%)	92/117(79%)	0/117(0%)
gb AAT38711.2	60S ribosomal protein L34, putative [Solanum demissum]	15291 bits(419)	1,81E-43	79/114(69%)	89/114(78%)	6/114(5%)
gb ABK21342.1	unknown [Picea sitchensis]	150984 bits(419)	8,13E-43	74/108(69%)	86/108(80%)	0/108(0%)
gb ABD66516.1	ribosomal protein [Gymnadenia conopsea]	146362 bits(419)	5,31E-41	75/108(69%)	85/108(79%)	0/108(0%)
ref XP_002305630.1	predicted protein [Populus trichocarpa]	142124 bits(419)	1,97E-39	72/106(68%)	83/106(78%)	0/106(0%)
ref XP_002318778.1	predicted protein [Populus trichocarpa]	139043 bits(419)	2,79E-38	73/108(68%)	82/108(76%)	10/108(9%)
gb AAO19740.1	60S ribosomal protein L34, partial [Orobanche cernua var. cumana]	120168 bits(419)	3,62E-31	61/90(68%)	70/90(78%)	0/90(0%)
gb AFW62295.1	hypothetical protein ZEAMMB73_590401 [Zea mays]	109383 bits(419)	2,79E-26	55/82(67%)	64/82(78%)	0/82(0%)
ref XP_001751741.1	predicted protein [Physcomitrella patens subsp. patens]	13288 bits(419)	9,71E-36	69/112(62%)	80/112(71%)	0/112(0%)
ref XP_001762793.1	predicted protein [Physcomitrella patens subsp. patens]	132494 bits(419)	1,06E-35	69/112(62%)	80/112(71%)	0/112(0%)
ref XP_001765474.1	predicted protein [Physcomitrella patens subsp. patens]	131339 bits(419)	3,18E-35	68/112(61%)	80/112(71%)	0/112(0%)
ref XP_001759832.1	predicted protein [Physcomitrella patens subsp. patens]	130568 bits(419)	6,16E-35	68/112(61%)	81/112(72%)	0/112(0%)
ref XP_001759827.1	predicted protein [Physcomitrella patens subsp. patens]	130183 bits(419)	1,02E-34	68/112(61%)	80/112(71%)	0/112(0%)
ref XP_001776049.1	predicted protein [Physcomitrella patens subsp. patens]	129798 bits(419)	1,31E-34	67/112(60%)	81/112(72%)	0/112(0%)
ref XP_002980840.1	hypothetical protein SELMODRAFT_113433 [Selaginella moellendorffii]	114005 bits(419)	6,07E-29	52/88(59%)	72/88(82%)	0/88(0%)
gb ACO13005.1	60S ribosomal protein L34 [Lepeophtheirus salmonis]	101293 bits(419)	9,93E-24	54/96(56%)	67/96(70%)	2/96(2%)
gb AFK44937.1	unknown [Lotus japonicus]	947449 bits(419)	2,05E-21	53/96(55%)	63/96(66%)	2/96(2%)
ref XP_003621256.1	60S ribosomal protein L34 [Medicago truncatula]	970561 bits(419)	2,08E-22	52/96(54%)	59/96(61%)	16/96(17%)
gb EIE22072.1	hypothetical protein COCSUDRAFT_16876 [Coccomyxa subellipsoidea C-169]	951301 bits(419)	6,47E-21	51/96(53%)	65/96(68%)	0/96(0%)
gb ACO10322.1	60S ribosomal protein L34 [Caligus rogercresseyi]	978265 bits(419)	1,72E-22	50/96(52%)	69/96(72%)	2/96(2%)

gb ACO14672.1	60S ribosomal protein L34 [Caligus clemensi]	970561 bits(419)	2,92E-22	50/96(52%)	70/96(73%)	2/96(2%)
gb EFN58323.1	hypothetical protein CHLNCDRAFT_20596 [Chlorella variabilis]	100908 bits(419)	4,23E-23	49/96(51%)	67/96(70%)	2/96(2%)
emb CAR63541.1	putative Ribosomal Protein [Angiostrongylus cantonensis]	889669 bits(419)	3,33E-19	49/96(51%)	63/96(66%)	2/96(2%)
dbj GAA96860.1	hypothetical protein E5Q_03533 [Mixia osmundae IAM 14324]	889669 bits(419)	3,17E-19	48/96(50%)	61/96(64%)	2/96(2%)
emb CBY09263.1	unnamed protein product [Oikopleura dioica]	870409 bits(419)	1,49E-18	48/96(50%)	66/96(69%)	2/96(2%)
ref XP_001751736.1	predicted protein [Physcomitrella patens subsp. patens]	912781 bits(419)	3,58E-20	49/100(49%)	62/100(62%)	0/100(0%)
gb EKM78285.1	hypothetical protein AGABI1DRAFT_85942 [Agaricus bisporus var. burnettii JB137-S8]	893521 bits(419)	2,01E-19	49/101(49%)	61/101(60%)	2/101(2%)
gb ADV03019.1	ribosomal protein L34 [Karlodinium veneficum]	874261 bits(419)	1,40E-18	46/96(48%)	58/96(60%)	2/96(2%)
dbj GAA96859.1	hypothetical protein E5Q_03532 [Mixia osmundae IAM 14324]	932041 bits(419)	2,14E-20	56/125(45%)	73/125(58%)	21/125(17%)
gb EGG00349.1	hypothetical protein MELLADRAFT_50435 [Melampsora larici-populina 98AG31]	885817 bits(419)	5,35E-19	51/116(44%)	69/116(59%)	2/116(2%)

Contig12

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_003537949.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	204142 bits(518)	1,88E-57	96/131(73%)	111/131(85%)	0/131(0%)
gb ABK94688.1	unknown [Populus trichocarpa]	145591 bits(518)	3,65E-41	66/91(73%)	78/91(86%)	0/91(0%)
ref XP_002513443.1	lipoxygenase, putative [Ricinus communis]	199519 bits(518)	7,32E-56	93/130(72%)	107/130(82%)	0/130(0%)
gb AAZ57444.1	lipoxygenase LOX1 [Populus deltoides]	199134 bits(518)	1,15E-55	93/130(72%)	108/130(83%)	0/130(0%)
ref XP_002323952.1	predicted protein [Populus trichocarpa]	195667 bits(518)	1,51E-54	91/130(70%)	107/130(82%)	0/130(0%)
gb ABM88259.1	lipoxygenase [Phaseolus vulgaris]	191815 bits(518)	4,26E-53	92/131(70%)	105/131(80%)	0/131(0%)
gb EMJ07637.1	hypothetical protein PRUPE_ppa000968mg [Prunus persica]	196052 bits(518)	2,18E-54	91/131(69%)	106/131(81%)	0/131(0%)
ref XP_004306503.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	194897 bits(518)	4,11E-54	90/131(69%)	108/131(82%)	0/131(0%)

dbj BAA13542.1	CPRD46 protein [Vigna unguiculata]	192586 bits(518)	2,64E-53	89/131(68%)	107/131(82%)	0/131(0%)
gb ADZ73653.1	putative 13-lipoxygenase [Capsicum annuum]	191045 bits(518)	8,86E-53	89/131(68%)	105/131(80%)	0/131(0%)
dbj BAB84352.1	lipoxygenase [Citrus jambhiri]	189504 bits(518)	2,81E-52	89/130(68%)	104/130(80%)	0/130(0%)
ref XP_002297796.1	predicted protein [Populus trichocarpa]	189119 bits(518)	4,02E-52	89/131(68%)	106/131(81%)	0/131(0%)
emb CAN80268.1	hypothetical protein VITISV_027685 [Vitis vinifera]	193741 bits(518)	1,11E-55	88/131(67%)	109/131(83%)	0/131(0%)
ref XP_002285574.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic [Vitis vinifera]	193356 bits(518)	1,37E-53	88/131(67%)	109/131(83%)	0/131(0%)
gb ACZ17391.1	lipoxygenase [Vitis vinifera]	193356 bits(518)	1,45E-53	88/131(67%)	109/131(83%)	0/131(0%)
emb CAA05278.1	loxc homologue [Solanum pimpinellifolium]	19143 bits(518)	3,85E-53	88/131(67%)	104/131(79%)	0/131(0%)
ref NP_001233812.1	lipoxygenase [Solanum lycopersicum]	189889 bits(518)	2,49E-52	87/130(67%)	104/130(80%)	0/130(0%)
gb AAZ57445.1	lipoxygenase LOX2 [Populus deltoides]	187963 bits(518)	1,05E-51	88/131(67%)	105/131(80%)	0/131(0%)
gb EMJ21452.1	hypothetical protein PRUPE_ppa001634mg [Prunus persica]	182956 bits(518)	3,73E-50	88/131(67%)	103/131(79%)	0/131(0%)
gb ACJ54281.1	lipoxygenase [Camellia sinensis]	179104 bits(518)	1,84E-48	88/131(67%)	103/131(79%)	0/131(0%)
gb AAT77551.1	LoxC-like [Solanum pimpinellifolium]	184882 bits(518)	2,24E-54	86/131(66%)	105/131(80%)	0/131(0%)
emb CAA05270.1	unnamed protein product [Solanum habrochaites]	184496 bits(518)	4,45E-54	86/131(66%)	105/131(80%)	0/131(0%)
gb AAD13306.1	lipoxygenase [Solanum lycopersicum]	185652 bits(518)	1,27E-53	86/131(66%)	105/131(80%)	0/131(0%)
emb CAN81242.1	hypothetical protein VITISV_001963 [Vitis vinifera]	187963 bits(518)	1,39E-53	87/131(66%)	104/131(79%)	0/131(0%)
emb CAA05277.1	loxc homologue [Solanum pimpinellifolium]	184882 bits(518)	2,19E-53	86/131(66%)	105/131(80%)	0/131(0%)
gb ABR13283.1	putative lipoxygenase [Prunus dulcis]	177178 bits(518)	6,45E-53	84/127(66%)	100/127(79%)	0/127(0%)
ref XP_002314229.1	predicted protein [Populus trichocarpa]	19066 bits(518)	6,69E-53	87/131(66%)	107/131(82%)	0/131(0%)
ref XP_004228859.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Solanum lycopersicum]	186037 bits(518)	8,92E-53	86/131(66%)	105/131(80%)	0/131(0%)
emb CAA05280.1	loxc homologue [Solanum lycopersicum]	185652 bits(518)	1,22E-52	86/131(66%)	105/131(80%)	0/131(0%)
ref XP_002513441.1	lipoxygenase, putative [Ricinus communis]	189889 bits(518)	1,25E-52	87/131(66%)	106/131(81%)	0/131(0%)
ref XP_002263854.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic [Vitis vinifera]	188734 bits(518)	7,05E-52	87/131(66%)	104/131(79%)	0/131(0%)
ref XP_002283123.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Vitis vinifera]	187963 bits(518)	8,44E-52	87/131(66%)	104/131(79%)	0/131(0%)
emb CBI16438.3	unnamed protein product [Vitis vinifera]	187578 bits(518)	1,12E-51	87/131(66%)	104/131(79%)	0/131(0%)
gb ADO51752.1	lipoxygenase [Camellia sinensis]	187193 bits(518)	2,25E-51	87/131(66%)	104/131(79%)	0/131(0%)

ref XP_004296344.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	186808 bits(518)	5,37E-51	86/131(66%)	104/131(79%)	0/131(0%)
ref NP_001234259.1	lipoxygenase [Solanum lycopersicum]	186037 bits(518)	6,22E-51	86/131(66%)	105/131(80%)	0/131(0%)
gb ACQ76787.1	lipoxygenase [Camellia sinensis]	185267 bits(518)	1,00E-50	86/131(66%)	103/131(79%)	0/131(0%)
ref XP_003555640.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	184111 bits(518)	2,47E-50	87/131(66%)	101/131(77%)	0/131(0%)
ref XP_003539375.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	183341 bits(518)	5,51E-50	86/131(66%)	103/131(79%)	0/131(0%)
gb AFK49450.1	unknown [Lotus japonicus]	185267 bits(518)	1,21E-53	85/131(65%)	105/131(80%)	0/131(0%)
sp O24370.1 LOX21_SOLTU	RecName: Full=Linoleate 13S-lipoxygenase 2-1, chloroplastic; AltName: Full=Lipoxygenase 2-1; Flags: Precursor	188348 bits(518)	8,02E-52	85/131(65%)	104/131(79%)	0/131(0%)
gb AAP83137.1	lipoxygenase, partial [Nicotiana attenuata]	186808 bits(518)	3,03E-51	85/130(65%)	104/130(80%)	0/130(0%)
ref XP_004301613.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	186037 bits(518)	5,25E-51	85/131(65%)	104/131(79%)	0/131(0%)
ref XP_004295226.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	185652 bits(518)	1,26E-50	85/131(65%)	102/131(78%)	0/131(0%)
gb EMJ22184.1	hypothetical protein PRUPE_ppa001631mg [Prunus persica]	179874 bits(518)	4,23E-49	85/131(65%)	103/131(79%)	0/131(0%)
ref XP_002283135.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Vitis vinifera]	184496 bits(518)	1,45E-50	84/131(64%)	103/131(79%)	0/131(0%)
emb CBI16439.3	unnamed protein product [Vitis vinifera]	184496 bits(518)	1,45E-49	84/131(64%)	103/131(79%)	0/131(0%)
ref XP_003537948.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	177948 bits(518)	4,73E-48	84/131(64%)	99/131(76%)	0/131(0%)
ref XP_002299250.1	predicted protein [Populus trichocarpa]	176407 bits(518)	1,55E-47	84/131(64%)	102/131(78%)	0/131(0%)
gb AAD31897.1 AF145479_1	lipoxygenase [Mesembryanthemum crystallinum]	170629 bits(518)	1,92E-48	83/131(63%)	98/131(75%)	0/131(0%)
gb EMJ21799.1	hypothetical protein PRUPE_ppa001316mg [Prunus persica]	178333 bits(518)	2,36E-48	82/131(63%)	102/131(78%)	0/131(0%)
ref XP_004168530.1	PREDICTED: LOW QUALITY PROTEIN: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Cucumis sativus]	176022 bits(518)	2,71E-47	83/132(63%)	102/132(77%)	0/132(0%)
gb ADL28386.1	lipoxygenase [Carya cathayensis]	154836 bits(518)	8,32E-44	73/116(63%)	89/116(77%)	0/116(0%)
dbj BAJ34928.1	lipoxygenase [Vitis hybrid cultivar]	165622 bits(518)	1,12E-43	74/118(63%)	96/118(81%)	0/118(0%)
gb ACR38701.1	unknown [Zea mays]	154066 bits(518)	1,17E-43	71/113(63%)	88/113(78%)	0/113(0%)
gb ACF79556.1	unknown [Zea mays]	154066 bits(518)	6,19E-42	71/113(63%)	88/113(78%)	0/113(0%)
gb ACN26379.1	unknown [Zea mays]	154451 bits(518)	1,06E-40	71/113(63%)	88/113(78%)	0/113(0%)

gb AFW66200.1	hypothetical protein ZEAMMB73_503041 [Zea mays]	154451 bits(518)	8,97E-40	71/113(63%)	88/113(78%)	0/113(0%)
gb AFW66201.1	hypothetical protein ZEAMMB73_503041 [Zea mays]	154451 bits(518)	1,12E-39	71/113(63%)	88/113(78%)	0/113(0%)
gb ABF01001.1	lipoxygenase [Zea mays]	154451 bits(518)	1,13E-39	71/113(63%)	88/113(78%)	0/113(0%)
ref NP_001105980.1	lipoxygenase10 [Zea mays]	154451 bits(518)	1,15E-39	71/113(63%)	88/113(78%)	0/113(0%)
gb ACS28586.1	lipoxygenase [Passiflora edulis]	17217 bits(518)	9,98E-51	81/131(62%)	104/131(79%)	0/131(0%)
gb AAA16093.1	lipoxygenase, partial [Cuscuta reflexa]	176022 bits(518)	1,38E-49	81/130(62%)	99/130(76%)	0/130(0%)
ref XP_002520644.1	lipoxygenase, putative [Ricinus communis]	179489 bits(518)	9,05E-49	82/133(62%)	105/133(79%)	1/133(1%)
ref XP_003543820.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	179489 bits(518)	9,48E-49	81/131(62%)	101/131(77%)	0/131(0%)
ref XP_004142137.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Cucumis sativus]	173326 bits(518)	2,20E-46	82/132(62%)	100/132(76%)	0/132(0%)
ref XP_004142135.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Cucumis sativus]	17217 bits(518)	2,23E-46	82/132(62%)	101/132(77%)	0/132(0%)
ref XP_004169303.1	PREDICTED: LOW QUALITY PROTEIN: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Cucumis sativus]	173326 bits(518)	2,28E-46	82/132(62%)	100/132(76%)	0/132(0%)
ref XP_003519005.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	1714 bits(518)	1,22E-45	81/131(62%)	100/131(76%)	0/131(0%)
ref XP_003601322.1	Lipoxygenase [Medicago truncatula]	156762 bits(518)	1,56E-40	81/131(62%)	100/131(76%)	0/131(0%)
ref XP_004251816.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Solanum lycopersicum]	177178 bits(518)	7,91E-48	80/131(61%)	99/131(76%)	0/131(0%)
gb ABF59998.1	lipoxygenase 4 [Actinidia deliciosa]	167548 bits(518)	9,71E-48	81/133(61%)	101/133(76%)	2/133(2%)
ref XP_004172896.1	PREDICTED: LOW QUALITY PROTEIN: linoleate 13S-lipoxygenase 2-1, chloroplastic-like, partial [Cucumis sativus]	170629 bits(518)	9,51E-46	80/132(61%)	99/132(75%)	0/132(0%)
ref XP_003571227.1	PREDICTED: lipoxygenase 2.3, chloroplastic-like [Brachypodium distachyon]	169474 bits(518)	5,88E-45	80/131(61%)	101/131(77%)	0/131(0%)
ref XP_003571069.1	PREDICTED: lipoxygenase 2.3, chloroplastic-like [Brachypodium distachyon]	169088 bits(518)	8,15E-45	80/131(61%)	100/131(76%)	0/131(0%)
sp Q8GSM2.1 LOX23_HORVU	RecName: Full=Lipoxygenase 2.3, chloroplastic; AltName: Full=LOX2:Hv:3; Flags: Precursor	167162 bits(518)	3,69E-44	80/131(61%)	101/131(77%)	0/131(0%)
dbj BAK01953.1	predicted protein [Hordeum vulgare subsp. vulgare]	167162 bits(518)	3,84E-44	80/131(61%)	101/131(77%)	0/131(0%)
gb EEC72668.1	hypothetical protein Osl_06220 [Oryza sativa Indica Group]	167162 bits(518)	3,90E-44	80/131(61%)	99/131(76%)	0/131(0%)
ref XP_003556820.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	176022 bits(518)	1,70E-47	78/131(60%)	99/131(76%)	0/131(0%)
ref XP_002513228.1	lipoxygenase, putative [Ricinus communis]	174481 bits(518)	9,58E-47	79/131(60%)	104/131(79%)	0/131(0%)

ref XP_004308211.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	170244 bits(518)	2,40E-45	79/131(60%)	97/131(74%)	0/131(0%)
gb ACR38419.1	unknown [Zea mays]	160999 bits(518)	1,46E-44	79/131(60%)	100/131(76%)	2/131(2%)
ref XP_002875714.1	hypothetical protein ARALYDRAFT_484907 [Arabidopsis lyrata subsp. lyrata]	166007 bits(518)	9,35E-44	78/130(60%)	94/130(72%)	0/130(0%)
ref XP_002283147.2	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Vitis vinifera]	165236 bits(518)	1,52E-43	78/131(60%)	98/131(75%)	3/131(2%)
gb ACN33257.1	unknown [Zea mays]	162155 bits(518)	1,31E-42	79/131(60%)	100/131(76%)	2/131(2%)
gb ABF01002.1	lipoxygenase [Zea mays]	16254 bits(518)	1,47E-42	79/131(60%)	100/131(76%)	2/131(2%)
gb ACL53739.1	unknown [Zea mays]	162155 bits(518)	1,48E-42	79/131(60%)	100/131(76%)	2/131(2%)
ref NP_001105981.1	lipoxygenase11 [Zea mays]	16254 bits(518)	1,51E-42	79/131(60%)	100/131(76%)	2/131(2%)
gb EAZ22076.1	hypothetical protein OsJ_05740 [Oryza sativa Japonica Group]	154451 bits(518)	1,23E-39	79/131(60%)	98/131(75%)	0/131(0%)
ref XP_004172829.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like, partial [Cucumis sativus]	168318 bits(518)	5,64E-45	78/132(59%)	98/132(74%)	0/132(0%)
ref XP_004142236.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Cucumis sativus]	168703 bits(518)	1,11E-44	78/132(59%)	98/132(74%)	0/132(0%)
dbj BAF01939.1	lipoxygenase [Arabidopsis thaliana]	16254 bits(518)	9,71E-46	76/130(58%)	95/130(73%)	0/130(0%)
ref XP_004308210.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Fragaria vesca subsp. vesca]	169859 bits(518)	4,16E-45	77/132(58%)	97/132(73%)	0/132(0%)
dbj BAD94917.1	lipoxygenase [Arabidopsis thaliana]	16254 bits(518)	6,32E-44	76/130(58%)	95/130(73%)	0/130(0%)
gb AAL32689.1	lipoxygenase AtLOX2 [Arabidopsis thaliana]	162155 bits(518)	2,42E-42	76/130(58%)	95/130(73%)	0/130(0%)
ref NP_566875.1	lipoxygenase 2 [Arabidopsis thaliana]	162155 bits(518)	2,57E-42	76/130(58%)	95/130(73%)	0/130(0%)
emb CAB72152.1	lipoxygenase AtLOX2 [Arabidopsis thaliana]	16177 bits(518)	2,72E-42	76/130(58%)	95/130(73%)	0/130(0%)
ref XP_003555620.1	PREDICTED: linoleate 13S-lipoxygenase 2-1, chloroplastic-like [Glycine max]	169859 bits(518)	4,33E-45	75/131(57%)	98/131(75%)	0/131(0%)
gb ACD43485.1	lipoxygenase 2 [Olea europaea]	166392 bits(518)	7,19E-44	74/131(56%)	102/131(78%)	0/131(0%)
ref XP_004171869.1	PREDICTED: LOW QUALITY PROTEIN: linoleate 13S-lipoxygenase 2-1, chloroplastic-like, partial [Cucumis sativus]	147517 bits(518)	7,88E-40	73/131(56%)	89/131(68%)	1/131(1%)
Contig16						
Sequence Id	Name	Score	Expect	Identities	Posities	Gaps
ref XP_002529997.1	cystathionine gamma-synthase, putative [Ricinus communis]	148288 bits(373)	5,40E-40	71/78(91%)	73/78(94%)	0/78(0%)

gb EMJ17201.1	hypothetical protein PRUPE_ppa004232mg [Prunus persica]	147132 bits(373)	3,91E-39	69/78(88%)	72/78(92%)	0/78(0%)
ref NP_001237572.1	cystathionine-gamma-synthase precursor [Glycine max]	147132 bits(373)	4,50E-39	68/78(87%)	72/78(92%)	0/78(0%)
pdb 1QGN A	Chain A, Cystathionine Gamma-Synthase From Nicotiana Tabacum	147132 bits(373)	1,84E-39	68/79(86%)	72/79(91%)	0/79(0%)
gb AAF74981.1 AF082891_1	cystathionine gamma-synthase isoform 1 [Solanum tuberosum]	146362 bits(373)	7,26E-39	68/79(86%)	73/79(92%)	0/79(0%)
ref NP_001234489.1	cystathionine gamma synthase [Solanum lycopersicum]	146362 bits(373)	7,66E-39	68/79(86%)	73/79(92%)	0/79(0%)
ref XP_004304284.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like [Fragaria vesca subsp. vesca]	145976 bits(373)	1,34E-38	67/78(86%)	72/78(92%)	0/78(0%)
emb CAA04772.2	cystathionine gamma synthase [Fragaria vesca]	145976 bits(373)	1,34E-38	67/78(86%)	72/78(92%)	0/78(0%)
ref XP_004136690.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like [Cucumis sativus]	145591 bits(373)	1,66E-38	67/78(86%)	73/78(94%)	0/78(0%)
ref XP_004159653.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like, partial [Cucumis sativus]	143665 bits(373)	2,08E-38	67/78(86%)	73/78(94%)	0/78(0%)
ref XP_003552541.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like [Glycine max]	14405 bits(373)	5,65E-38	67/78(86%)	71/78(91%)	0/78(0%)
gb AAT81295.1	cystathionine gamma synthase [Medicago sativa]	993673 bits(373)	3,84E-23	46/54(85%)	49/54(91%)	0/54(0%)
ref XP_003621273.1	Cystathionine gamma-synthase [Medicago truncatula]	138658 bits(373)	3,96E-37	63/75(84%)	69/75(92%)	0/75(0%)
gb AAC19395.1	cystathionine gamma-synthase [Mesembryanthemum crystallinum]	14251 bits(373)	1,95E-37	65/78(83%)	72/78(92%)	0/78(0%)
gb AGA60136.1	cystathionine gamma-synthase [Cicer arietinum]	141354 bits(373)	5,51E-37	65/78(83%)	70/78(90%)	0/78(0%)
gb AAZ32856.1	cystathionine-gamma-synthase [Medicago sativa]	115161 bits(373)	4,77E-30	54/65(83%)	59/65(91%)	0/65(0%)
emb CBI22246.3	unnamed protein product [Vitis vinifera]	141739 bits(373)	2,76E-37	65/79(82%)	71/79(90%)	0/79(0%)
ref XP_002283866.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like [Vitis vinifera]	141739 bits(373)	3,24E-37	65/79(82%)	71/79(90%)	0/79(0%)
ref XP_002882199.1	hypothetical protein ARALYDRAFT_477417 [Arabidopsis lyrata subsp. lyrata]	138272 bits(373)	7,94E-36	64/78(82%)	69/78(88%)	0/78(0%)
gb AAD31520.2 AF144102_1	cystathionine-gamma-synthase [Solanum tuberosum]	130568 bits(373)	5,22E-33	61/74(82%)	66/74(89%)	0/74(0%)
gb AAF74982.1 AF082892_1	cystathionine gamma-synthase isoform 2 [Solanum tuberosum]	143665 bits(373)	7,37E-38	64/79(81%)	73/79(92%)	0/79(0%)
gb ABI34092.1	cystathionine gamma-synthase [Medicago sativa]	139813 bits(373)	1,86E-36	63/78(81%)	70/78(90%)	0/78(0%)
ref XP_003621272.1	Cystathionine gamma-synthase [Medicago truncatula]	139428 bits(373)	2,28E-36	63/78(81%)	70/78(90%)	0/78(0%)
dbj BAH57133.1	AT3G01120 [Arabidopsis thaliana]	137117 bits(373)	4,23E-36	63/78(81%)	69/78(88%)	0/78(0%)
gb AAM13883.1	putative cystathionine gamma-synthase [Arabidopsis thaliana]	137502 bits(373)	1,39E-35	63/78(81%)	69/78(88%)	0/78(0%)

gb AAC25687.1	cystathionine gamma-synthase precursor [Arabidopsis thaliana]	137117 bits(373)	1,58E-35	63/78(81%)	69/78(88%)	0/78(0%)
ref NP_186761.1	cystathionine gamma-synthase [Arabidopsis thaliana]	137117 bits(373)	1,58E-35	63/78(81%)	69/78(88%)	0/78(0%)
emb CAA64383.1	cystathionine gamma-synthase [Arabidopsis thaliana]	137117 bits(373)	1,69E-35	63/78(81%)	69/78(88%)	0/78(0%)
gb AAC49574.1	similar to the metB gene product of Escherichia coli; cloned by functional complementation of a metB mutant strain of Escherichia coli LE392 [Arabidopsis thaliana]	137117 bits(373)	1,85E-35	63/78(81%)	69/78(88%)	0/78(0%)
ref XP_003618534.1	Cystathionine-gamma-synthase [Medicago truncatula]	133265 bits(373)	3,25E-34	61/78(78%)	68/78(87%)	0/78(0%)
gb AGA60137.1	cystathionine gamma-synthase [Lens culinaris]	138658 bits(373)	4,07E-36	61/80(76%)	72/80(90%)	0/80(0%)
ref XP_002323872.1	predicted protein [Populus trichocarpa]	132109 bits(373)	9,73E-34	60/79(76%)	68/79(86%)	0/79(0%)
gb ABR16369.1	unknown [Picea sitchensis]	130568 bits(373)	3,24E-33	59/78(76%)	70/78(90%)	0/78(0%)
ref XP_001761633.1	predicted protein [Physcomitrella patens subsp. patens]	127102 bits(373)	5,69E-32	59/78(76%)	66/78(85%)	0/78(0%)
gb ABK96139.1	unknown [Populus trichocarpa]	75485 bits(373)	4,55E-14	33/44(75%)	37/44(84%)	0/44(0%)
ref XP_001772529.1	predicted protein [Physcomitrella patens subsp. patens]	125946 bits(373)	1,59E-31	58/78(74%)	65/78(83%)	0/78(0%)
ref XP_002983120.1	hypothetical protein SELMODRAFT_155550 [Selaginella moellendorffii]	125561 bits(373)	1,67E-31	57/78(73%)	64/78(82%)	0/78(0%)
gb AAG38873.1 AF076495_1	cystathionine gamma-synthase [Oryza sativa]	111694 bits(373)	7,32E-27	53/73(73%)	60/73(82%)	0/73(0%)
ref XP_002965872.1	hypothetical protein SELMODRAFT_143363 [Selaginella moellendorffii]	124405 bits(373)	4,32E-31	56/78(72%)	64/78(82%)	0/78(0%)
gb ABF96221.1	Cys/Met metabolism PLP-dependent enzyme family protein, expressed [Oryza sativa Japonica Group]	119013 bits(373)	4,39E-29	56/79(71%)	63/79(80%)	0/79(0%)
gb AAL82522.1 AC084766_8	putative cystathionine gamma synthase [Oryza sativa Japonica Group]	117857 bits(373)	6,70E-29	56/79(71%)	63/79(80%)	0/79(0%)
ref NP_001130426.1	uncharacterized protein LOC100191523 [Zea mays]	117472 bits(373)	1,27E-28	56/79(71%)	63/79(80%)	0/79(0%)
gb AFW87953.1	hypothetical protein ZEAMMB73_590146 [Zea mays]	117087 bits(373)	1,66E-28	56/79(71%)	63/79(80%)	0/79(0%)
ref NP_001050220.1	Os03g0376100 [Oryza sativa Japonica Group]	115546 bits(373)	2,97E-28	56/79(71%)	63/79(80%)	0/79(0%)
gb AAB61348.1	cystathionine gamma-synthase [Zea mays]	115931 bits(373)	5,28E-28	56/79(71%)	62/79(78%)	0/79(0%)
gb EFN54424.1	hypothetical protein CHLNCRAFT_24830 [Chlorella variabilis]	119013 bits(373)	2,63E-29	55/79(70%)	63/79(80%)	0/79(0%)
ref NP_001104875.1	cystathionine gamma-synthase1 [Zea mays]	115931 bits(373)	5,73E-28	55/79(70%)	62/79(78%)	0/79(0%)
gb ABF96223.1	Cys/Met metabolism PLP-dependent enzyme family protein, expressed [Oryza sativa Japonica Group]	847297 bits(373)	9,99E-18	42/60(70%)	48/60(80%)	0/60(0%)

gb ABF96222.1	Cys/Met metabolism PLP-dependent enzyme family protein, expressed [Oryza sativa Japonica Group]	855001 bits(373)	1,36E-17	42/60(70%)	48/60(80%)	0/60(0%)
gb ACF83809.1	unknown [Zea mays]	851149 bits(373)	1,49E-17	42/60(70%)	48/60(80%)	0/60(0%)
gb AFW87952.1	hypothetical protein ZEAMMB73_590146 [Zea mays]	851149 bits(373)	1,75E-17	42/60(70%)	48/60(80%)	0/60(0%)
ref XP_002465206.1	hypothetical protein SORBIDRAFT_01g034130 [Sorghum bicolor]	851149 bits(373)	1,92E-17	42/60(70%)	48/60(80%)	0/60(0%)
tpg DAA45549.1	TPA: hypothetical protein ZEAMMB73_196241 [Zea mays]	828037 bits(373)	7,57E-17	40/57(70%)	46/57(81%)	0/57(0%)
gb AAG51206.1 AC051630_3	cystathionine gamma-synthase, putative; 4884-7220 [Arabidopsis thaliana]	114775 bits(373)	6,33E-28	53/77(69%)	62/77(81%)	0/77(0%)
ref NP_174600.1	putative cystathionine gamma-synthase [Arabidopsis thaliana]	114775 bits(373)	8,38E-28	53/77(69%)	62/77(81%)	0/77(0%)
ref XP_001419722.1	cystathionine gamma synthase [Ostreococcus lucimarinus CCE9901]	119013 bits(373)	2,61E-29	54/79(68%)	61/79(77%)	0/79(0%)
tpg DAA45550.1	TPA: hypothetical protein ZEAMMB73_196241 [Zea mays]	839593 bits(373)	3,00E-17	41/60(68%)	47/60(78%)	0/60(0%)
gb ACF82119.1	unknown [Zea mays]	835741 bits(373)	5,20E-17	41/60(68%)	47/60(78%)	0/60(0%)
emb CCO18050.1	cystathionine gamma-synthase [Bathycoccus prasinos]	120553 bits(373)	7,50E-30	53/79(67%)	63/79(80%)	0/79(0%)
gb EIE26818.1	cystathionine gamma-synthase [Coccomyxa subellipsoidea C-169]	115161 bits(373)	8,39E-28	53/79(67%)	62/79(78%)	0/79(0%)
ref XP_003057119.1	predicted protein [Micromonas pusilla CCMP1545]	108612 bits(373)	1,21E-25	49/78(63%)	59/78(76%)	0/78(0%)
ref XP_003561880.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like [Brachypodium distachyon]	108227 bits(373)	2,40E-25	49/78(63%)	60/78(77%)	0/78(0%)
ref XP_002506028.1	predicted protein [Micromonas sp. RCC299]	110153 bits(373)	2,96E-26	49/79(62%)	60/79(76%)	0/79(0%)
gb EKX31485.1	hypothetical protein GUIHDRAFT_159042 [Guillardia theta CCMP2712]	105145 bits(373)	2,65E-24	48/78(62%)	58/78(74%)	0/78(0%)
ref NP_001064539.1	Os10g0399200 [Oryza sativa Japonica Group]	103605 bits(373)	1,17E-23	48/78(62%)	58/78(74%)	0/78(0%)
gb EAY78387.1	hypothetical protein Osl_33474 [Oryza sativa Indica Group]	103219 bits(373)	1,21E-23	48/78(62%)	58/78(74%)	0/78(0%)
tpg DAA50062.1	TPA: hypothetical protein ZEAMMB73_527983 [Zea mays]	100523 bits(373)	1,48E-22	49/79(62%)	59/79(75%)	1/79(1%)
dbj BAJ93384.1	predicted protein [Hordeum vulgare subsp. vulgare]	997525 bits(373)	3,30E-22	49/79(62%)	59/79(75%)	1/79(1%)
ref XP_002953537.1	hypothetical protein VOLCADRAFT_75793 [Volvox carteri f. nagariensis]	101293 bits(373)	9,29E-23	46/75(61%)	55/75(73%)	0/75(0%)
tpg DAA50063.1	TPA: hypothetical protein ZEAMMB73_244612 [Zea mays]	970561 bits(373)	3,16E-21	48/79(61%)	58/79(73%)	1/79(1%)
ref XP_002467292.1	hypothetical protein SORBIDRAFT_01g023050 [Sorghum bicolor]	105145 bits(373)	3,60E-24	47/78(60%)	60/78(77%)	0/78(0%)

gb EAY78385.1	hypothetical protein Osl_33472 [Oryza sativa Indica Group]	102449 bits(373)	3,72E-24	47/78(60%)	57/78(73%)	0/78(0%)
gb EAZ15990.1	hypothetical protein OsJ_31434 [Oryza sativa Japonica Group]	101679 bits(373)	4,95E-23	47/78(60%)	57/78(73%)	0/78(0%)
gb EAY78389.1	hypothetical protein Osl_33476 [Oryza sativa Indica Group]	101679 bits(373)	6,51E-23	47/78(60%)	57/78(73%)	0/78(0%)
gb EAY78386.1	hypothetical protein Osl_33473 [Oryza sativa Indica Group]	101679 bits(373)	7,35E-23	47/78(60%)	57/78(73%)	0/78(0%)
ref NP_001064540.1	Os10g0399700 [Oryza sativa Japonica Group]	101679 bits(373)	8,56E-23	47/78(60%)	57/78(73%)	0/78(0%)
gb ABB47533.1	Cys/Met metabolism PLP-dependent enzyme family protein [Oryza sativa Japonica Group]	100523 bits(373)	1,17E-22	47/78(60%)	57/78(73%)	0/78(0%)
ref XP_001695742.1	cystathionine gamma-synthase [Chlamydomonas reinhardtii]	978265 bits(373)	1,70E-21	45/75(60%)	54/75(72%)	0/75(0%)
gb EEC80304.1	hypothetical protein Osl_22339 [Oryza sativa Indica Group]	101293 bits(373)	4,69E-23	45/78(58%)	58/78(74%)	0/78(0%)
gb EEE65425.1	hypothetical protein OsJ_20780 [Oryza sativa Japonica Group]	101293 bits(373)	4,78E-23	45/78(58%)	58/78(74%)	0/78(0%)
dbj BAD37853.1	putative O-succinylhomoserine (thiol)-lyase [Oryza sativa Japonica Group]	997525 bits(373)	1,95E-22	45/78(58%)	58/78(74%)	0/78(0%)
gb AAK50404.1 AC021891_5	Putative O-succinylhomoserine (thiol)-lyase (cystathionine gamma synthase) [Oryza sativa Japonica Group]	997525 bits(373)	4,01E-22	46/79(58%)	56/79(71%)	0/79(0%)
gb EME26477.1	cystathionine gamma-synthase [Galdieria sulphuraria]	916633 bits(373)	5,19E-19	43/76(57%)	54/76(71%)	0/76(0%)
ref XP_003557832.1	PREDICTED: cystathionine gamma-synthase, chloroplastic-like [Brachypodium distachyon]	113235 bits(373)	5,08E-27	59/113(52%)	73/113(65%)	0/113(0%)
dbj BAJ84799.1	predicted protein [Hordeum vulgare subsp. vulgare]	113235 bits(373)	6,10E-27	59/113(52%)	73/113(65%)	0/113(0%)
ref YP_003267806.1	cystathionine gamma-synthase [Haliangium ochraceum DSM 14365]	843445 bits(373)	6,13E-17	36/76(47%)	54/76(71%)	0/76(0%)
ref YP_001529282.1	cystathionine gamma-synthase [Desulfococcus oleovorans Hxd3]	781814 bits(373)	1,13E-14	36/77(47%)	47/77(61%)	0/77(0%)
ref ZP_09657302.1	Cys/Met metabolism pyridoxal-phosphate-dependent protein [Leptonema illini DSM 21528]	785666 bits(373)	7,58E-15	39/85(46%)	50/85(59%)	0/85(0%)
ref YP_001618247.1	cystathionine gamma-synthase [Sorangium cellulosum So ce56]	766406 bits(373)	3,66E-14	34/76(45%)	50/76(66%)	0/76(0%)
ref ZP_01912348.1	cystathionine gamma-synthase [Plesiocystis pacifica SIR-1]	762554 bits(373)	4,79E-14	36/80(45%)	49/80(61%)	6/80(8%)
ref YP_005441536.1	cystathionine gamma-lyase [Caldilinea aerophila DSM 14535 = NBRC 104270]	69707 bits(373)	1,10E-11	33/79(42%)	46/79(58%)	0/79(0%)
ref ZP_10868942.1	hypothetical protein IMCC14465_01760 [alpha proteobacterium IMCC14465]	712478 bits(373)	3,70E-12	32/78(41%)	48/78(62%)	0/78(0%)
ref ZP_23727941.1	methionine gamma-lyase [Rhodopirellula sp. SWK7]	689366 bits(373)	2,31E-11	32/79(41%)	46/79(58%)	0/79(0%)

ref ZP_23708016.1	methionine gamma-lyase [Rhodopirellula maiorica SM1]	685514 bits(373)	3,25E-11	32/79(41%)	46/79(58%)	0/79(0%)
ref YP_003372175.1	cystathionine gamma-synthase [Pirellula staleyi DSM 6068]	685514 bits(373)	3,31E-11	29/73(40%)	45/73(62%)	0/73(0%)
ref YP_297783.1	cystathionine beta-lyase [Ralstonia eutropha JMP134]	693218 bits(373)	1,60E-11	34/88(39%)	50/88(57%)	1/88(1%)
ref ZP_09997629.1	O-succinylhomoserine sulfhydrylase [Acidithiobacillus thiooxidans ATCC 19377]	689366 bits(373)	2,60E-11	32/82(39%)	49/82(60%)	0/82(0%)
ref ZP_07686650.1	cystathionine gamma-synthase [Oscillochloris trichoides DG-6]	812629 bits(373)	8,77E-16	44/120(37%)	67/120(56%)	3/120(2%)
ref YP_001635075.1	cystathionine gamma-synthase [Chloroflexus aurantiacus J-10-fl]	75485 bits(373)	1,02E-13	40/116(34%)	65/116(56%)	9/116(8%)
ref YP_002463279.1	cystathionine gamma-synthase [Chloroflexus aggregans DSM 9485]	762554 bits(373)	5,83E-14	37/114(32%)	63/114(55%)	3/114(3%)

Contig19

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
emb CAL07979.1	putative ubiquitin protein 2 [Platanus x acerifolia]	916633 bits(226)	2,90E-21	44/48(92%)	46/48(96%)	0/48(0%)
gb ADW83727.1	ubiquitin 2 [Musa acuminata AAA Group]	932041 bits(226)	2,87E-19	44/48(92%)	47/48(98%)	0/48(0%)
emb CBI37752.3	unnamed protein product [Vitis vinifera]	897373 bits(226)	8,59E-19	43/48(90%)	45/48(94%)	0/48(0%)
ref XP_002282473.2	PREDICTED: ubiquilin-1-like [Vitis vinifera]	901225 bits(226)	4,87E-18	43/48(90%)	45/48(94%)	0/48(0%)
ref XP_002328127.1	predicted protein [Populus trichocarpa]	893521 bits(226)	7,59E-18	43/48(90%)	45/48(94%)	0/48(0%)
emb CAN59899.1	hypothetical protein VITISV_002886 [Vitis vinifera]	893521 bits(226)	7,64E-18	43/48(90%)	44/48(92%)	0/48(0%)
ref XP_004290059.1	PREDICTED: uncharacterized protein LOC101299722 [Fragaria vesca subsp. vesca]	851149 bits(226)	2,28E-16	40/45(89%)	43/45(96%)	0/45(0%)
emb CBI38418.3	unnamed protein product [Vitis vinifera]	947449 bits(226)	1,83E-20	46/52(88%)	47/52(90%)	0/52(0%)
ref XP_002263194.2	PREDICTED: uncharacterized protein LOC100250759 [Vitis vinifera]	939745 bits(226)	1,23E-19	46/52(88%)	47/52(90%)	0/52(0%)
ref XP_004173937.1	PREDICTED: ubiquilin-2-like, partial [Cucumis sativus]	881965 bits(226)	1,60E-18	42/48(88%)	45/48(94%)	0/48(0%)
ref NP_001169509.1	uncharacterized protein LOC100383383 [Zea mays]	889669 bits(226)	6,45E-18	42/48(88%)	45/48(94%)	0/48(0%)

ref XP_002465921.1	hypothetical protein SORBIDRAFT_01g048260 [Sorghum bicolor]	889669 bits(226)	9,62E-18	42/48(88%)	45/48(94%)	0/48(0%)
tpg DAA43181.1	TPA: hypothetical protein ZEAMMB73_616463 [Zea mays]	889669 bits(226)	1,04E-17	42/48(88%)	45/48(94%)	0/48(0%)
tpg DAA43180.1	TPA: hypothetical protein ZEAMMB73_616463 [Zea mays]	889669 bits(226)	1,07E-17	42/48(88%)	45/48(94%)	0/48(0%)
ref XP_004144885.1	PREDICTED: ubiquilin-2-like isoform 2 [Cucumis sativus]	885817 bits(226)	1,59E-17	42/48(88%)	45/48(94%)	0/48(0%)
ref XP_004144884.1	PREDICTED: ubiquilin-2-like isoform 1 [Cucumis sativus]	885817 bits(226)	1,63E-17	42/48(88%)	45/48(94%)	0/48(0%)
ref XP_003517138.1	PREDICTED: ubiquilin-4-like [Glycine max]	939745 bits(226)	2,03E-19	45/52(87%)	48/52(92%)	0/52(0%)
ref XP_002521706.1	plant ubiquilin, putative [Ricinus communis]	928189 bits(226)	4,29E-19	45/52(87%)	48/52(92%)	0/52(0%)
emb CCI55438.1	PH01B031C15.21 [Phyllostachys edulis]	870409 bits(226)	2,92E-17	41/47(87%)	44/47(94%)	0/47(0%)
ref NP_001048862.1	Os03g0131300 [Oryza sativa Japonica Group]	870409 bits(226)	4,86E-17	41/47(87%)	44/47(94%)	0/47(0%)
gb EEC74450.1	hypothetical protein Osl_09858 [Oryza sativa Indica Group]	866557 bits(226)	5,67E-17	41/47(87%)	44/47(94%)	0/47(0%)
ref XP_003537681.1	PREDICTED: ubiquilin-1-like [Glycine max]	924337 bits(226)	6,44E-19	44/52(85%)	47/52(90%)	0/52(0%)
ref XP_003547896.1	PREDICTED: ubiquilin-1-like isoform 1 [Glycine max]	920485 bits(226)	7,45E-19	44/52(85%)	48/52(92%)	0/52(0%)
ref XP_003547897.1	PREDICTED: ubiquilin-1-like isoform 2 [Glycine max]	920485 bits(226)	8,20E-19	44/52(85%)	48/52(92%)	0/52(0%)
gb ADE77928.1	unknown [Picea sitchensis]	835741 bits(226)	3,34E-18	40/47(85%)	43/47(91%)	0/47(0%)
gb AEA50963.1	putative PDF1-interacting protein 1, partial [Gossypium barbadense]	862705 bits(226)	8,61E-17	41/48(85%)	44/48(92%)	0/48(0%)
gb EMJ00976.1	hypothetical protein PRUPE_ppa003711mg [Prunus persica]	858853 bits(226)	1,19E-16	40/47(85%)	44/47(94%)	0/47(0%)
ref NP_001150353.1	LOC100283983 [Zea mays]	878113 bits(226)	2,67E-17	41/49(84%)	45/49(92%)	0/49(0%)
ref XP_001771888.1	predicted protein [Physcomitrella patens subsp. patens]	831889 bits(226)	1,06E-15	38/45(84%)	42/45(93%)	0/45(0%)
ref XP_002985299.1	hypothetical protein SELMODRAFT_268972 [Selaginella moellendorffii]	824185 bits(226)	2,02E-15	38/45(84%)	43/45(96%)	0/45(0%)
ref XP_002970108.1	hypothetical protein SELMODRAFT_171198 [Selaginella moellendorffii]	824185 bits(226)	2,04E-15	38/45(84%)	43/45(96%)	0/45(0%)
gb ABK95433.1	unknown [Populus trichocarpa]	901225 bits(226)	6,54E-19	43/52(83%)	48/52(92%)	0/52(0%)
ref XP_003629155.1	Ubiquilin-1 [Medicago truncatula]	920485 bits(226)	9,42E-19	43/52(83%)	48/52(92%)	0/52(0%)
ref XP_003519686.1	PREDICTED: ubiquilin-1-like [Glycine max]	908929 bits(226)	2,35E-18	43/52(83%)	48/52(92%)	0/52(0%)
ref XP_002306425.1	predicted protein [Populus trichocarpa]	901225 bits(226)	3,57E-18	43/52(83%)	48/52(92%)	0/52(0%)
gb ADM18306.1	ubiquitin-like protein 1 [Gladiolus grandiflorus]	843445 bits(226)	5,94E-17	39/47(83%)	44/47(94%)	0/47(0%)

ref XP_001764740.1	predicted protein [Physcomitrella patens subsp. patens]	847297 bits(226)	3,34E-16	39/47(83%)	43/47(91%)	0/47(0%)
gb ACN54191.1	ubiquitin-like protein [Triticum aestivum]	843445 bits(226)	4,92E-16	40/48(83%)	43/48(90%)	0/48(0%)
gb ACN54190.1	ubiquitin-like protein [Triticum aestivum]	843445 bits(226)	5,02E-16	40/48(83%)	43/48(90%)	0/48(0%)
dbj BAJ95849.1	predicted protein [Hordeum vulgare subsp. vulgare]	843445 bits(226)	5,21E-16	40/48(83%)	43/48(90%)	0/48(0%)
dbj BAJ87273.1	predicted protein [Hordeum vulgare subsp. vulgare]	835741 bits(226)	6,81E-16	40/48(83%)	43/48(90%)	0/48(0%)
ref XP_003558890.1	PREDICTED: uncharacterized protein LOC100825963 isoform 1 [Brachypodium distachyon]	835741 bits(226)	7,10E-16	40/48(83%)	44/48(92%)	0/48(0%)
dbj BAK06552.1	predicted protein [Hordeum vulgare subsp. vulgare]	835741 bits(226)	7,14E-16	40/48(83%)	43/48(90%)	0/48(0%)
ref XP_003558891.1	PREDICTED: uncharacterized protein LOC100825963 isoform 2 [Brachypodium distachyon]	835741 bits(226)	7,51E-16	40/48(83%)	44/48(92%)	0/48(0%)
ref XP_003558889.1	PREDICTED: uncharacterized protein LOC100825650 [Brachypodium distachyon]	835741 bits(226)	8,14E-16	40/48(83%)	44/48(92%)	0/48(0%)
gb EMJ04822.1	hypothetical protein PRUPE_ppa003863mg [Prunus persica]	866557 bits(226)	7,50E-17	42/51(82%)	44/51(86%)	0/51(0%)
ref XP_003612325.1	Ubiquilin [Medicago truncatula]	897373 bits(226)	6,33E-18	42/52(81%)	47/52(90%)	0/52(0%)
ref XP_004234514.1	PREDICTED: ubiquilin-2-like [Solanum lycopersicum]	885817 bits(226)	1,38E-17	42/52(81%)	47/52(90%)	0/52(0%)
tpg DAA46118.1	TPA: hypothetical protein ZEAMMB73_654243 [Zea mays]	820333 bits(226)	1,56E-17	39/48(81%)	43/48(90%)	0/48(0%)
ref XP_004232558.1	PREDICTED: ubiquilin-1-like [Solanum lycopersicum]	885817 bits(226)	1,67E-17	42/52(81%)	47/52(90%)	0/52(0%)
ref XP_002465005.1	hypothetical protein SORBIDRAFT_01g030340 [Sorghum bicolor]	816481 bits(226)	3,95E-15	39/48(81%)	43/48(90%)	0/48(0%)
gb ACN33873.1	unknown [Zea mays]	812629 bits(226)	4,72E-15	39/48(81%)	43/48(90%)	0/48(0%)
ref NP_001168254.1	uncharacterized protein LOC100382017 [Zea mays]	808777 bits(226)	7,33E-15	39/48(81%)	43/48(90%)	0/48(0%)
ref NP_001148622.1	ubiquilin-1 [Zea mays]	79337 bits(226)	2,04E-14	38/47(81%)	42/47(89%)	0/47(0%)
gb ADW83726.1	ubiquitin 1 [Musa acuminata AAA Group]	77411 bits(226)	9,59E-14	36/45(80%)	42/45(93%)	0/45(0%)
ref XP_003629154.1	Ubiquilin [Medicago truncatula]	750998 bits(226)	5,12E-13	36/46(78%)	41/46(89%)	0/46(0%)
ref XP_004146171.1	PREDICTED: ubiquilin-1-like [Cucumis sativus]	851149 bits(226)	2,39E-16	40/52(77%)	45/52(87%)	0/52(0%)
ref XP_004287736.1	PREDICTED: uncharacterized protein LOC101302002 [Fragaria vesca subsp. vesca]	843445 bits(226)	3,61E-16	40/52(77%)	44/52(85%)	0/52(0%)
ref NP_001065193.1	Os10g0542200 [Oryza sativa Japonica Group]	77411 bits(226)	1,05E-13	37/48(77%)	43/48(90%)	0/48(0%)
gb EAY79386.1	hypothetical protein Osl_34513 [Oryza sativa Indica Group]	77411 bits(226)	1,05E-13	37/48(77%)	43/48(90%)	0/48(0%)
gb EEE51337.1	hypothetical protein OsJ_32327 [Oryza sativa Japonica Group]	77411 bits(226)	1,06E-13	37/48(77%)	43/48(90%)	0/48(0%)

gb ABX75362.1	hypothetical protein LBL8 [Panax quinquefolius]	739442 bits(226)	1,90E-13	36/48(75%)	43/48(90%)	0/48(0%)
ref XP_002886102.1	ubiquitin family protein [Arabidopsis lyrata subsp. lyrata]	743294 bits(226)	1,06E-12	36/48(75%)	41/48(85%)	0/48(0%)
ref XP_002886103.1	ubiquitin family protein [Arabidopsis lyrata subsp. lyrata]	743294 bits(226)	1,11E-12	36/48(75%)	41/48(85%)	0/48(0%)
ref NP_179311.1	ubiquitin-like protein [Arabidopsis thaliana]	743294 bits(226)	1,13E-12	36/48(75%)	41/48(85%)	0/48(0%)
gb AAK59681.1	putative ubiquitin protein [Arabidopsis thaliana]	743294 bits(226)	1,15E-12	36/48(75%)	41/48(85%)	0/48(0%)
gb ABX75361.1	hypothetical protein LBL7 [Panax quinquefolius]	720182 bits(226)	1,13E-12	35/48(73%)	42/48(88%)	0/48(0%)
ref NP_565407.1	ubiquilin [Arabidopsis thaliana]	727886 bits(226)	3,62E-12	35/48(73%)	41/48(85%)	0/48(0%)
ref XP_001759037.1	predicted protein [Physcomitrella patens subsp. patens]	689366 bits(226)	6,08E-11	32/47(68%)	39/47(83%)	0/47(0%)

Contig20

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb AEK31196.1	miraculin-like protein 2 [Citrus limonia]	935893 bits(269)	9,85E-21	42/68(62%)	50/68(74%)	2/68(3%)
gb AEK31197.1	miraculin-like protein 2 [Citrus aurantiifolia]	935893 bits(269)	1,03E-20	42/68(62%)	50/68(74%)	2/68(3%)
gb AEK31199.1	miraculin-like protein 2 [Aegle marmelos]	924337 bits(269)	2,39E-20	43/69(62%)	51/69(74%)	3/69(4%)
gb AEK31201.1	miraculin-like protein 2 [Murraya paniculata]	893521 bits(269)	3,23E-19	42/68(62%)	50/68(74%)	3/68(4%)
gb AEK31202.1	miraculin-like protein 2 [Murraya koenigii]	885817 bits(269)	7,26E-19	42/68(62%)	48/68(71%)	2/68(3%)
dbj BAE79511.1	miraculin-like protein 2 [Citrus jambhiri]	108227 bits(269)	9,73E-26	52/85(61%)	61/85(72%)	4/85(5%)
gb AFR54471.1	miraculin-like protein 2 [Citrus japonica]	107071 bits(269)	2,01E-25	52/85(61%)	60/85(71%)	4/85(5%)
gb AEO27900.1	miraculin-like protein 2 [Lagenaria siceraria]	847297 bits(269)	1,66E-17	40/66(61%)	47/66(71%)	3/66(5%)
gb AEO27899.1	miraculin-like protein [Citrullus lanatus]	881965 bits(269)	1,02E-18	41/68(60%)	49/68(72%)	3/68(4%)
gb AEO27901.1	miraculin-like protein 2 [Citrus aurantium]	912781 bits(269)	6,51E-20	43/73(59%)	49/73(67%)	2/73(3%)
gb AEK31190.1	miraculin-like protein 1 [Citrus limonia]	808777 bits(269)	4,27E-16	39/66(59%)	47/66(71%)	3/66(5%)
gb AEK31191.1	miraculin-like protein 1 [Citrus aurantiifolia]	816481 bits(269)	2,01E-16	39/67(58%)	48/67(72%)	3/67(4%)
gb AFL91233.1	Kunitz-like protease inhibitor, partial [Helianthus annuus]	808777 bits(269)	4,22E-16	39/69(57%)	48/69(70%)	2/69(3%)

gb AFL91232.1	Kunitz-like protease inhibitor, partial [<i>Helianthus annuus</i>]	804925 bits(269)	5,27E-16	39/69(57%)	48/69(70%)	2/69(3%)
gb AEK31192.1	miraculin-like protein 1 [<i>Citrus maxima</i>]	777962 bits(269)	5,38E-15	38/67(57%)	48/67(72%)	3/67(4%)
gb AAG38519.1 AF283534_1	miraculin-like protein 3 [<i>Citrus x paradisi</i>]	935893 bits(269)	1,88E-20	38/68(56%)	50/68(74%)	0/68(0%)
gb AEK26933.1	Kunitz-type trypsin inhibitor B2 [<i>Populus nigra</i>]	73559 bits(269)	2,28E-13	36/66(55%)	44/66(67%)	3/66(5%)
gb ACS92509.1	Kunitz-type protease inhibitor KPI-D2.3 [<i>Populus trichocarpa</i>]	724034 bits(269)	5,83E-13	36/66(55%)	43/66(65%)	3/66(5%)
gb ACS92483.1	Kunitz-type protease inhibitor KPI-D2.2 [<i>Populus trichocarpa</i>]	724034 bits(269)	6,11E-13	36/66(55%)	43/66(65%)	3/66(5%)
ref XP_002305032.1	predicted protein [<i>Populus trichocarpa</i>]	720182 bits(269)	6,89E-13	36/66(55%)	43/66(65%)	3/66(5%)
dbj BAA82840.1	miraculin homologue [<i>Youngia japonica</i>]	693218 bits(269)	3,35E-12	30/55(55%)	38/55(69%)	1/55(2%)
gb ABG47463.1	vegetative storage protein [<i>Litchi chinensis</i>]	962857 bits(269)	2,28E-21	44/82(54%)	56/82(68%)	3/82(4%)
gb ABL67650.1	putative miraculin-like protein 2 [<i>Citrus hybrid cultivar</i>]	935893 bits(269)	2,19E-20	44/82(54%)	56/82(68%)	3/82(4%)
gb EMJ12410.1	hypothetical protein PRUPE_ppa011496mg [<i>Prunus persica</i>]	835741 bits(269)	8,90E-17	36/67(54%)	47/67(70%)	1/67(1%)
gb AEK31200.1	miraculin-like protein 2 [<i>Citrus reticulata</i>]	79337 bits(269)	1,52E-15	37/69(54%)	46/69(67%)	3/69(4%)
ref XP_002266430.1	PREDICTED: miraculin [<i>Vitis vinifera</i>]	79337 bits(269)	2,08E-15	36/67(54%)	45/67(67%)	1/67(1%)
gb AFL91228.1	Kunitz-like protease inhibitor, partial [<i>Helianthus annuus</i>]	762554 bits(269)	2,08E-14	37/68(54%)	45/68(66%)	2/68(3%)
gb AAL85643.1	trypsin inhibitor [<i>Theobroma angustifolium</i>]	71633 bits(269)	7,59E-13	29/54(54%)	36/54(67%)	0/54(0%)
gb AAL85644.1	trypsin inhibitor [<i>Theobroma subincanum</i>]	704774 bits(269)	1,50E-12	29/54(54%)	36/54(67%)	0/54(0%)
gb AAL85645.1	trypsin inhibitor [<i>Theobroma grandiflorum</i>]	704774 bits(269)	1,52E-12	29/54(54%)	36/54(67%)	0/54(0%)
gb AAL85648.1	trypsin inhibitor [<i>Theobroma obovatum</i>]	704774 bits(269)	1,55E-12	29/54(54%)	36/54(67%)	0/54(0%)
gb AAL85652.1	trypsin inhibitor [<i>Theobroma mammosum</i>]	700922 bits(269)	2,01E-12	29/54(54%)	36/54(67%)	0/54(0%)
gb AAL85640.1	trypsin inhibitor [<i>Herrania albiflora</i>]	681662 bits(269)	1,10E-11	29/54(54%)	35/54(65%)	0/54(0%)
gb AFL91230.1	Kunitz-like protease inhibitor, partial [<i>Helianthus annuus</i>]	804925 bits(269)	4,61E-16	43/81(53%)	53/81(65%)	3/81(4%)
gb AFL91227.1	Kunitz-like protease inhibitor, partial [<i>Helianthus annuus</i>]	808777 bits(269)	4,62E-16	43/81(53%)	53/81(65%)	3/81(4%)
gb AFL91231.1	Kunitz-like protease inhibitor, partial [<i>Helianthus annuus</i>]	808777 bits(269)	4,76E-16	43/81(53%)	53/81(65%)	3/81(4%)
gb AFL91229.1	Kunitz-like protease inhibitor, partial [<i>Helianthus annuus</i>]	804925 bits(269)	4,77E-16	43/81(53%)	53/81(65%)	3/81(4%)
emb CAN65022.1	hypothetical protein VITISV_027379 [<i>Vitis vinifera</i>]	885817 bits(269)	1,08E-18	41/79(52%)	53/79(67%)	2/79(3%)

gb AFL91234.1	Kunitz-like protease inhibitor, partial [Helianthus annuus]	801073 bits(269)	7,20E-16	42/81(52%)	53/81(65%)	3/81(4%)
gb AFL91226.1	Kunitz-like protease inhibitor, partial [Helianthus annuus]	797221 bits(269)	1,02E-15	42/81(52%)	53/81(65%)	3/81(4%)
sp P13087.3 MIRA_RICDU	RecName: Full=Miraculin; Short=MIR; Flags: Precursor	720182 bits(269)	9,92E-13	36/69(52%)	46/69(67%)	2/69(3%)
ref XP_003620175.1	Pathogen-inducible trypsin-inhibitor-like protein [Medicago truncatula]	69707 bits(269)	4,89E-12	34/66(52%)	44/66(67%)	2/66(3%)
ref XP_003620174.1	Pathogen-inducible trypsin-inhibitor-like protein [Medicago truncatula]	69707 bits(269)	4,94E-12	34/66(52%)	44/66(67%)	2/66(3%)
gb AAL85650.1	trypsin inhibitor [Theobroma mammosum]	689366 bits(269)	6,25E-12	28/54(52%)	35/54(65%)	0/54(0%)
gb AAL85651.1	trypsin inhibitor [Theobroma mammosum]	689366 bits(269)	6,62E-12	28/54(52%)	35/54(65%)	0/54(0%)
emb CAN81015.1	hypothetical protein VITISV_025776 [Vitis vinifera]	847297 bits(269)	3,25E-17	40/79(51%)	52/79(66%)	2/79(3%)
ref XP_002266302.2	PREDICTED: miraculin-like, partial [Vitis vinifera]	831889 bits(269)	4,94E-17	40/79(51%)	52/79(66%)	2/79(3%)
gb ADD51186.1	tumor-related protein [Vitis cinerea var. helleri x Vitis riparia]	839593 bits(269)	5,51E-17	40/79(51%)	52/79(66%)	2/79(3%)
ref XP_002270111.1	PREDICTED: miraculin [Vitis vinifera]	839593 bits(269)	6,14E-17	40/79(51%)	52/79(66%)	2/79(3%)
emb CBI35471.3	unnamed protein product [Vitis vinifera]	843445 bits(269)	1,03E-15	40/79(51%)	52/79(66%)	2/79(3%)
ref XP_004299457.1	PREDICTED: miraculin-like [Fragaria vesca subsp. vesca]	797221 bits(269)	2,17E-15	35/68(51%)	46/68(68%)	1/68(1%)
ref XP_002335653.1	predicted protein [Populus trichocarpa]	77411 bits(269)	1,19E-14	35/69(51%)	42/69(61%)	2/69(3%)
gb ACS92516.1	Kunitz-type protease inhibitor KPI-D12 [Populus trichocarpa x Populus deltoides]	77411 bits(269)	1,19E-14	35/69(51%)	43/69(62%)	2/69(3%)
ref XP_002331924.1	predicted protein [Populus trichocarpa]	77411 bits(269)	1,20E-14	35/69(51%)	42/69(61%)	2/69(3%)
dbj BAA82843.1	miraculin homologue [Solanum melongena]	750998 bits(269)	4,52E-14	34/67(51%)	44/67(66%)	1/67(1%)
ref XP_003534100.1	PREDICTED: miraculin-like [Glycine max]	724034 bits(269)	7,36E-13	34/67(51%)	45/67(67%)	2/67(3%)
gb ACJ51124.1	talisin [Talisia esculenta]	855001 bits(269)	1,57E-17	41/82(50%)	52/82(63%)	4/82(5%)
ref XP_002265965.1	PREDICTED: miraculin [Vitis vinifera]	808777 bits(269)	5,89E-16	39/78(50%)	48/78(62%)	1/78(1%)
sp P32765.1 ASP_THECC	RecName: Full=21 kDa seed protein; Flags: Precursor	727886 bits(269)	4,59E-13	34/68(50%)	43/68(63%)	1/68(1%)
prf 1802409A	albumin	727886 bits(269)	4,86E-13	34/68(50%)	43/68(63%)	1/68(1%)
dbj BAA82842.1	miraculin homologue [Taraxacum officinale]	69707 bits(269)	2,09E-12	32/64(50%)	38/64(59%)	1/64(2%)
emb CAZ64526.1	Kunitz trypsin inhibitor [Populus trichocarpa]	69707 bits(269)	3,27E-12	31/62(50%)	37/62(60%)	0/62(0%)
gb ACO40082.1	kunitz trypsin inhibitor 4 [Populus balsamifera]	689366 bits(269)	6,96E-12	31/62(50%)	36/62(58%)	0/62(0%)

gb ACO40119.1	kunitz trypsin inhibitor 4 [Populus balsamifera]	689366 bits(269)	7,03E-12	31/62(50%)	36/62(58%)	0/62(0%)
gb ACO40144.1	kunitz trypsin inhibitor 4 [Populus balsamifera]	689366 bits(269)	7,23E-12	31/62(50%)	36/62(58%)	0/62(0%)
emb CBI35464.3	unnamed protein product [Vitis vinifera]	816481 bits(269)	4,25E-16	39/79(49%)	51/79(65%)	2/79(3%)
emb CAN75396.1	hypothetical protein VITISV_028636 [Vitis vinifera]	812629 bits(269)	4,33E-16	39/79(49%)	51/79(65%)	2/79(3%)
gb ACV52580.1	endopeptidase inhibitor [Nicotiana benthamiana]	731738 bits(269)	6,85E-14	33/67(49%)	44/67(66%)	1/67(1%)
ref XP_003548288.1	PREDICTED: miraculin-like [Glycine max]	750998 bits(269)	6,94E-14	33/67(49%)	45/67(67%)	1/67(1%)
ref NP_173228.1	kunitz type trypsin and protease inhibitor domain-containing protein [Arabidopsis thaliana]	71633 bits(269)	9,54E-13	32/65(49%)	43/65(66%)	1/65(2%)
ref XP_004301214.1	PREDICTED: miraculin-like [Fragaria vesca subsp. vesca]	712478 bits(269)	1,59E-12	33/67(49%)	43/67(64%)	1/67(1%)
gb ACL12055.1	Kunitz trypsin inhibitor [Nicotiana tabacum]	689366 bits(269)	9,24E-12	32/65(49%)	41/65(63%)	1/65(2%)
pdb 3IIR A	Chain A, Crystal Structure Of Miraculin Like Protein From Seeds Of Mu Koenigii	885817 bits(269)	8,26E-19	47/97(48%)	61/97(63%)	12/97(12%)
gb ACS37303.1	trypsin inhibitor [Murraya koenigii]	889669 bits(269)	8,35E-19	47/97(48%)	61/97(63%)	12/97(12%)
ref XP_003534080.1	PREDICTED: miraculin-like [Glycine max]	75485 bits(269)	5,25E-14	32/67(48%)	46/67(69%)	1/67(1%)
ref XP_003534078.1	PREDICTED: miraculin-like [Glycine max]	747146 bits(269)	9,01E-14	32/67(48%)	46/67(69%)	1/67(1%)
ref NP_001237312.1	uncharacterized protein LOC100527782 precursor [Glycine max]	747146 bits(269)	1,00E-13	32/67(48%)	46/67(69%)	1/67(1%)
gb AAV41233.1	putative 21 kDa trypsin inhibitor [Theobroma bicolor]	739442 bits(269)	2,13E-13	34/71(48%)	44/71(62%)	1/71(1%)
ref XP_003534079.1	PREDICTED: miraculin-like [Glycine max]	71633 bits(269)	9,86E-13	32/67(48%)	46/67(69%)	1/67(1%)
ref XP_004234331.1	PREDICTED: miraculin-like [Solanum lycopersicum]	785666 bits(269)	4,51E-15	37/79(47%)	51/79(65%)	2/79(3%)
dbj BAA05474.1	tumor-related protein [Nicotiana glauca x Nicotiana langsdorffii]	743294 bits(269)	3,89E-14	37/79(47%)	50/79(63%)	2/79(3%)
ref NP_001238098.1	uncharacterized protein LOC100306134 precursor [Glycine max]	708626 bits(269)	1,93E-12	33/70(47%)	46/70(66%)	2/70(3%)
dbj BAA82841.1	miraculin homologue [Youngia japonica]	693218 bits(269)	3,00E-12	30/64(47%)	41/64(64%)	1/64(2%)
gb AEK26927.1	Kunitz-type trypsin inhibitor B4 [Populus nigra]	700922 bits(269)	4,43E-12	36/76(47%)	44/76(58%)	3/76(4%)
gb ADW95380.1	Kunitz-type trypsin inhibitor [Populus nigra]	69707 bits(269)	4,98E-12	36/76(47%)	44/76(58%)	3/76(4%)
ref XP_002305034.1	predicted protein [Populus trichocarpa]	693218 bits(269)	6,26E-12	36/76(47%)	44/76(58%)	3/76(4%)
gb AAV41234.1	putative 21 kDa trypsin inhibitor [Theobroma microcarpum]	77411 bits(269)	1,50E-14	33/71(46%)	46/71(65%)	1/71(1%)
gb AAQ96377.1	miraculin-like protein [Solanum palustre]	762554 bits(269)	3,40E-14	36/79(46%)	51/79(65%)	2/79(3%)

gb AAC49969.1	tumor-related protein [Nicotiana tabacum]	75485 bits(269)	4,76E-14	37/80(46%)	51/80(64%)	2/80(2%)
gb ADK62529.1	miraculin-like protein [Nicotiana benthamiana]	750998 bits(269)	7,53E-14	36/79(46%)	50/79(63%)	2/79(3%)
ref XP_003634323.1	PREDICTED: miraculin-like [Vitis vinifera]	727886 bits(269)	4,71E-13	36/79(46%)	48/79(61%)	2/79(3%)
ref XP_002305033.1	predicted protein [Populus trichocarpa]	704774 bits(269)	3,12E-12	35/77(45%)	45/77(58%)	1/77(1%)
gb AEK26926.1	Kunitz-type trypsin inhibitor B1 [Populus nigra]	693218 bits(269)	7,87E-12	35/77(45%)	45/77(58%)	1/77(1%)
ref XP_004137534.1	PREDICTED: miraculin-like [Cucumis sativus]	75485 bits(269)	5,97E-14	35/79(44%)	48/79(61%)	2/79(3%)
gb ACL78790.1	putative miraculin-like protein 2 [Citrus unshiu]	720182 bits(269)	1,14E-12	38/87(44%)	46/87(53%)	6/87(7%)
gb AAG38518.1 AF283533_1	miraculin-like protein 2 [Citrus x paradisi]	71633 bits(269)	1,43E-12	38/87(44%)	46/87(53%)	6/87(7%)
ref XP_003548289.1	PREDICTED: miraculin-like [Glycine max]	704774 bits(269)	2,84E-12	30/68(44%)	43/68(63%)	1/68(1%)
ref XP_002890252.1	trypsin and protease inhibitor family protein [Arabidopsis lyrata subsp. lyrata]	693218 bits(269)	5,80E-12	34/79(43%)	48/79(61%)	2/79(3%)
ref XP_004137639.1	PREDICTED: miraculin-like [Cucumis sativus]	77411 bits(269)	1,07E-14	34/80(42%)	52/80(65%)	2/80(2%)
ref NP_001237333.1	uncharacterized protein LOC100500648 precursor [Glycine max]	69707 bits(269)	5,74E-12	30/73(41%)	43/73(59%)	1/73(1%)

Contig22

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_003542118.1	PREDICTED: LOW QUALITY PROTEIN: glucan endo-1,3-beta-glucosidase 14-like [Glycine max]	708626 bits(228)	1,53E-11	32/34(94%)	33/34(97%)	0/34(0%)
ref XP_002527196.1	Glucan endo-1,3-beta-glucosidase precursor, putative [Ricinus communis]	643142 bits(228)	4,20E-10	29/32(91%)	31/32(97%)	0/32(0%)
ref XP_002308978.1	predicted protein [Populus trichocarpa]	654698 bits(228)	9,04E-10	32/37(86%)	32/37(86%)	0/37(0%)
gb EMJ18529.1	hypothetical protein PRUPE_ppa020393mg, partial [Prunus persica]	766406 bits(228)	1,41E-13	38/46(83%)	42/46(91%)	1/46(2%)
ref XP_002327527.1	predicted protein [Populus trichocarpa]	739442 bits(228)	1,30E-12	37/45(82%)	40/45(89%)	1/45(2%)
ref XP_003546926.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Glycine max]	812629 bits(228)	3,75E-15	39/49(80%)	42/49(86%)	0/49(0%)
gb ABI49503.1	Glycosyl hydrolases family 17 protein [Solanum demissum]	69707 bits(228)	7,15E-11	36/45(80%)	38/45(84%)	2/45(4%)

gb EMJ05555.1	hypothetical protein PRUPE_ppa007158mg [Prunus persica]	65855 bits(228)	8,80E-10	31/39(79%)	34/39(87%)	0/39(0%)
ref XP_002517634.1	Glucan endo-1,3-beta-glucosidase, acidic isoform GL153 precursor, putative [Ricinus communis]	720182 bits(228)	6,24E-12	35/45(78%)	37/45(82%)	0/45(0%)
ref XP_002299750.1	predicted protein [Populus trichocarpa]	712478 bits(228)	1,28E-11	35/45(78%)	36/45(80%)	0/45(0%)
ref XP_004242119.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Solanum lycopersicum]	662402 bits(228)	6,35E-10	35/46(76%)	37/46(80%)	1/46(2%)
ref XP_004251202.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Solanum lycopersicum]	666254 bits(228)	5,43E-10	34/46(74%)	37/46(80%)	1/46(2%)
ref XP_003604379.1	Glucan endo-1,3-beta-glucosidase [Medicago truncatula]	727886 bits(228)	4,32E-12	35/49(71%)	39/49(80%)	0/49(0%)
ref XP_004242775.1	PREDICTED: glucan endo-1,3-beta-glucosidase 11-like [Solanum lycopersicum]	666254 bits(228)	4,08E-10	32/45(71%)	37/45(82%)	0/45(0%)
ref XP_002265779.2	PREDICTED: glucan endo-1,3-beta-glucosidase 11-like [Vitis vinifera]	828037 bits(228)	1,39E-15	49/72(68%)	52/72(72%)	8/72(11%)
emb CBI36326.3	unnamed protein product [Vitis vinifera]	828037 bits(228)	1,41E-15	49/72(68%)	52/72(72%)	8/72(11%)
ref XP_004305510.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Fragaria vesca subsp. vesca]	924337 bits(228)	5,17E-19	48/73(66%)	55/73(75%)	4/73(5%)
emb CAN70046.1	hypothetical protein VITISV_032962 [Vitis vinifera]	743294 bits(228)	1,54E-12	44/68(65%)	48/68(71%)	7/68(10%)
ref XP_003529808.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Glycine max]	831889 bits(228)	1,03E-15	44/69(64%)	51/69(74%)	4/69(6%)
ref XP_003533065.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like, partial [Glycine max]	855001 bits(228)	1,72E-16	43/68(63%)	49/68(72%)	3/68(4%)
ref XP_004152347.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Cucumis sativus]	747146 bits(228)	9,59E-13	44/70(63%)	48/70(69%)	4/70(6%)
ref XP_004139302.1	PREDICTED: glucan endo-1,3-beta-glucosidase 11-like [Cucumis sativus]	693218 bits(228)	3,28E-11	44/70(63%)	49/70(70%)	1/70(1%)
ref XP_004159384.1	PREDICTED: glucan endo-1,3-beta-glucosidase 11-like [Cucumis sativus]	689366 bits(228)	7,51E-11	44/70(63%)	49/70(70%)	1/70(1%)
emb CBI28222.3	unnamed protein product [Vitis vinifera]	820333 bits(228)	1,62E-15	43/73(59%)	48/73(66%)	3/73(4%)
ref XP_002282272.1	PREDICTED: glucan endo-1,3-beta-glucosidase 11-like [Vitis vinifera]	816481 bits(228)	3,95E-15	43/73(59%)	48/73(66%)	3/73(4%)
ref NP_001058028.1	Os06g0607000 [Oryza sativa Japonica Group]	670106 bits(228)	4,62E-10	31/53(58%)	38/53(72%)	0/53(0%)
ref XP_002453496.1	hypothetical protein SORBIDRAFT_04g006850 [Sorghum bicolor]	700922 bits(228)	3,49E-11	40/72(56%)	49/72(68%)	4/72(6%)
ref XP_004287297.1	PREDICTED: glucan endo-1,3-beta-glucosidase 14-like [Fragaria vesca subsp. vesca]	666254 bits(228)	4,42E-10	40/72(56%)	44/72(61%)	2/72(3%)

Contig26						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_004237852.1	PREDICTED: laccase-14-like [Solanum lycopersicum]	192586 bits(478)	2,87E-54	84/102(82%)	94/102(92%)	0/102(0%)
emb CAN75766.1	hypothetical protein VITISV_032561 [Vitis vinifera]	188734 bits(478)	3,55E-56	81/100(81%)	90/100(90%)	0/100(0%)
ref XP_003634549.1	PREDICTED: LOW QUALITY PROTEIN: putative laccase-9-like [Vitis vinifera]	194512 bits(478)	1,39E-54	81/100(81%)	90/100(90%)	0/100(0%)
ref XP_002267373.1	PREDICTED: putative laccase-9 [Vitis vinifera]	190274 bits(478)	1,60E-53	80/100(80%)	92/100(92%)	0/100(0%)
ref XP_003634919.1	PREDICTED: laccase-14-like [Vitis vinifera]	1922 bits(478)	1,83E-53	80/100(80%)	91/100(91%)	0/100(0%)
emb CAN77307.1	hypothetical protein VITISV_000786 [Vitis vinifera]	185652 bits(478)	1,04E-51	80/100(80%)	91/100(91%)	0/100(0%)
emb CBI18288.3	unnamed protein product [Vitis vinifera]	189889 bits(478)	4,14E-53	79/100(79%)	90/100(90%)	0/100(0%)
ref XP_002268847.2	PREDICTED: putative laccase-9-like [Vitis vinifera]	190274 bits(478)	4,57E-53	79/100(79%)	90/100(90%)	0/100(0%)
ref XP_003634547.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	187193 bits(478)	2,63E-52	79/100(79%)	90/100(90%)	0/100(0%)
ref XP_002266043.1	PREDICTED: putative laccase-9 [Vitis vinifera]	186808 bits(478)	3,77E-52	79/100(79%)	90/100(90%)	0/100(0%)
ref XP_004288102.1	PREDICTED: laccase-15-like [Fragaria vesca subsp. vesca]	165236 bits(478)	2,34E-44	73/92(79%)	79/92(86%)	0/92(0%)
emb CAN61420.1	hypothetical protein VITISV_023544 [Vitis vinifera]	195282 bits(478)	1,28E-53	78/100(78%)	89/100(89%)	0/100(0%)
emb CBI18286.3	unnamed protein product [Vitis vinifera]	19143 bits(478)	1,99E-52	76/98(78%)	89/98(91%)	0/98(0%)
emb CBI18289.3	unnamed protein product [Vitis vinifera]	172555 bits(478)	1,65E-51	76/98(78%)	87/98(89%)	0/98(0%)
ref XP_002266971.1	PREDICTED: laccase-14 [Vitis vinifera]	184882 bits(478)	1,94E-51	78/100(78%)	89/100(89%)	0/100(0%)
ref XP_002266331.1	PREDICTED: laccase-14-like [Vitis vinifera]	184496 bits(478)	2,13E-51	78/100(78%)	90/100(90%)	0/100(0%)
gb EMJ21848.1	hypothetical protein PRUPE_ppa003529mg [Prunus persica]	166777 bits(478)	7,34E-45	72/92(78%)	80/92(87%)	0/92(0%)
emb CAN82952.1	hypothetical protein VITISV_026918 [Vitis vinifera]	1818 bits(478)	1,26E-50	77/100(77%)	87/100(87%)	0/100(0%)
ref XP_002267700.1	PREDICTED: laccase-14-like [Vitis vinifera]	182185 bits(478)	1,62E-50	77/100(77%)	87/100(87%)	0/100(0%)
dbj BAK07490.1	predicted protein [Hordeum vulgare subsp. vulgare]	181415 bits(478)	5,55E-50	77/101(76%)	85/101(84%)	0/101(0%)
ref XP_002268802.2	PREDICTED: laccase-14-like [Vitis vinifera]	176792 bits(478)	1,83E-48	76/100(76%)	85/100(85%)	0/100(0%)
emb CAN72660.1	hypothetical protein VITISV_042718 [Vitis vinifera]	187578 bits(478)	9,01E-53	81/109(74%)	95/109(87%)	1/109(1%)

ref XP_002266921.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	186808 bits(478)	2,15E-52	81/109(74%)	95/109(87%)	1/109(1%)
emb CAN68975.1	hypothetical protein VITISV_035998 [Vitis vinifera]	186422 bits(478)	2,36E-52	81/109(74%)	95/109(87%)	1/109(1%)
ref XP_003634552.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	186808 bits(478)	2,54E-52	81/109(74%)	95/109(87%)	1/109(1%)
ref XP_003634922.1	PREDICTED: LOW QUALITY PROTEIN: putative laccase-9-like [Vitis vinifera]	166392 bits(478)	8,39E-45	74/100(74%)	84/100(84%)	0/100(0%)
emb CBI18287.3	unnamed protein product [Vitis vinifera]	183726 bits(478)	4,41E-53	80/109(73%)	93/109(85%)	1/109(1%)
ref XP_002265713.1	PREDICTED: putative laccase-9 [Vitis vinifera]	18257 bits(478)	1,42E-50	80/109(73%)	93/109(85%)	1/109(1%)
ref XP_003634924.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	175252 bits(478)	2,91E-48	74/101(73%)	87/101(86%)	0/101(0%)
emb CAN63600.1	hypothetical protein VITISV_043805 [Vitis vinifera]	169859 bits(478)	3,11E-46	74/101(73%)	88/101(87%)	1/101(1%)
emb CAN68784.1	hypothetical protein VITISV_021537 [Vitis vinifera]	185267 bits(478)	1,12E-51	79/109(72%)	95/109(87%)	1/109(1%)
ref XP_002264235.1	PREDICTED: laccase-14-like [Vitis vinifera]	185267 bits(478)	1,15E-51	79/109(72%)	95/109(87%)	1/109(1%)
ref XP_003603035.1	Laccase-15 [Medicago truncatula]	169088 bits(478)	1,42E-44	71/99(72%)	83/99(84%)	0/99(0%)
ref XP_002264394.2	PREDICTED: laccase-14 [Vitis vinifera]	194897 bits(478)	2,79E-55	87/123(71%)	102/123(83%)	5/123(4%)
ref XP_002264702.1	PREDICTED: laccase-14-like [Vitis vinifera]	192586 bits(478)	2,25E-54	86/123(70%)	102/123(83%)	5/123(4%)
ref XP_003634918.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	191815 bits(478)	7,11E-54	86/123(70%)	100/123(81%)	5/123(4%)
gb EMJ28500.1	hypothetical protein PRUPE_ppa016865mg [Prunus persica]	168318 bits(478)	2,03E-45	72/103(70%)	85/103(83%)	0/103(0%)
ref XP_002299682.1	predicted protein [Populus trichocarpa]	166777 bits(478)	6,62E-45	70/100(70%)	83/100(83%)	0/100(0%)
ref XP_002452547.1	hypothetical protein SORBIDRAFT_04g027850 [Sorghum bicolor]	164851 bits(478)	4,00E-44	71/101(70%)	81/101(80%)	0/101(0%)
ref XP_003634921.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	186808 bits(478)	3,13E-52	85/123(69%)	98/123(80%)	5/123(4%)
emb CBI29319.3	unnamed protein product [Vitis vinifera]	185267 bits(478)	5,20E-52	85/123(69%)	98/123(80%)	5/123(4%)
ref XP_002269593.2	PREDICTED: laccase-14-like [Vitis vinifera]	185652 bits(478)	7,72E-52	85/123(69%)	98/123(80%)	5/123(4%)
ref XP_002449707.1	hypothetical protein SORBIDRAFT_05g021895 [Sorghum bicolor]	16254 bits(478)	4,19E-47	70/101(69%)	82/101(81%)	0/101(0%)
ref XP_003635327.1	PREDICTED: laccase-14-like [Vitis vinifera]	19066 bits(478)	1,22E-53	84/123(68%)	99/123(80%)	5/123(4%)
ref XP_003635465.1	PREDICTED: laccase-14-like [Vitis vinifera]	187578 bits(478)	1,62E-52	84/123(68%)	101/123(82%)	5/123(4%)
emb CBI29324.3	unnamed protein product [Vitis vinifera]	186422 bits(478)	1,93E-52	84/123(68%)	100/123(81%)	5/123(4%)
emb CAN60212.1	hypothetical protein VITISV_002465 [Vitis vinifera]	186422 bits(478)	3,77E-52	84/123(68%)	100/123(81%)	5/123(4%)
ref XP_002271713.1	PREDICTED: putative laccase-9 [Vitis vinifera]	186422 bits(478)	5,02E-52	84/123(68%)	99/123(80%)	5/123(4%)

ref XP_003635347.1	PREDICTED: laccase-14-like [Vitis vinifera]	185652 bits(478)	7,56E-52	84/123(68%)	100/123(81%)	5/123(4%)
emb CAN68492.1	hypothetical protein VITISV_042551 [Vitis vinifera]	18103 bits(478)	3,51E-50	84/124(68%)	98/124(79%)	6/124(5%)
ref XP_002307537.1	predicted protein [Populus trichocarpa]	177563 bits(478)	8,71E-49	77/114(68%)	93/114(82%)	5/114(4%)
emb CBI29315.3	unnamed protein product [Vitis vinifera]	168318 bits(478)	9,41E-49	77/114(68%)	87/114(76%)	5/114(4%)
ref XP_002307536.1	predicted protein [Populus trichocarpa]	176022 bits(478)	1,65E-48	78/114(68%)	92/114(81%)	5/114(4%)
ref XP_002272618.1	PREDICTED: laccase-14-like [Vitis vinifera]	170244 bits(478)	4,15E-46	77/113(68%)	87/113(77%)	5/113(4%)
ref XP_003616564.1	Laccase [Medicago truncatula]	166392 bits(478)	1,75E-44	68/100(68%)	81/100(81%)	0/100(0%)
emb CBI18499.3	unnamed protein product [Vitis vinifera]	19066 bits(478)	9,02E-56	82/123(67%)	101/123(82%)	5/123(4%)
emb CBI18500.3	unnamed protein product [Vitis vinifera]	184111 bits(478)	2,38E-54	83/123(67%)	97/123(79%)	5/123(4%)
emb CBI18291.3	unnamed protein product [Vitis vinifera]	189504 bits(478)	7,83E-54	83/123(67%)	100/123(81%)	5/123(4%)
ref XP_002264812.2	PREDICTED: putative laccase-9-like [Vitis vinifera]	19143 bits(478)	1,91E-53	83/123(67%)	100/123(81%)	5/123(4%)
ref XP_002271411.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	189504 bits(478)	3,60E-53	82/123(67%)	101/123(82%)	5/123(4%)
ref XP_002265415.1	PREDICTED: laccase-14-like [Vitis vinifera]	189504 bits(478)	3,99E-53	83/123(67%)	100/123(81%)	5/123(4%)
emb CAN64853.1	hypothetical protein VITISV_027067 [Vitis vinifera]	189504 bits(478)	4,38E-53	83/123(67%)	100/123(81%)	5/123(4%)
ref XP_002265095.2	PREDICTED: laccase-14-like [Vitis vinifera]	188734 bits(478)	6,64E-53	83/123(67%)	101/123(82%)	5/123(4%)
ref XP_003634563.1	PREDICTED: LOW QUALITY PROTEIN: putative laccase-9-like [Vitis vinifera]	187578 bits(478)	1,58E-52	83/123(67%)	98/123(80%)	5/123(4%)
ref XP_003635635.1	PREDICTED: laccase-14-like, partial [Vitis vinifera]	183341 bits(478)	1,65E-52	82/123(67%)	98/123(80%)	5/123(4%)
emb CBI29318.3	unnamed protein product [Vitis vinifera]	187193 bits(478)	2,30E-52	82/123(67%)	99/123(80%)	5/123(4%)
emb CAN68491.1	hypothetical protein VITISV_042550 [Vitis vinifera]	187193 bits(478)	2,32E-52	82/123(67%)	99/123(80%)	5/123(4%)
ref XP_002272088.1	PREDICTED: laccase-14 [Vitis vinifera]	187193 bits(478)	2,41E-52	82/123(67%)	99/123(80%)	5/123(4%)
ref XP_003634928.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	186808 bits(478)	4,26E-52	82/123(67%)	100/123(81%)	5/123(4%)
ref XP_003634564.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	186037 bits(478)	6,83E-52	82/123(67%)	97/123(79%)	5/123(4%)
ref XP_003635523.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	185652 bits(478)	8,04E-52	83/123(67%)	100/123(81%)	5/123(4%)
ref XP_003634923.1	PREDICTED: laccase-14-like [Vitis vinifera]	185267 bits(478)	1,26E-51	83/123(67%)	99/123(80%)	5/123(4%)
emb CAN68468.1	hypothetical protein VITISV_012531 [Vitis vinifera]	185267 bits(478)	1,48E-51	83/123(67%)	99/123(80%)	5/123(4%)
ref XP_003634927.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	184111 bits(478)	3,33E-51	82/123(67%)	100/123(81%)	5/123(4%)

ref XP_002265445.2	PREDICTED: putative laccase-9-like [Vitis vinifera]	183726 bits(478)	5,02E-51	83/123(67%)	99/123(80%)	5/123(4%)
ref XP_003634555.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	183726 bits(478)	5,03E-51	83/124(67%)	98/124(79%)	6/124(5%)
ref XP_002271127.1	PREDICTED: putative laccase-9-like [Vitis vinifera]	182956 bits(478)	9,28E-51	83/123(67%)	98/123(80%)	5/123(4%)
gb EMJ06199.1	hypothetical protein PRUPE_ppa003546mg [Prunus persica]	18103 bits(478)	4,05E-50	83/123(67%)	96/123(78%)	5/123(4%)
gb EMJ01390.1	hypothetical protein PRUPE_ppa016123mg [Prunus persica]	167548 bits(478)	3,86E-45	76/114(67%)	86/114(75%)	5/114(4%)
gb EEC74012.1	hypothetical protein Osl_08946 [Oryza sativa Indica Group]	166007 bits(478)	1,47E-44	70/106(66%)	86/106(81%)	0/106(0%)
dbj BAF01776.1	laccase (diphenol oxidase)-like protein [Arabidopsis thaliana]	160614 bits(478)	4,27E-44	67/101(66%)	81/101(80%)	0/101(0%)
ref XP_002873397.1	laccase family protein/diphenol oxidase family protein [Arabidopsis lyrata subsp. lyrata]	164466 bits(478)	5,53E-44	68/103(66%)	83/103(81%)	0/103(0%)
emb CBI29314.3	unnamed protein product [Vitis vinifera]	16331 bits(478)	2,12E-47	74/114(65%)	88/114(77%)	5/114(4%)
ref NP_001242162.1	uncharacterized protein LOC100788550 precursor [Glycine max]	171785 bits(478)	9,52E-47	79/122(65%)	94/122(77%)	6/122(5%)
sp Q6Z8L2.1 LAC9_ORYSJ	RecName: Full=Putative laccase-9; AltName: Full=Benzenediol:oxygen oxidoreductase 9; AltName: Full=Diphenol oxidase 9; AltName: Full=Urishiol oxidase 9; Flags: Precursor	165236 bits(478)	2,45E-44	69/106(65%)	86/106(81%)	0/106(0%)
gb ACC78283.1	putative laccase [Rosa hybrid cultivar]	165236 bits(478)	2,57E-44	74/114(65%)	90/114(79%)	5/114(4%)
ref XP_002272689.1	PREDICTED: laccase-14-like [Vitis vinifera]	164851 bits(478)	3,25E-44	74/114(65%)	88/114(77%)	5/114(4%)
gb AFG57867.1	Pinus taeda anonymous locus 0_9423_01 genomic sequence	154451 bits(478)	5,10E-44	66/101(65%)	80/101(79%)	0/101(0%)
ref NP_001241720.1	uncharacterized protein LOC100856898 [Zea mays]	174481 bits(478)	1,28E-48	77/120(64%)	94/120(78%)	2/120(2%)
ref NP_001105915.1	putative laccase precursor [Zea mays]	173711 bits(478)	2,46E-47	77/120(64%)	94/120(78%)	2/120(2%)
gb AFW64473.1	putative laccase family protein [Zea mays]	173326 bits(478)	3,57E-47	77/120(64%)	94/120(78%)	2/120(2%)
ref XP_002452900.1	hypothetical protein SORBIDRAFT_04g034610 [Sorghum bicolor]	169859 bits(478)	6,51E-46	77/120(64%)	91/120(76%)	2/120(2%)
sp Q7XE50.1 LAC16_ORYSJ	RecName: Full=Putative laccase-16; AltName: Full=Benzenediol:oxygen oxidoreductase 16; AltName: Full=Diphenol oxidase 16; AltName: Full=Urishiol oxidase 16	163696 bits(478)	3,45E-44	68/106(64%)	86/106(81%)	0/106(0%)
ref XP_003634930.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	179874 bits(478)	6,25E-50	78/123(63%)	96/123(78%)	5/123(4%)
ref XP_003634542.1	PREDICTED: LOW QUALITY PROTEIN: laccase-14-like [Vitis vinifera]	179104 bits(478)	2,24E-49	78/123(63%)	99/123(80%)	5/123(4%)

ref XP_002452548.1	hypothetical protein SORBIDRAFT_04g027860 [Sorghum bicolor]	172555 bits(478)	6,77E-47	76/120(63%)	93/120(78%)	2/120(2%)
gb AFW73452.1	putative laccase family protein [Zea mays]	1714 bits(478)	1,72E-46	76/120(63%)	93/120(78%)	2/120(2%)
ref XP_004252634.1	PREDICTED: laccase-15-like [Solanum lycopersicum]	170629 bits(478)	2,74E-46	76/123(62%)	91/123(74%)	5/123(4%)
gb EEC70671.1	hypothetical protein Osl_01989 [Oryza sativa Indica Group]	161384 bits(478)	4,72E-44	66/106(62%)	84/106(79%)	0/106(0%)
ref XP_002452901.1	hypothetical protein SORBIDRAFT_04g034620 [Sorghum bicolor]	167162 bits(478)	5,46E-45	73/120(61%)	91/120(76%)	2/120(2%)

Contig30

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb ADB02892.1	metallothionein-like MT-3 [Jatropha curcas]	10476 bits(260)	4,33E-26	51/64(80%)	56/64(88%)	0/64(0%)
ref XP_002525821.1	conserved hypothetical protein [Ricinus communis]	985969 bits(260)	9,41E-24	49/64(77%)	56/64(88%)	1/64(2%)
ref XP_002531714.1	conserved hypothetical protein [Ricinus communis]	966709 bits(260)	5,06E-23	49/64(77%)	56/64(88%)	1/64(2%)
sp Q96386.1 MT3_CARPA	RecName: Full=Metallothionein-like protein type 3; Short=MT-3	920485 bits(260)	3,62E-21	50/66(76%)	56/66(85%)	3/66(5%)
gb ABK96303.1	unknown [Populus trichocarpa x Populus deltoides]	785666 bits(260)	3,52E-16	40/53(75%)	45/53(85%)	1/53(2%)
ref XP_002316894.1	predicted protein [Populus trichocarpa]	770258 bits(260)	1,23E-15	40/53(75%)	44/53(83%)	1/53(2%)
ref XP_003527252.1	PREDICTED: uncharacterized protein LOC100305954 [Glycine max]	612326 bits(260)	7,40E-10	38/51(75%)	41/51(80%)	3/51(6%)
dbj BAD95608.1	metallothionein 3 [Populus alba x Populus glandulosa]	75485 bits(260)	4,46E-15	39/53(74%)	43/53(81%)	1/53(2%)
ref XP_002329134.1	predicted protein [Populus trichocarpa]	73559 bits(260)	3,05E-14	36/50(72%)	43/50(86%)	1/50(2%)
ref XP_002284956.1	PREDICTED: metallothionein-like protein-like isoform 1 [Vitis vinifera]	947449 bits(260)	2,62E-22	46/65(71%)	56/65(86%)	1/65(2%)
emb CAA07565.1	metallothionein-like protein [Ribes nigrum]	893521 bits(260)	3,76E-20	46/65(71%)	52/65(80%)	1/65(2%)
gb AEQ54919.1	metallothionin 3 [Salvia miltiorrhiza]	862705 bits(260)	5,34E-19	47/66(71%)	54/66(82%)	5/66(8%)
gb ABK96100.1	unknown [Populus trichocarpa]	708626 bits(260)	2,35E-13	35/50(70%)	42/50(84%)	1/50(2%)
gb ADR30789.1	metallothionein 3-like protein [Hevea brasiliensis]	908929 bits(260)	8,14E-21	44/64(69%)	53/64(83%)	1/64(2%)

sp P43389.1 MT3_ACTDE	RecName: Full=Metallothionein-like protein type 3	77411 bits(260)	1,05E-15	44/65(68%)	52/65(80%)	4/65(6%)
gb AAT02527.1	metallothionein 3b [Populus trichocarpa x Populus deltoides]	67781 bits(260)	3,88E-12	34/50(68%)	40/50(80%)	1/50(2%)
gb ABN46986.1	metallothionein-like protein 3 [Nelumbo nucifera]	928189 bits(260)	1,44E-21	45/67(67%)	55/67(82%)	1/67(1%)
gb ABA43635.1	metallothionein-like protein [Metroxylon sagu]	889669 bits(260)	5,43E-20	43/65(66%)	54/65(83%)	1/65(2%)
gb AAO92264.1	metallothionein-like protein [Arachis hypogaea]	804925 bits(260)	7,27E-17	41/62(66%)	49/62(79%)	2/62(3%)
emb CAA11845.1	metallothionein-like protein [Rubus idaeus]	739442 bits(260)	1,97E-14	41/62(66%)	50/62(81%)	4/62(6%)
gb AAK08208.1 AF320905_1	metallothionein-like protein [Citrus unshiu]	897373 bits(260)	2,60E-20	44/68(65%)	56/68(82%)	1/68(1%)
gb ABL67648.1	putative metallothionein-like protein [Citrus hybrid cultivar]	893521 bits(260)	3,52E-20	44/68(65%)	56/68(82%)	1/68(1%)
emb CAB52585.1	metallothionein-like protein [Elaeis guineensis]	862705 bits(260)	4,88E-19	42/65(65%)	52/65(80%)	1/65(2%)
gb AAW47577.1	metallothionein-like [Gossypium hirsutum]	812629 bits(260)	3,98E-17	44/68(65%)	55/68(81%)	6/68(9%)
gb AAD22465.1 AF118230_1	metallothionein-like protein [Gossypium hirsutum]	777962 bits(260)	6,04E-16	43/66(65%)	50/66(76%)	5/66(8%)
gb AFP54303.1	metallothionein-like protein [Pyrus x bretschneideri]	758702 bits(260)	3,99E-15	40/62(65%)	45/62(73%)	1/62(2%)
gb AAW52725.1	metallothionein-like protein [Fragaria x ananassa]	693218 bits(260)	9,28E-13	40/62(65%)	49/62(79%)	4/62(6%)
gb AFP57435.1	putative metallothionein type 3 [Brassica napus]	858853 bits(260)	6,29E-19	41/64(64%)	52/64(81%)	0/64(0%)
gb ACC77568.1	type 3 metallothionein [Prosopis juliflora]	851149 bits(260)	1,46E-18	42/66(64%)	53/66(80%)	1/66(2%)
gb AFK33631.1	unknown [Lotus japonicus]	781814 bits(260)	6,00E-16	41/64(64%)	50/64(78%)	2/64(3%)
gb ADO12864.1	metallothionein type 3-like protein [Ipomoea aquatica]	708626 bits(260)	2,55E-13	43/67(64%)	51/67(76%)	4/67(6%)
gb ACB10219.2	metallothionein-like protein [Elaeis guineensis]	831889 bits(260)	7,61E-18	41/65(63%)	51/65(78%)	1/65(2%)
sp Q40256.1 MT3_MUSAC	RecName: Full=Metallothionein-like protein type 3; Short=MT-3; AltName: Full=MWMT3	831889 bits(260)	7,69E-18	41/65(63%)	50/65(77%)	1/65(2%)
gb ACV51811.1	metallothionein type 3 [Typha angustifolia]	804925 bits(260)	7,72E-17	41/65(63%)	50/65(77%)	2/65(3%)
gb ABV58319.1	metallothionein class I type 3 [Avicennia marina]	789518 bits(260)	2,47E-16	39/62(63%)	48/62(77%)	2/62(3%)
sp Q24059.1 MT3_MALDO	RecName: Full=Metallothionein-like protein type 3; Short=MT-3	750998 bits(260)	6,85E-15	39/62(63%)	46/62(74%)	1/62(2%)
gb AAF78526.1 AF195234_1	metallothionein-like protein [Pyrus pyrifolia]	720182 bits(260)	8,48E-14	38/60(63%)	43/60(72%)	1/60(2%)
ref XP_004295176.1	PREDICTED: metallothionein-like protein 1-like [Fragaria vesca subsp. vesca]	693218 bits(260)	9,28E-13	39/62(63%)	48/62(77%)	4/62(6%)
dbj BAB85599.1	metallothionein type 3 [Brassica juncea]	866557 bits(260)	4,23E-19	40/64(62%)	52/64(81%)	0/64(0%)
dbj BAB85600.1	metallothionein type 3 [Brassica juncea]	835741 bits(260)	5,37E-18	40/64(62%)	51/64(80%)	0/64(0%)

dbj BAB85601.1	metallothionein type 3 [Brassica juncea]	75485 bits(260)	5,26E-15	40/65(62%)	52/65(80%)	2/65(3%)
gb ACR46965.1	metallothionein 3 [Noccaea caerulescens]	739442 bits(260)	1,73E-14	40/65(62%)	51/65(78%)	2/65(3%)
gb ACL80667.1	metallothionein [Solanum nigrum]	681662 bits(260)	2,50E-12	41/66(62%)	50/66(76%)	5/66(8%)
gb AAK27970.1 AF242374_1	metallothionein-like protein [Ipomoea batatas]	666254 bits(260)	8,25E-12	41/66(62%)	47/66(71%)	4/66(6%)
gb AFK34468.1	unknown [Lotus japonicus]	627734 bits(260)	1,85E-10	31/50(62%)	36/50(72%)	1/50(2%)
ref NP_566509.1	metallothionein 3 [Arabidopsis thaliana]	781814 bits(260)	5,17E-16	39/64(61%)	51/64(80%)	1/64(2%)
dbj BAD52241.1	metallothionein [Oryza rufipogon]	689366 bits(260)	1,37E-12	40/66(61%)	47/66(71%)	6/66(9%)
dbj BAD52234.1	metallothionein [Oryza rufipogon]	681662 bits(260)	2,02E-12	40/66(61%)	47/66(71%)	6/66(9%)
gb AAF68995.1	metallothionein [Oryza coarctata]	666254 bits(260)	8,09E-12	41/67(61%)	49/67(73%)	6/67(9%)
gb ACL80666.1	metallothionein [Solanum nigrum]	666254 bits(260)	8,50E-12	40/66(61%)	50/66(76%)	5/66(8%)
ref XP_002885083.1	hypothetical protein ARALYDRAFT_897819 [Arabidopsis lyrata subsp. lyrata]	758702 bits(260)	3,51E-15	39/65(60%)	53/65(82%)	3/65(5%)
gb AAM19713.1 AF499726_1	metallothionein-like protein [Eutrema halophilum]	73559 bits(260)	2,43E-14	39/66(59%)	50/66(76%)	2/66(3%)
sp O48951.1 MT3_PRUAV	RecName: Full=Metallothionein-like protein 1; Short=MT-1	708626 bits(260)	2,28E-13	37/63(59%)	43/63(68%)	1/63(2%)
ref NP_001189899.1	metallothionein 3 [Arabidopsis thaliana]	708626 bits(260)	2,99E-13	38/64(59%)	49/64(77%)	3/64(5%)
ref NP_001042319.1	Os01g0200700 [Oryza sativa Japonica Group]	700922 bits(260)	5,03E-13	39/66(59%)	47/66(71%)	6/66(9%)
ref XP_004133709.1	PREDICTED: metallothionein-like protein type 3-like [Cucumis sativus]	67781 bits(260)	3,20E-12	38/64(59%)	43/64(67%)	2/64(3%)
gb AAB95220.1	metallothionein type I [Fritillaria agrestis]	673958 bits(260)	3,83E-12	38/64(59%)	47/64(73%)	4/64(6%)
gb AAX39388.1	metallothionein-like protein [Oryza officinalis]	666254 bits(260)	7,54E-12	39/66(59%)	47/66(71%)	6/66(9%)
gb AAC13291.1	metallothionein-like protein [Fagopyrum esculentum]	635438 bits(260)	8,62E-11	37/63(59%)	47/63(75%)	6/63(10%)
gb EMJ13542.1	hypothetical protein PRUPE_ppa014506mg [Prunus persica]	739442 bits(260)	1,86E-14	38/65(58%)	45/65(69%)	1/65(2%)
ref XP_003628521.1	Metallothionein-like protein [Medicago truncatula]	67781 bits(260)	2,76E-12	37/64(58%)	48/64(75%)	4/64(6%)
gb ACL80668.1	metallothionein [Solanum nigrum]	67781 bits(260)	3,30E-12	38/66(58%)	47/66(71%)	3/66(5%)
gb AAB95219.1	metallothionein type I [Fritillaria agrestis]	643142 bits(260)	4,81E-11	37/64(58%)	46/64(72%)	4/64(6%)
dbj BAI39990.1	metallothionein [Tamarix androssowii]	77411 bits(260)	1,03E-15	37/65(57%)	48/65(74%)	1/65(2%)
gb ADN33916.1	metallothionein-like protein [Cucumis melo subsp. melo]	662402 bits(260)	1,20E-11	36/64(56%)	44/64(69%)	2/64(3%)
gb ABG75913.1	metallothionein [Fagopyrum esculentum]	654698 bits(260)	2,27E-11	37/66(56%)	50/66(76%)	6/66(9%)

gb AAX39389.1	metallothionein-like protein [Oryza officinalis]	650846 bits(260)	2,58E-11	37/66(56%)	45/66(68%)	4/66(6%)
gb ABG75912.1	metallothionein [Fagopyrum esculentum]	646994 bits(260)	3,79E-11	36/66(55%)	50/66(76%)	6/66(9%)
gb AAS99234.1	methallothionein type 3 [Noccaea caerulea]	63929 bits(260)	8,26E-11	36/65(55%)	47/65(72%)	2/65(3%)
sp A3B0Y1.1 MT3B_ORYSJ	RecName: Full=Metallothionein-like protein 3B; AltName: Full=Class I metallothionein-like protein 3B; AltName: Full=OsMT-I-3b	635438 bits(260)	1,02E-10	36/66(55%)	45/66(68%)	3/66(5%)
gb ACJ09673.1	putative metallothionein-like protein [Cupressus sempervirens]	65855 bits(260)	6,25E-11	39/73(53%)	52/73(71%)	5/73(7%)
sp A2Y1D7.1 MT3B_ORYSI	RecName: Full=Metallothionein-like protein 3B; AltName: Full=Class I metallothionein-like protein 3B; AltName: Full=OsMT-I-3b	62003 bits(260)	3,51E-10	35/66(53%)	45/66(68%)	3/66(5%)

Contig33

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_004066697.1	PREDICTED: protein unc-80 homolog [Oryzias latipes]	824185 bits(194)	1,81E-14	62/124(50%)	68/124(55%)	10/124(8%)
gb ELW62064.1	Foot protein 1 variant 1 [Tupaia chinensis]	69707 bits(194)	1,76E-10	47/124(38%)	73/124(59%)	10/124(8%)

Contig34

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
dbj BAD94561.1	adenylate translocator [Arabidopsis thaliana]	75485 bits(193)	4,20E-15	48/55(87%)	51/55(93%)	0/55(0%)
gb AFK48077.1	unknown [Lotus japonicus]	789518 bits(193)	7,82E-16	50/60(83%)	54/60(90%)	0/60(0%)
ref XP_002299948.1	predicted protein [Populus trichocarpa]	766406 bits(193)	9,64E-14	50/60(83%)	53/60(88%)	0/60(0%)
dbj BAH20415.1	AT3G08580 [Arabidopsis thaliana]	770258 bits(193)	1,29E-14	49/60(82%)	54/60(90%)	0/60(0%)
ref NP_187470.1	ADP,ATP carrier protein 1 [Arabidopsis thaliana]	770258 bits(193)	5,61E-14	49/60(82%)	54/60(90%)	0/60(0%)
ref XP_002884685.1	adenosine nucleotide translocator [Arabidopsis lyrata subsp. lyrata]	766406 bits(193)	9,72E-14	49/60(82%)	53/60(88%)	0/60(0%)

ref XP_002517886.1	ADP,ATP carrier protein, putative [<i>Ricinus communis</i>]	758702 bits(193)	1,53E-13	49/60(82%)	53/60(88%)	0/60(0%)
emb CAA05979.1	adenine nucleotide translocator [<i>Lupinus albus</i>]	75485 bits(193)	2,22E-13	49/60(82%)	53/60(88%)	0/60(0%)
gb ADN33720.1	adenine nucleotide translocator [<i>Cucumis melo</i> subsp. melo]	75485 bits(193)	2,68E-13	49/60(82%)	53/60(88%)	0/60(0%)
gb ADN33719.1	adenine nucleotide translocator [<i>Cucumis melo</i> subsp. melo]	75485 bits(193)	2,73E-13	49/60(82%)	53/60(88%)	0/60(0%)
ref XP_002531911.1	ADP,ATP carrier protein, putative [<i>Ricinus communis</i>]	747146 bits(193)	3,96E-13	49/60(82%)	52/60(87%)	0/60(0%)
gb AAM65696.1	ADP,ATP carrier-like protein [<i>Arabidopsis thaliana</i>]	689366 bits(193)	3,80E-11	46/56(82%)	50/56(89%)	0/56(0%)
ref XP_002303893.1	predicted protein [<i>Populus trichocarpa</i>]	731738 bits(193)	1,39E-12	48/59(81%)	52/59(88%)	0/59(0%)
gb AAL15894.1 AF417306_1	putative adenine nucleotide translocase [<i>Castanea sativa</i>]	739442 bits(193)	5,75E-14	48/60(80%)	51/60(85%)	0/60(0%)
ref XP_003542855.1	PREDICTED: ADP,ATP carrier protein 2, mitochondrial-like [<i>Glycine max</i>]	747146 bits(193)	3,67E-13	48/60(80%)	53/60(88%)	0/60(0%)
ref XP_004142869.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [<i>Cucumis sativus</i>]	747146 bits(193)	4,32E-13	48/60(80%)	53/60(88%)	0/60(0%)
ref XP_003531442.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [<i>Glycine max</i>]	747146 bits(193)	4,38E-13	48/60(80%)	53/60(88%)	0/60(0%)
ref XP_003546882.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [<i>Glycine max</i>]	747146 bits(193)	4,59E-13	48/60(80%)	53/60(88%)	0/60(0%)
ref XP_004142870.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [<i>Cucumis sativus</i>]	747146 bits(193)	4,90E-13	48/60(80%)	53/60(88%)	0/60(0%)
ref XP_002299232.1	predicted protein [<i>Populus trichocarpa</i>]	747146 bits(193)	4,96E-13	48/60(80%)	52/60(87%)	0/60(0%)
gb EMJ12607.1	hypothetical protein PRUPE_ppa006913mg [<i>Prunus persica</i>]	739442 bits(193)	7,29E-13	48/60(80%)	53/60(88%)	0/60(0%)
emb CAN81107.1	hypothetical protein VITISV_008170 [<i>Vitis vinifera</i>]	73559 bits(193)	9,67E-13	48/60(80%)	52/60(87%)	0/60(0%)
ref XP_002279748.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [<i>Vitis vinifera</i>]	73559 bits(193)	1,01E-12	48/60(80%)	52/60(87%)	0/60(0%)
gb ACU18447.1	unknown [<i>Glycine max</i>]	73559 bits(193)	1,28E-12	48/60(80%)	53/60(88%)	0/60(0%)
emb CBI16399.3	unnamed protein product [<i>Vitis vinifera</i>]	720182 bits(193)	1,62E-12	47/59(80%)	52/59(88%)	0/59(0%)
gb AFK44525.1	unknown [<i>Medicago truncatula</i>]	724034 bits(193)	2,07E-12	47/59(80%)	52/59(88%)	0/59(0%)
ref XP_003627715.1	ADP,ATP carrier protein [<i>Medicago truncatula</i>]	727886 bits(193)	2,38E-12	47/59(80%)	52/59(88%)	0/59(0%)
ref XP_002285503.1	PREDICTED: ADP,ATP carrier protein, mitochondrial-like [<i>Vitis vinifera</i>]	720182 bits(193)	3,44E-12	47/59(80%)	52/59(88%)	0/59(0%)
emb CAN66307.1	hypothetical protein VITISV_009087 [<i>Vitis vinifera</i>]	71633 bits(193)	5,54E-12	47/59(80%)	51/59(86%)	0/59(0%)
gb EMJ19324.1	hypothetical protein PRUPE_ppa007062mg [<i>Prunus persica</i>]	762554 bits(193)	1,09E-13	47/60(78%)	54/60(90%)	0/60(0%)

ref XP_004306530.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [Fragaria vesca subsp. vesca]	747146 bits(193)	4,24E-13	47/60(78%)	53/60(88%)	0/60(0%)
ref XP_004306529.1	PREDICTED: ADP,ATP carrier protein 1, mitochondrial-like [Fragaria vesca subsp. vesca]	747146 bits(193)	4,33E-13	47/60(78%)	52/60(87%)	0/60(0%)
ref XP_002310404.1	predicted protein [Populus trichocarpa]	73559 bits(193)	9,19E-13	47/60(78%)	53/60(88%)	0/60(0%)
sp O22342.1 ADT1_GOSHI	RecName: Full=ADP,ATP carrier protein 1, mitochondrial; AltName: Full=ADP/ATP translocase 1; AltName: Full=Adenine nucleotide translocator 1; Short=ANT 1; Flags: Precursor	739442 bits(193)	9,39E-13	47/60(78%)	52/60(87%)	0/60(0%)
ref XP_002871570.1	hypothetical protein ARALYDRAFT_488169 [Arabidopsis lyrata subsp. lyrata]	724034 bits(193)	2,53E-12	47/60(78%)	52/60(87%)	0/60(0%)
emb CAA46518.1	adenylate translocator [Arabidopsis thaliana]	720182 bits(193)	3,14E-12	47/60(78%)	52/60(87%)	0/60(0%)
ref NP_196853.1	ADP,ATP carrier protein 2 [Arabidopsis thaliana]	71633 bits(193)	4,58E-12	47/60(78%)	51/60(85%)	0/60(0%)
emb CAA48579.1	adenosine nucleotide translocator [Arabidopsis thaliana]	71633 bits(193)	4,71E-12	47/60(78%)	51/60(85%)	0/60(0%)
prf 1908224A	nucleotide translocator	71633 bits(193)	5,63E-12	47/60(78%)	51/60(85%)	0/60(0%)
ref XP_002327912.1	predicted protein [Populus trichocarpa]	712478 bits(193)	6,09E-12	46/59(78%)	52/59(88%)	0/59(0%)
ref XP_002275525.2	PREDICTED: ADP,ATP carrier protein 3, mitochondrial [Vitis vinifera]	71633 bits(193)	6,86E-12	47/60(78%)	51/60(85%)	0/60(0%)
ref XP_004136845.1	PREDICTED: ADP,ATP carrier protein 3, mitochondrial-like [Cucumis sativus]	712478 bits(193)	7,43E-12	46/59(78%)	52/59(88%)	0/59(0%)
ref NP_194568.1	ADP/ATP carrier 3 protein [Arabidopsis thaliana]	704774 bits(193)	1,12E-11	47/60(78%)	51/60(85%)	0/60(0%)
ref XP_002867450.1	hypothetical protein ARALYDRAFT_491934 [Arabidopsis lyrata subsp. lyrata]	704774 bits(193)	1,20E-11	47/60(78%)	51/60(85%)	0/60(0%)
gb AFX67036.1	ADP, ATP carrier protein [Solanum tuberosum]	720182 bits(193)	3,94E-12	46/60(77%)	51/60(85%)	0/60(0%)
sp P27081.1 ADT2_SOLTU	RecName: Full=ADP,ATP carrier protein, mitochondrial; AltName: Full=ADP/ATP translocase; AltName: Full=Adenine nucleotide translocator; Short=ANT; Flags: Precursor	720182 bits(193)	3,95E-12	46/60(77%)	51/60(85%)	0/60(0%)
ref XP_004303109.1	PREDICTED: ADP,ATP carrier protein 3, mitochondrial-like [Fragaria vesca subsp. vesca]	693218 bits(193)	3,46E-11	46/60(77%)	52/60(87%)	0/60(0%)
ref XP_003539937.1	PREDICTED: ADP,ATP carrier protein 3, mitochondrial-like [Glycine max]	693218 bits(193)	3,46E-11	46/60(77%)	51/60(85%)	0/60(0%)
gb ABB29945.1	ADP/ATP translocator-like [Solanum tuberosum]	689366 bits(193)	4,22E-11	45/59(76%)	51/59(86%)	0/59(0%)
gb ABB55387.1	ADP,ATP carrier protein precursor-like [Solanum tuberosum]	689366 bits(193)	4,26E-11	45/59(76%)	51/59(86%)	0/59(0%)
sp P25083.1 ADT1_SOLTU	RecName: Full=ADP,ATP carrier protein, mitochondrial; AltName: Full=ADP/ATP translocase; AltName: Full=Adenine nucleotide translocator; Short=ANT; Flags: Precursor	689366 bits(193)	4,34E-11	45/59(76%)	51/59(86%)	0/59(0%)

ref XP_004229620.1	PREDICTED: ADP,ATP carrier protein, mitochondrial-like [Solanum lycopersicum]	704774 bits(193)	1,11E-11	44/59(75%)	51/59(86%)	0/59(0%)
ref XP_004243366.1	PREDICTED: ADP,ATP carrier protein 3, mitochondrial-like [Solanum lycopersicum]	700922 bits(193)	1,59E-11	45/60(75%)	51/60(85%)	0/60(0%)
ref NP_001055139.1	Os05g0302700 [Oryza sativa Japonica Group]	69707 bits(193)	2,26E-11	45/60(75%)	52/60(87%)	0/60(0%)
emb CAC27140.1	ADP, ATP carrier protein precursor [Picea abies]	681662 bits(193)	2,66E-11	44/59(75%)	51/59(86%)	0/59(0%)
ref NP_001241092.1	uncharacterized protein LOC100802974 [Glycine max]	689366 bits(193)	4,69E-11	45/60(75%)	51/60(85%)	0/60(0%)
ref XP_002468591.1	hypothetical protein SORBIDRAFT_01g048660 [Sorghum bicolor]	685514 bits(193)	5,44E-11	45/60(75%)	51/60(85%)	0/60(0%)
gb ABK24587.1	unknown [Picea sitchensis]	685514 bits(193)	5,50E-11	44/59(75%)	51/59(86%)	0/59(0%)
dbj BAD91181.1	putative mitochondrial adenylate transporter [Mesembryanthemum crystallinum]	65855 bits(193)	4,36E-10	44/59(75%)	49/59(83%)	0/59(0%)
ref XP_002513476.1	ADP,ATP carrier protein, putative [Ricinus communis]	654698 bits(193)	7,06E-10	44/59(75%)	51/59(86%)	0/59(0%)
sp Q41629.1 ADT1_WHEAT	RecName: Full=ADP,ATP carrier protein 1, mitochondrial; AltName: Full=ADP/ATP translocase 1; AltName: Full=Adenine nucleotide translocator 1; Short=ANT 1; Flags: Precursor	704774 bits(193)	8,88E-12	44/60(73%)	52/60(87%)	0/60(0%)
ref XP_003543539.1	PREDICTED: ADP,ATP carrier protein, mitochondrial-like [Glycine max]	685514 bits(193)	6,54E-11	43/59(73%)	51/59(86%)	0/59(0%)
ref NP_001150672.1	LOC100284305 [Zea mays]	681662 bits(193)	7,87E-11	44/60(73%)	51/60(85%)	0/60(0%)
gb ABR25608.1	mitochondrial ADP, ATP carrier protein precursor [Oryza sativa Indica Group]	700922 bits(193)	2,56E-12	43/60(72%)	52/60(87%)	0/60(0%)
gb EEC73899.1	hypothetical protein Osl_08713 [Oryza sativa Indica Group]	69707 bits(193)	2,20E-11	43/60(72%)	52/60(87%)	0/60(0%)
ref NP_001047941.1	Os02g0718900 [Oryza sativa Japonica Group]	69707 bits(193)	2,29E-11	43/60(72%)	52/60(87%)	0/60(0%)
gb ACJ84693.1	unknown [Medicago truncatula]	662402 bits(193)	2,51E-11	43/60(72%)	50/60(83%)	0/60(0%)
emb CAA33743.1	adenine nucleotide translocator [Zea mays]	69707 bits(193)	2,64E-11	43/60(72%)	51/60(85%)	0/60(0%)
gb AFW73142.1	adenine nucleotide translocator, mRNA [Zea mays]	689366 bits(193)	3,35E-11	43/60(72%)	51/60(85%)	0/60(0%)
emb CAA26600.1	unnamed protein product [Zea mays]	685514 bits(193)	3,62E-11	43/60(72%)	51/60(85%)	0/60(0%)
gb AFW63697.1	adenine nucleotide translocator2 [Zea mays]	693218 bits(193)	4,41E-11	43/60(72%)	51/60(85%)	0/60(0%)
ref NP_001105431.1	ADP,ATP carrier protein 1, mitochondrial precursor [Zea mays]	689366 bits(193)	4,53E-11	43/60(72%)	51/60(85%)	0/60(0%)
ref NP_001105434.1	ADP,ATP carrier protein 2, mitochondrial precursor [Zea mays]	689366 bits(193)	4,57E-11	43/60(72%)	51/60(85%)	0/60(0%)
emb CAA33742.1	adenine nucleotide translocator [Zea mays]	689366 bits(193)	4,75E-11	43/60(72%)	51/60(85%)	0/60(0%)

ref XP_003570256.1	PREDICTED: ADP,ATP carrier protein, mitochondrial-like isoform 1 [Brachypodium distachyon]	685514 bits(193)	1,04E-10	43/60(72%)	50/60(83%)	0/60(0%)
gb EMJ00030.1	hypothetical protein PRUPE_ppa007166mg [Prunus persica]	673958 bits(193)	1,65E-10	43/60(72%)	52/60(87%)	0/60(0%)
gb ACJ84706.1	unknown [Medicago truncatula]	654698 bits(193)	7,08E-10	43/60(72%)	50/60(83%)	0/60(0%)
ref NP_001234018.1	ADP,ATP translocator [Solanum lycopersicum]	689366 bits(193)	4,10E-11	45/63(71%)	52/63(83%)	0/63(0%)
ref XP_004291800.1	PREDICTED: ADP,ATP carrier protein, mitochondrial-like [Fragaria vesca subsp. vesca]	662402 bits(193)	3,28E-10	42/59(71%)	50/59(85%)	0/59(0%)
ref XP_003597438.1	ADP,ATP carrier protein [Medicago truncatula]	65855 bits(193)	5,11E-10	42/59(71%)	49/59(83%)	0/59(0%)
sp Q41630.1 ADT2_WHEAT	RecName: Full=ADP,ATP carrier protein 2, mitochondrial; AltName: Full=ADP/ATP translocase 2; AltName: Full=Adenine nucleotide translocator 2; Short=ANT 2; Flags: Precursor	65855 bits(193)	4,04E-10	42/60(70%)	50/60(83%)	0/60(0%)
gb ACF87847.1	unknown [Zea mays]	67781 bits(193)	7,81E-12	41/60(68%)	50/60(83%)	0/60(0%)
gb AAN11327.1	ADP-ATP translocase [Gaeumannomyces graminis var. tritici]	67781 bits(193)	6,22E-11	41/60(68%)	49/60(82%)	0/60(0%)
gb EJT71808.1	ADP,ATP carrier protein [Gaeumannomyces graminis var. tritici R3-111a-1]	67781 bits(193)	6,28E-11	41/60(68%)	49/60(82%)	0/60(0%)
ref NP_001142273.1	uncharacterized protein LOC100274442 [Zea mays]	681662 bits(193)	7,09E-11	41/60(68%)	50/60(83%)	0/60(0%)
gb ACL53828.1	unknown [Zea mays]	67781 bits(193)	7,16E-11	41/60(68%)	50/60(83%)	0/60(0%)
gb EFX06704.1	carnitine acyl carrier [Grosmanina clavigera kw1407]	673958 bits(193)	9,03E-11	41/60(68%)	49/60(82%)	0/60(0%)
gb AFW86475.1	hypothetical protein ZEAMMB73_378887 [Zea mays]	681662 bits(193)	1,11E-10	41/60(68%)	50/60(83%)	0/60(0%)
dbj BAJ99879.1	predicted protein [Hordeum vulgare subsp. vulgare]	673958 bits(193)	1,54E-10	41/60(68%)	52/60(87%)	0/60(0%)
dbj BAJ98173.1	predicted protein [Hordeum vulgare subsp. vulgare]	673958 bits(193)	1,54E-10	41/60(68%)	52/60(87%)	0/60(0%)
ref XP_002965475.1	hypothetical protein SELMODRAFT_230650 [Selaginella moellendorffii]	654698 bits(193)	3,30E-10	41/60(68%)	50/60(83%)	0/60(0%)
ref XP_002438320.1	hypothetical protein SORBIDRAFT_10g012150 [Sorghum bicolor]	670106 bits(193)	1,55E-10	40/60(67%)	50/60(83%)	0/60(0%)
ref XP_001909827.1	hypothetical protein [Podospora anserina S mat+]	666254 bits(193)	1,86E-10	39/60(65%)	49/60(82%)	0/60(0%)
gb EGS23048.1	putative ADP/ATP carrier protein [Chaetomium thermophilum var. thermophilum DSM 1495]	65855 bits(193)	3,16E-10	39/60(65%)	49/60(82%)	0/60(0%)
ref XP_001228473.1	ADP/ATP carrier protein [Chaetomium globosum CBS 148.51]	650846 bits(193)	6,05E-10	39/60(65%)	48/60(80%)	0/60(0%)
ref XP_002837794.1	ADP/ATP carrier protein [Tuber melanosporum Mel28]	646994 bits(193)	6,49E-10	39/60(65%)	49/60(82%)	0/60(0%)
ref XP_661668.1	ADT_NEUCR ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINE	646994 bits(193)	7,51E-10	39/60(65%)	48/60(80%)	0/60(0%)

	NUCLEOTIDE TRANSLOCATOR) (ANT) [Aspergillus nidulans FGSC A4]					
gb AA07662.1	ADP/ATP carrier protein-like protein [Magnaporthe grisea]	646994 bits(193)	7,67E-10	39/60(65%)	48/60(80%)	0/60(0%)
ref XP_003658188.1	ADP/ATP carrier protein [Thielavia terrestris NRRL 8126]	650846 bits(193)	6,09E-10	38/60(63%)	48/60(80%)	0/60(0%)
Contig44						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb ABK95490.1	unknown [Populus trichocarpa]	670106 bits(164)	1,64E-10	29/30(97%)	30/30(100%)	0/30(0%)
ref XP_002322593.1	predicted protein [Populus trichocarpa]	670106 bits(164)	1,71E-10	29/30(97%)	30/30(100%)	0/30(0%)
emb CAA06853.1	26S protease regulatory subunit 6 [Cicer arietinum]	67781 bits(164)	1,46E-11	33/43(77%)	34/43(79%)	5/43(12%)
ref XP_003600548.1	26S protease regulatory subunit 6B [Medicago truncatula]	681662 bits(164)	6,17E-11	33/43(77%)	34/43(79%)	5/43(12%)
gb AFK43608.1	unknown [Medicago truncatula]	681662 bits(164)	6,47E-11	33/43(77%)	34/43(79%)	5/43(12%)
ref XP_003538340.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Glycine max]	67781 bits(164)	8,04E-11	33/43(77%)	34/43(79%)	5/43(12%)
emb CBI15803.3	unnamed protein product [Vitis vinifera]	670106 bits(164)	1,07E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_004154554.1	PREDICTED: LOW QUALITY PROTEIN: 26S protease regulatory subunit 6B homolog [Cucumis sativus]	673958 bits(164)	1,28E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_004242453.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Solanum lycopersicum]	673958 bits(164)	1,30E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_004139931.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Cucumis sativus]	673958 bits(164)	1,30E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_004242474.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Solanum lycopersicum]	673958 bits(164)	1,36E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_002276130.2	PREDICTED: 26S protease regulatory subunit 6B homolog [Vitis vinifera]	673958 bits(164)	1,42E-10	32/43(74%)	34/43(79%)	5/43(12%)
sp P54778.1 PRS6B_SOLTU	RecName: Full=26S protease regulatory subunit 6B homolog	670106 bits(164)	1,47E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_002523664.1	26S protease regulatory subunit 6b, putative [Ricinus communis]	670106 bits(164)	1,50E-10	32/43(74%)	34/43(79%)	5/43(12%)

ref XP_002307945.1	predicted protein [Populus trichocarpa]	670106 bits(164)	1,61E-10	32/43(74%)	34/43(79%)	5/43(12%)
gb AFK37554.1	unknown [Lotus japonicus]	643142 bits(164)	1,83E-10	32/43(74%)	33/43(77%)	5/43(12%)
ref XP_003553190.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Glycine max]	670106 bits(164)	1,83E-10	32/43(74%)	34/43(79%)	5/43(12%)
ref XP_003538341.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Glycine max]	662402 bits(164)	2,98E-10	32/43(74%)	34/43(79%)	5/43(12%)
gb ACA58350.1	putative 26S proteasome regulatory complex protein [Sandersonia aurantiaca]	643142 bits(164)	4,01E-11	31/43(72%)	34/43(79%)	5/43(12%)
ref NP_200637.1	regulatory particle triple-A ATPase 3 [Arabidopsis thaliana]	654698 bits(164)	4,69E-10	31/43(72%)	33/43(77%)	5/43(12%)
sp P85200.1 PRS6B_HELAN	RecName: Full=26S protease regulatory subunit 6B homolog	654698 bits(164)	5,56E-10	31/43(72%)	34/43(79%)	5/43(12%)
ref XP_004294715.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Fragaria vesca subsp. vesca]	654698 bits(164)	6,18E-10	31/43(72%)	34/43(79%)	5/43(12%)
ref XP_004307102.1	PREDICTED: 26S protease regulatory subunit 6B homolog [Fragaria vesca subsp. vesca]	650846 bits(164)	6,27E-10	31/43(72%)	34/43(79%)	5/43(12%)
gb ADE75925.1	unknown [Picea sitchensis]	635438 bits(164)	5,14E-10	30/43(70%)	33/43(77%)	5/43(12%)
gb ACA61610.1	hypothetical protein AP2_E06.1 [Arabidopsis lyrata subsp. petraea]	662402 bits(164)	5,20E-11	35/59(59%)	40/59(68%)	7/59(12%)
Contig47						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_003540856.1	PREDICTED: isoprene synthase, chloroplastic-like [Glycine max]	966709 bits(239)	7,41E-20	42/75(56%)	57/75(76%)	0/75(0%)
ref XP_002309854.1	predicted protein [Populus trichocarpa]	870409 bits(239)	1,06E-16	42/76(55%)	54/76(71%)	1/76(1%)
sp Q672F7.1 TPS2_LOTJA	RecName: Full=Tricyclene synthase EBOS, chloroplastic; AltName: Full=(E)-beta-ocimene synthase; Short=LjEBOS; Short=LjEbetaOS; AltName: Full=Terpenoid synthase 2; Short=LjTPS2; Flags: Precursor	866557 bits(239)	1,41E-16	41/75(55%)	51/75(68%)	0/75(0%)
emb CAN65805.1	hypothetical protein VITISV_042691 [Vitis vinifera]	858853 bits(239)	2,83E-16	43/80(54%)	56/80(70%)	2/80(2%)
ref XP_002275444.1	PREDICTED: isoprene synthase, chloroplastic [Vitis vinifera]	858853 bits(239)	2,71E-16	43/82(52%)	57/82(70%)	2/82(2%)

gb ABY65110.1	beta-ocimene synthase [<i>Phaseolus lunatus</i>]	959005 bits(239)	1,15E-19	43/84(51%)	61/84(73%)	0/84(0%)
ref XP_003526313.1	PREDICTED: LOW QUALITY PROTEIN: isoprene synthase, chloroplastic-like [<i>Glycine max</i>]	939745 bits(239)	6,11E-19	43/84(51%)	59/84(70%)	1/84(1%)
sp Q5UB07.1 TPS4_MEDTR	RecName: Full=Tricyclene synthase TPS4, chloroplastic; AltName: Full=(E)-beta-ocimene synthase; Short=MtEBOS; Short=MtEbetaOS; AltName: Full=Terpenoid synthase 4; Short=MtTPS4; Flags: Precursor	843445 bits(239)	9,07E-16	38/74(51%)	52/74(70%)	0/74(0%)
gb AGB14628.1	(E)-beta-ocimene synthase [<i>Malus x domestica</i>]	839593 bits(239)	1,05E-15	37/73(51%)	50/73(68%)	0/73(0%)
gb ACD70404.1	putative chloroplast isoprene synthase [<i>Populus trichocarpa</i>]	770258 bits(239)	2,39E-13	37/73(51%)	47/73(64%)	0/73(0%)
sp Q50L36.1 ISPS_POPAL	RecName: Full=Isoprene synthase, chloroplastic; Short=PalspS; Flags: Precursor	770258 bits(239)	2,62E-13	37/73(51%)	47/73(64%)	0/73(0%)
ref XP_002327829.1	predicted protein [<i>Populus trichocarpa</i>]	770258 bits(239)	2,68E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb AEK70966.1	isoprene synthase [<i>Populus deltoides</i>]	770258 bits(239)	2,71E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb AEK70965.1	isoprene synthase [<i>Populus grandidentata</i>]	770258 bits(239)	2,72E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb AEK70964.1	isoprene synthase [<i>Populus balsamifera</i>]	770258 bits(239)	2,74E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb ABV04402.1	chloroplast isoprene synthase 1 [<i>Populus alba</i>]	770258 bits(239)	2,74E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb ABK95568.1	unknown [<i>Populus trichocarpa</i>]	770258 bits(239)	2,74E-13	37/73(51%)	47/73(64%)	0/73(0%)
sp Q7XAS7.1 ISPS_POPTM	RecName: Full=Isoprene synthase, chloroplastic; Flags: Precursor	770258 bits(239)	2,90E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb ADG96474.1	chloroplast isoprene synthase 3 [<i>Populus alba</i>]	770258 bits(239)	2,92E-13	37/73(51%)	47/73(64%)	0/73(0%)
emb CAL69918.1	isoprene synthase [<i>Populus nigra</i>]	770258 bits(239)	2,98E-13	37/73(51%)	47/73(64%)	0/73(0%)
pdb 3N0F A	Chain A, Crystal Structure Of Isoprene Synthase From Grey Poplar Leaves (<i>Populus X Canescens</i>)	766406 bits(239)	2,98E-13	37/73(51%)	47/73(64%)	0/73(0%)
gb ADG96473.1	chloroplast isoprene synthase 2 [<i>Populus alba</i>]	766406 bits(239)	3,54E-13	37/73(51%)	47/73(64%)	0/73(0%)
ref XP_002534394.1	(R)-limonene synthase, putative [<i>Ricinus communis</i>]	762554 bits(239)	4,35E-13	39/78(50%)	50/78(64%)	0/78(0%)
emb CBI40170.3	unnamed protein product [<i>Vitis vinifera</i>]	785666 bits(239)	6,42E-15	37/76(49%)	50/76(66%)	0/76(0%)
ref XP_002266808.1	PREDICTED: myrcene synthase, chloroplastic [<i>Vitis vinifera</i>]	789518 bits(239)	6,08E-14	37/76(49%)	50/76(66%)	0/76(0%)
gb AEK70967.1	isoprene synthase [<i>Populus fremontii</i>]	747146 bits(239)	1,43E-12	36/73(49%)	46/73(63%)	0/73(0%)
sp Q9AR86.1 ISPS_POPCN	RecName: Full=Isoprene synthase, chloroplastic; Flags: Precursor	71633 bits(239)	1,58E-11	36/73(49%)	46/73(63%)	0/73(0%)
emb CBI17967.3	unnamed protein product [<i>Vitis vinifera</i>]	797221 bits(239)	3,30E-15	37/77(48%)	54/77(70%)	2/77(3%)

gb AEK70968.1	isoprene synthase, partial [Robinia pseudoacacia]	804925 bits(239)	1,48E-14	35/75(47%)	51/75(68%)	0/75(0%)
emb CBI40171.3	unnamed protein product [Vitis vinifera]	777962 bits(239)	4,61E-14	36/76(47%)	50/76(66%)	0/76(0%)
ref XP_002267019.2	PREDICTED: myrcene synthase, chloroplastic-like [Vitis vinifera]	777962 bits(239)	2,29E-13	36/76(47%)	50/76(66%)	0/76(0%)
sp Q93X23.1 MYRS_QUEIL	RecName: Full=Myrcene synthase, chloroplastic; Flags: Precursor	770258 bits(239)	2,75E-13	35/75(47%)	48/75(64%)	0/75(0%)
emb CAN82377.1	hypothetical protein VITISV_019347 [Vitis vinifera]	708626 bits(239)	2,58E-11	34/73(47%)	46/73(63%)	0/73(0%)
ref XP_004161822.1	PREDICTED: (-)-alpha-terpineol synthase-like [Cucumis sativus]	681662 bits(239)	2,31E-10	36/77(47%)	46/77(60%)	2/77(3%)
ref XP_004150681.1	PREDICTED: myrcene synthase, chloroplastic-like [Cucumis sativus]	67781 bits(239)	2,94E-10	36/77(47%)	46/77(60%)	2/77(3%)
gb ADR74205.1	(E)-beta-ocimene synthase [Vitis vinifera]	797221 bits(239)	3,14E-14	38/82(46%)	56/82(68%)	2/82(2%)
ref XP_002281379.1	PREDICTED: isoprene synthase, chloroplastic-like isoform 1 [Vitis vinifera]	797221 bits(239)	3,17E-14	38/82(46%)	56/82(68%)	2/82(2%)
sp Q6EJ97.1 ISPS_PUELO	RecName: Full=Isoprene synthase, chloroplastic; Flags: Precursor	797221 bits(239)	3,63E-14	35/76(46%)	52/76(68%)	1/76(1%)
ref XP_003633325.1	PREDICTED: LOW QUALITY PROTEIN: isoprene synthase, chloroplastic-like [Vitis vinifera]	77411 bits(239)	1,77E-13	39/84(46%)	54/84(64%)	2/84(2%)
gb AEK70969.1	isoprene synthase, partial [Wisteria sp. 101210T2Dc1]	77411 bits(239)	1,57E-13	33/74(45%)	54/74(73%)	1/74(1%)
emb CAN62729.1	hypothetical protein VITISV_003091 [Vitis vinifera]	708626 bits(239)	2,99E-11	34/75(45%)	46/75(61%)	0/75(0%)
ref XP_002334026.1	predicted protein [Populus trichocarpa]	704774 bits(239)	3,50E-11	33/73(45%)	44/73(60%)	0/73(0%)
dbj BAF02832.1	monoterpene synthase [Eucalyptus globulus]	704774 bits(239)	3,56E-11	33/74(45%)	47/74(64%)	0/74(0%)
dbj BAF02831.1	monoterpene synthase [Eucalyptus globulus]	704774 bits(239)	3,69E-11	33/74(45%)	47/74(64%)	0/74(0%)
ref XP_002533356.1	(R)-limonene synthase [Ricinus communis]	685514 bits(239)	3,71E-11	35/77(45%)	49/77(64%)	3/77(4%)
gb AFI47928.1	terpene synthase 3 [Olea europaea]	71633 bits(239)	1,66E-11	33/75(44%)	45/75(60%)	0/75(0%)
gb ADR74206.1	(E)-beta-ocimene/myrcene synthase [Vitis vinifera]	69707 bits(239)	6,01E-11	33/75(44%)	46/75(61%)	0/75(0%)
gb AAP40638.1	putative monoterpene synthase [Melaleuca alternifolia]	712478 bits(239)	2,31E-11	32/75(43%)	46/75(61%)	0/75(0%)
gb AEP82768.1	terpene synthase [Solanum lycopersicum]	700922 bits(239)	5,10E-11	32/75(43%)	45/75(60%)	0/75(0%)
ref XP_004233802.1	PREDICTED: (-)-alpha-terpineol synthase-like [Solanum lycopersicum]	69707 bits(239)	5,27E-11	32/75(43%)	45/75(60%)	0/75(0%)
ref XP_003634879.1	PREDICTED: myrcene synthase, chloroplastic-like, partial [Vitis vinifera]	681662 bits(239)	1,06E-10	32/75(43%)	45/75(60%)	0/75(0%)
emb CBI40592.3	unnamed protein product [Vitis vinifera]	681662 bits(239)	1,95E-10	32/75(43%)	45/75(60%)	0/75(0%)
gb ACI32638.1	MTS2 [Humulus lupulus]	67781 bits(239)	3,23E-10	32/75(43%)	46/75(61%)	0/75(0%)

ref XP_003634885.1	PREDICTED: LOW QUALITY PROTEIN: myrcene synthase, chloroplastic-like [Vitis vinifera]	731738 bits(239)	5,48E-12	34/80(42%)	51/80(64%)	0/80(0%)
dbj BAG82825.1	linalool synthase [Backhousia citriodora]	693218 bits(239)	8,92E-11	32/77(42%)	48/77(62%)	2/77(3%)
ref XP_002512868.1	(R)-limonene synthase, putative [Ricinus communis]	693218 bits(239)	7,98E-11	31/75(41%)	47/75(63%)	0/75(0%)
ref XP_002329451.1	predicted protein [Populus trichocarpa]	646994 bits(239)	1,07E-10	30/73(41%)	43/73(59%)	0/73(0%)
gb AFO63173.1	chloroplast cineolsynthase, partial [Nicotiana bonariensis]	67781 bits(239)	2,38E-10	30/74(41%)	49/74(66%)	1/74(1%)
ref XP_003634881.1	PREDICTED: LOW QUALITY PROTEIN: myrcene synthase, chloroplastic-like [Vitis vinifera]	693218 bits(239)	8,84E-11	32/80(40%)	49/80(61%)	0/80(0%)
ref XP_002329449.1	predicted protein [Populus trichocarpa]	643142 bits(239)	1,52E-10	29/73(40%)	44/73(60%)	0/73(0%)
ref XP_002337303.1	predicted protein [Populus trichocarpa]	63929 bits(239)	2,37E-10	29/73(40%)	44/73(60%)	0/73(0%)
gb AEX09186.1	plastid terpeneol synthase [Nicotiana langsdorffii]	666254 bits(239)	6,90E-10	29/74(39%)	49/74(66%)	1/74(1%)
gb AFB82540.1	plastid monoterpene synthase [Nicotiana glauca]	662402 bits(239)	7,51E-10	29/74(39%)	49/74(66%)	1/74(1%)
ref XP_002875283.1	hypothetical protein ARALYDRAFT_322722 [Arabidopsis lyrata subsp. lyrata]	627734 bits(239)	3,82E-10	31/82(38%)	47/82(57%)	5/82(6%)
Contig51						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb ACD69680.1	metallothionein [Mangifera indica]	905077 bits(223)	3,13E-20	42/54(78%)	46/54(85%)	0/54(0%)
gb ADH04476.1	metallothionein [Mangifera indica]	905077 bits(223)	3,03E-20	43/56(77%)	47/56(84%)	0/56(0%)
gb ABD97258.1	metallothionin 2 [Camellia sinensis]	724034 bits(223)	1,22E-13	38/52(73%)	41/52(79%)	1/52(2%)
ref XP_002265058.1	PREDICTED: metallothionein isoform 1 [Vitis vinifera]	666254 bits(223)	1,68E-11	37/53(70%)	41/53(77%)	2/53(4%)
dbj BAA96449.1	metallothionein-like protein [Pyrus pyrifolia]	720182 bits(223)	1,89E-13	36/54(67%)	43/54(80%)	1/54(2%)
gb AFP93964.1	metallothionein type 2 [Ilex paraguariensis]	700922 bits(223)	7,70E-13	35/52(67%)	41/52(79%)	1/52(2%)
ref XP_002285180.1	PREDICTED: metallothionein-like protein type 2 [Vitis vinifera]	654698 bits(223)	3,60E-11	35/53(66%)	40/53(75%)	2/53(4%)
gb EMJ06266.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	712478 bits(223)	2,66E-13	35/54(65%)	44/54(81%)	1/54(2%)

gb EMJ06265.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	704774 bits(223)	5,50E-13	35/54(65%)	44/54(81%)	1/54(2%)
gb ABD97257.1	metallothionin 1 [Camellia sinensis]	69707 bits(223)	1,34E-12	34/52(65%)	40/52(77%)	1/52(2%)
gb ACS14665.1	metallothionein [Camellia oleifera]	685514 bits(223)	2,76E-12	34/52(65%)	40/52(77%)	1/52(2%)
gb ABN46987.1	metallothionein-like protein 2a [Nelumbo nucifera]	65855 bits(223)	3,02E-11	35/54(65%)	42/54(78%)	3/54(6%)
gb ABN46988.1	metallothionein-like protein 2b [Nelumbo nucifera]	689366 bits(223)	2,13E-12	35/55(64%)	42/55(76%)	3/55(5%)
emb CAC39481.2	metallothionein-like protein [Quercus suber]	646994 bits(223)	7,21E-11	34/53(64%)	38/53(72%)	1/53(2%)
sp P43390.1 MT2_ACTDE	RecName: Full=Metallothionein-like protein type 2	673958 bits(223)	9,03E-12	33/52(63%)	39/52(75%)	1/52(2%)
ref XP_002314182.1	predicted protein [Populus trichocarpa]	689366 bits(223)	1,37E-12	32/52(62%)	41/52(79%)	1/52(2%)
gb ABK94874.1	unknown [Populus trichocarpa]	685514 bits(223)	2,73E-12	32/52(62%)	41/52(79%)	1/52(2%)
gb AAT02525.1	metallothionein 2b [Populus trichocarpa X Populus deltoides]	67781 bits(223)	6,67E-12	32/52(62%)	40/52(77%)	1/52(2%)
gb AAY16439.1	metallothionein-like protein [Betula platyphylla]	670106 bits(223)	9,20E-12	33/53(62%)	41/53(77%)	1/53(2%)
gb ABM21762.1	metallothionein-like protein MT2B [Salix matsudana]	670106 bits(223)	1,06E-11	32/52(62%)	40/52(77%)	1/52(2%)
ref XP_002329424.1	predicted protein [Populus trichocarpa]	65855 bits(223)	1,50E-11	32/52(62%)	39/52(75%)	1/52(2%)
ref XP_002299873.1	predicted protein [Populus trichocarpa]	654698 bits(223)	4,04E-11	32/52(62%)	40/52(77%)	1/52(2%)
gb AAT02522.1	metallothionein 1a [Populus trichocarpa x Populus deltoides]	650846 bits(223)	4,36E-11	32/52(62%)	39/52(75%)	1/52(2%)
gb AEE81756.1	metallothionein [Hevea brasiliensis]	623882 bits(223)	5,27E-10	31/51(61%)	36/51(71%)	3/51(6%)
gb ABS72197.1	metallothionein 2b, partial [Corchorus olitorius]	650846 bits(223)	3,47E-11	31/52(60%)	39/52(75%)	1/52(2%)
ref XP_002330684.1	predicted protein [Populus trichocarpa]	643142 bits(223)	9,12E-11	31/52(60%)	39/52(75%)	1/52(2%)
gb AAT02523.1	metallothionein 1b [Populus trichocarpa x Populus deltoides]	643142 bits(223)	9,77E-11	31/52(60%)	39/52(75%)	1/52(2%)
emb CAA10232.1	metallothionein-like protein class II [Fagus sylvatica]	631586 bits(223)	2,83E-10	32/53(60%)	38/53(72%)	2/53(4%)
gb AFK41306.1	unknown [Lotus japonicus]	646994 bits(223)	6,86E-11	32/54(59%)	42/54(78%)	2/54(4%)
dbj BAJ39944.1	metallothionein [Ziziphus jujuba]	643142 bits(223)	9,58E-11	30/51(59%)	38/51(75%)	2/51(4%)
gb AAB88276.1	metallothionein-like protein [Prunus armeniaca]	635438 bits(223)	2,04E-10	32/54(59%)	38/54(70%)	1/54(2%)
gb ACH58424.1	metallothionein-like protein [Prunus dulcis]	635438 bits(223)	2,04E-10	32/54(59%)	38/54(70%)	1/54(2%)
gb ADM86706.1	metallothionein-like protein [Cicer microphyllum]	646994 bits(223)	8,28E-11	31/53(58%)	40/53(75%)	2/53(4%)
dbj BAD18375.1	type 2 metallothionein [Vigna radiata var. radiata]	63929 bits(223)	1,49E-10	31/53(58%)	38/53(72%)	2/53(4%)

dbj BAD18379.1	type 2 metallothionein [<i>Vigna angularis</i>]	631586 bits(223)	2,75E-10	30/53(57%)	38/53(72%)	2/53(4%)
gb ABQ44281.1	metallothionein type 2 [<i>Sesbania drummondii</i>]	62003 bits(223)	7,44E-10	30/53(57%)	38/53(72%)	2/53(4%)
sp Q39459.2 MT2_CICAR	RecName: Full=Metallothionein-like protein 2; Short=MT-2	616178 bits(223)	9,16E-10	30/53(57%)	39/53(74%)	2/53(4%)
ref XP_002336333.1	predicted protein [<i>Populus trichocarpa</i>]	612326 bits(223)	7,58E-10	29/52(56%)	38/52(73%)	1/52(2%)
Contig53						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_004294730.1	PREDICTED: non-specific lipid-transfer protein 1-like [<i>Fragaria vesca</i> subsp. <i>vesca</i>]	134035 bits(336)	1,39E-34	67/123(54%)	90/123(73%)	4/123(3%)
gb EMJ07354.1	hypothetical protein PRUPE_ppa013519mg [<i>Prunus persica</i>]	119013 bits(336)	4,03E-29	59/111(53%)	77/111(69%)	1/111(1%)
gb AFK48111.1	unknown [<i>Lotus japonicus</i>]	125561 bits(336)	1,65E-31	60/117(51%)	79/117(68%)	3/117(3%)
gb AFM35695.1	nonspecific lipid transfer protein [<i>Vitis pseudoreticulata</i>]	108227 bits(336)	2,49E-25	49/96(51%)	67/96(70%)	0/96(0%)
gb AAT68263.1	lipid transfer protein [<i>Nicotiana glauca</i>]	107457 bits(336)	5,03E-25	47/93(51%)	64/93(69%)	0/93(0%)
sp P85894.1 LTP1_MORNI	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Allergen=Mor n 3	100908 bits(336)	5,10E-23	46/90(51%)	61/90(68%)	0/90(0%)
gb ACJ83942.1	unknown [<i>Medicago truncatula</i>]	115546 bits(336)	5,61E-28	57/115(50%)	75/115(65%)	4/115(3%)
gb AAS13435.1	lipid-transfer protein [<i>Nicotiana attenuata</i>]	105531 bits(336)	2,47E-24	47/94(50%)	63/94(67%)	0/94(0%)
gb AEI70828.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	103219 bits(336)	1,40E-23	45/90(50%)	62/90(69%)	0/90(0%)
emb CAH03799.1	lipid transfer protein [<i>Citrus sinensis</i>]	939745 bits(336)	1,74E-20	45/90(50%)	60/90(67%)	0/90(0%)
gb EMJ08287.1	hypothetical protein PRUPE_ppa023836mg [<i>Prunus persica</i>]	115546 bits(336)	7,70E-28	59/121(49%)	78/121(64%)	3/121(2%)
gb AAF28385.1 AF151214_1	lipid-transfer protein [<i>Nicotiana glauca</i>]	105145 bits(336)	2,89E-24	46/93(49%)	62/93(67%)	0/93(0%)
gb AAT68264.1	lipid transfer protein [<i>Nicotiana glauca</i>]	105145 bits(336)	3,45E-24	46/93(49%)	62/93(67%)	0/93(0%)
gb AAT45202.1	lipid transfer protein 1 precursor [<i>Nicotiana tabacum</i>]	103219 bits(336)	1,80E-23	46/93(49%)	61/93(66%)	0/93(0%)
dbj BAK19152.1	lipid transfer protein [<i>Nicotiana tabacum</i>]	102834 bits(336)	2,15E-23	46/93(49%)	61/93(66%)	0/93(0%)
gb AAO33394.1 AF467946_1	lipid transfer protein isoform 4 [<i>Vitis vinifera</i>]	100908 bits(336)	1,06E-22	44/90(49%)	63/90(70%)	0/90(0%)

gb AEI70827.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	985969 bits(336)	6,72E-22	44/90(49%)	59/90(66%)	0/90(0%)
gb AAO33357.1 AF465408_1	nonspecific lipid transfer protein 1 [<i>Vitis cinerea</i> var. <i>helleri</i> x <i>Vitis vinifera</i>]	978265 bits(336)	1,19E-21	44/90(49%)	62/90(69%)	0/90(0%)
ref XP_003553202.1	PREDICTED: non-specific lipid-transfer protein-like [<i>Glycine max</i>]	112079 bits(336)	1,17E-26	55/115(48%)	77/115(67%)	3/115(3%)
gb AAF35184.1 AF195863_1	lipid transfer protein precursor [<i>Gossypium hirsutum</i>]	107457 bits(336)	4,60E-25	59/124(48%)	77/124(62%)	7/124(6%)
gb AAT68265.1	lipid transfer protein precursor [<i>Nicotiana glauca</i>]	993673 bits(336)	3,21E-22	43/89(48%)	60/89(67%)	0/89(0%)
gb AEI70826.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	989821 bits(336)	4,84E-22	45/93(48%)	63/93(68%)	2/93(2%)
ref XP_002271116.1	PREDICTED: non-specific lipid-transfer protein [<i>Vitis vinifera</i>]	989821 bits(336)	5,01E-22	44/91(48%)	62/91(68%)	0/91(0%)
emb CBW38504.1	lipid transfer protein [<i>Helianthus annuus</i>]	978265 bits(336)	9,40E-22	43/90(48%)	59/90(66%)	0/90(0%)
emb CBW38503.1	lipid transfer protein [<i>Helianthus annuus</i>]	974413 bits(336)	1,05E-21	43/90(48%)	59/90(66%)	0/90(0%)
emb CBW38499.1	lipid transfer protein [<i>Helianthus annuus</i>]	974413 bits(336)	1,33E-21	43/90(48%)	59/90(66%)	0/90(0%)
gb AEI70829.1	non-specific lipid-transfer protein [<i>Helianthus annuus</i>]	974413 bits(336)	1,62E-21	43/90(48%)	59/90(66%)	0/90(0%)
emb CBW38497.1	lipid transfer protein [<i>Helianthus annuus</i>]	970561 bits(336)	1,65E-21	43/90(48%)	59/90(66%)	0/90(0%)
emb CAB96874.1	mal d 3 [<i>Malus x domestica</i>]	924337 bits(336)	5,39E-20	43/90(48%)	58/90(64%)	0/90(0%)
gb AAF61436.1 AF137353_1	lipid transfer protein precursor [<i>Pisum sativum</i>]	11362 bits(336)	2,64E-27	55/116(47%)	75/116(65%)	3/116(3%)
gb AAT68262.1	lipid transfer protein [<i>Nicotiana glauca</i>]	110538 bits(336)	3,37E-26	57/122(47%)	74/122(61%)	6/122(5%)
gb AGH13363.1	lipid transfer protein, partial [<i>Helianthus annuus</i>]	970561 bits(336)	1,49E-21	42/90(47%)	60/90(67%)	0/90(0%)
emb CBW38501.1	lipid transfer protein [<i>Helianthus annuus</i>]	970561 bits(336)	1,75E-21	42/90(47%)	60/90(67%)	0/90(0%)
ref XP_004245220.1	PREDICTED: non-specific lipid-transfer protein 3-like [<i>Solanum lycopersicum</i>]	974413 bits(336)	2,14E-21	45/95(47%)	63/95(66%)	1/95(1%)
emb CBW38498.1	lipid transfer protein [<i>Helianthus annuus</i>]	966709 bits(336)	2,70E-21	42/90(47%)	60/90(67%)	0/90(0%)
gb ABA29446.1	non-specific lipid transfer protein [<i>Vitis vinifera</i>]	959005 bits(336)	7,61E-21	43/91(47%)	61/91(67%)	0/91(0%)
sp P82534.1 NLTP1_PRUDO	RecName: Full=Non-specific lipid-transfer protein 1; Short=LTP 1; AltName: Full=Major allergen Pru d 3; AltName: Allergen=Pur d 3	932041 bits(336)	3,77E-20	42/90(47%)	60/90(67%)	0/90(0%)
ref XP_002276372.1	PREDICTED: non-specific lipid-transfer protein [<i>Vitis vinifera</i>]	122094 bits(336)	2,82E-30	56/123(46%)	81/123(66%)	4/123(3%)
gb AAL27855.1	lipid transfer protein precursor [<i>Davidia involucrata</i>]	112079 bits(336)	1,26E-26	54/117(46%)	78/117(67%)	4/117(3%)
gb AAF35186.1 AF195865_1	lipid transfer protein precursor [<i>Gossypium hirsutum</i>]	107842 bits(336)	3,78E-25	57/124(46%)	76/124(61%)	7/124(6%)
gb AEN70917.1	lipid transfer protein [<i>Gossypium schwendimanii</i>]	105531 bits(336)	2,24E-24	57/123(46%)	76/123(62%)	7/123(6%)

gb AEN70919.1	lipid transfer protein [Gossypium mustelinum]	103219 bits(336)	1,42E-23	57/123(46%)	77/123(63%)	7/123(6%)
gb AEN70935.1	lipid transfer protein [Gossypium aridum]	102834 bits(336)	2,56E-23	56/123(46%)	75/123(61%)	7/123(6%)
gb AEN70933.1	lipid transfer protein [Gossypium davidsonii]	102449 bits(336)	3,48E-23	56/123(46%)	76/123(62%)	7/123(6%)
gb ADC80503.1	non-specific lipid transfer protein 1b precursor [Phaseolus vulgaris]	978265 bits(336)	1,27E-21	44/95(46%)	63/95(66%)	0/95(0%)
sp Q39950.1 NLTP_HELAN	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; Short=NslTP; AltName: Full=SDI-9; Flags: Precursor	970561 bits(336)	2,75E-21	42/91(46%)	59/91(65%)	0/91(0%)
sp Q39794.1 NLTP_GERHY	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; Flags: Precursor	966709 bits(336)	3,87E-21	52/112(46%)	68/112(61%)	1/112(1%)
gb AEI70830.1	non-specific lipid-transfer protein [Helianthus annuus]	962857 bits(336)	4,36E-21	42/91(46%)	59/91(65%)	0/91(0%)
ref XP_002305878.1	predicted protein [Populus trichocarpa]	962857 bits(336)	4,45E-21	43/94(46%)	62/94(66%)	0/94(0%)
gb ABK96515.1	unknown [Populus trichocarpa x Populus deltoides]	962857 bits(336)	5,44E-21	43/94(46%)	61/94(65%)	0/94(0%)
gb AAK28533.1 AF329829_1	lipid transfer protein precursor [Corylus avellana]	955153 bits(336)	8,43E-21	42/92(46%)	65/92(71%)	0/92(0%)
gb AEC04836.1	lipid transfer protein [Dimocarpus longan]	928189 bits(336)	7,26E-20	43/93(46%)	59/93(63%)	0/93(0%)
gb AAV43800.1	FSLTP2 [Gossypium hirsutum]	102449 bits(336)	2,97E-23	55/123(45%)	76/123(62%)	7/123(6%)
gb AAG29777.1	lipid transfer protein 3 precursor [Gossypium hirsutum]	102064 bits(336)	3,95E-23	56/124(45%)	75/124(60%)	7/124(6%)
gb ADY68820.1	lipid transfer protein 3 [Gossypium herbaceum subsp. africanum]	101293 bits(336)	8,54E-23	55/123(45%)	75/123(61%)	7/123(6%)
gb AEN70918.1	lipid transfer protein [Gossypium turneri]	100908 bits(336)	1,04E-22	55/123(45%)	75/123(61%)	7/123(6%)
gb AEN70915.1	lipid transfer protein [Gossypium thurberi]	993673 bits(336)	3,39E-22	55/123(45%)	74/123(60%)	7/123(6%)
gb AAV43801.1	FSLTP3 [Gossypium hirsutum]	993673 bits(336)	4,13E-22	53/117(45%)	73/117(62%)	7/117(6%)
gb AEN70916.1	lipid transfer protein [Gossypium laxum]	989821 bits(336)	4,70E-22	55/123(45%)	74/123(60%)	7/123(6%)
sp P23802.1 NLTP_ELECO	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Full=Alpha-amylase inhibitor I-2	924337 bits(336)	7,01E-20	42/93(45%)	64/93(69%)	2/93(2%)
gb ACQ42221.1	lipid transfer protein [Chimonanthus praecox]	107457 bits(336)	4,95E-25	51/116(44%)	77/116(66%)	2/116(2%)
gb AAM21292.1 AF369931_1	lipid-transfer protein [Citrus sinensis]	102834 bits(336)	2,21E-23	49/112(44%)	71/112(63%)	3/112(3%)
emb CBW38502.1	lipid transfer protein [Helianthus annuus]	989821 bits(336)	3,49E-22	45/102(44%)	65/102(64%)	2/102(2%)
gb ADR66942.1	non-specific lipid transfer protein [Prunus avium]	955153 bits(336)	1,03E-20	40/90(44%)	61/90(68%)	0/90(0%)
sp Q9M5X8.1 NLTP_PRUAV	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Allergen=Pru av 3; Flags: Precursor	951301 bits(336)	1,08E-20	40/90(44%)	61/90(68%)	0/90(0%)

sp Q43129.1 NLTP2_GOSHI	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Full=GH3; Flags: Precursor	947449 bits(336)	1,59E-20	54/124(44%)	73/124(59%)	7/124(6%)
gb ADR66943.1	non-specific lipid transfer protein [<i>Prunus avium</i>]	943597 bits(336)	2,74E-20	40/90(44%)	61/90(68%)	0/90(0%)
gb ACP43542.1	lipid transfer protein [<i>Chimonanthus praecox</i>]	105145 bits(336)	3,01E-24	49/114(43%)	73/114(64%)	2/114(2%)
gb ABB96481.1	non-specific lipid transfer protein precursor [<i>Malus x domestica</i>]	974413 bits(336)	1,59E-21	49/114(43%)	69/114(61%)	0/114(0%)
gb AAR22488.1	allergen Mal d 3 [<i>Malus x domestica</i>]	993673 bits(336)	3,08E-22	51/122(42%)	72/122(59%)	8/122(7%)
gb AEN70936.1	lipid transfer protein [<i>Gossypium gossypioides</i>]	997525 bits(336)	3,17E-22	51/120(42%)	72/120(60%)	4/120(3%)
gb AAT80649.1	lipid transfer protein precursor [<i>Malus x domestica</i>]	989821 bits(336)	4,36E-22	51/122(42%)	72/122(59%)	8/122(7%)
sp Q9M5X7.1 NLTP_MALDO	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Allergen=Mal d 3; Flags: Precursor	989821 bits(336)	5,01E-22	51/122(42%)	71/122(58%)	8/122(7%)
ref XP_003517010.1	PREDICTED: non-specific lipid-transfer protein 1-like isoform 1 [<i>Glycine max</i>]	985969 bits(336)	8,20E-22	49/116(42%)	74/116(64%)	3/116(3%)
dbj BAH03575.1	nonspecific lipid transfer protein [<i>Citrus jambhiri</i>]	959005 bits(336)	5,87E-21	47/112(42%)	69/112(62%)	3/112(3%)
gb AAF35185.1 AF195864_1	lipid transfer protein precursor [<i>Gossypium hirsutum</i>]	997525 bits(336)	2,95E-22	50/122(41%)	69/122(57%)	5/122(4%)
gb AAL32039.1 AF439280_1	lipid transfer protein-like protein [<i>Retama raetam</i>]	974413 bits(336)	1,83E-21	47/116(41%)	72/116(62%)	4/116(3%)
gb AAQ96338.1	lipid transfer protein [<i>Vitis aestivalis</i>]	974413 bits(336)	2,11E-21	47/115(41%)	75/115(65%)	2/115(2%)
gb ADR66950.1	non-specific lipid transfer protein [<i>Prunus sargentii</i>]	951301 bits(336)	1,08E-20	47/116(41%)	75/116(65%)	2/116(2%)
gb ABK41612.1	lipid transfer protein [<i>Lilium longiflorum</i>]	947449 bits(336)	1,25E-20	47/116(41%)	71/116(61%)	8/116(7%)
gb ABY21307.1	pollen allergen Pla o 3 [<i>Platanus orientalis</i>]	951301 bits(336)	1,39E-20	50/122(41%)	69/122(57%)	5/122(4%)
gb AAV49759.1	non-specific lipid transfer protein 6 [<i>Hordeum vulgare</i> subsp. <i>vulgare</i>]	951301 bits(336)	1,60E-20	51/124(41%)	72/124(58%)	1/124(1%)
gb ADR66949.1	non-specific lipid transfer protein [<i>Prunus sargentii</i>]	947449 bits(336)	1,65E-20	47/116(41%)	75/116(65%)	2/116(2%)
gb ABF06565.1	non-specific lipid transfer-like protein [<i>Prosopis juliflora</i>]	101679 bits(336)	4,64E-23	48/120(40%)	75/120(62%)	5/120(4%)
ref XP_004302608.1	PREDICTED: non-specific lipid-transfer protein-like isoform 1 [<i>Fragaria vesca</i> subsp. <i>vesca</i>]	993673 bits(336)	3,19E-22	47/117(40%)	73/117(62%)	6/117(5%)
gb ADR66953.1	non-specific lipid transfer protein [<i>Prunus dulcis</i>]	951301 bits(336)	1,08E-20	46/116(40%)	75/116(65%)	2/116(2%)
ref XP_004302622.1	PREDICTED: non-specific lipid-transfer protein 3-like [<i>Fragaria vesca</i> subsp. <i>vesca</i>]	947449 bits(336)	1,78E-20	50/124(40%)	70/124(56%)	5/124(4%)
emb CAL07989.1	lipid transfer protein [<i>Platanus x acerifolia</i>]	939745 bits(336)	3,33E-20	49/122(40%)	69/122(57%)	5/122(4%)
gb AAT80665.1	lipid transfer protein precursor [<i>Malus x domestica</i>]	928189 bits(336)	7,11E-20	46/114(40%)	69/114(61%)	0/114(0%)

gb AAY83345.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	100908 bits(336)	8,91E-23	47/122(39%)	77/122(63%)	6/122(5%)
ref XP_004302624.1	PREDICTED: non-specific lipid-transfer protein 1-like [Fragaria vesca subsp. vesca]	989821 bits(336)	4,72E-22	47/122(39%)	76/122(62%)	6/122(5%)
gb ABG54494.1	putative allergen Rub i 3 [Rubus idaeus]	974413 bits(336)	1,56E-21	48/122(39%)	73/122(60%)	6/122(5%)
gb AGH27864.1	lipid transfer protein precursor [Prunus davidiana]	974413 bits(336)	1,71E-21	47/122(39%)	74/122(61%)	6/122(5%)
sp Q9M5X6.1 NLTP_PYRCO	RecName: Full=Non-specific lipid-transfer protein; Short=LTP; AltName: Allergen=Pyr c 3; Flags: Precursor	959005 bits(336)	6,11E-21	47/122(39%)	71/122(58%)	8/122(7%)
ref NP_001236948.1	uncharacterized protein LOC100306062 precursor [Glycine max]	962857 bits(336)	6,46E-21	49/125(39%)	75/125(60%)	12/125(10%)
gb AAY83343.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	997525 bits(336)	2,65E-22	46/122(38%)	76/122(62%)	6/122(5%)
gb AAY83344.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	962857 bits(336)	5,36E-21	46/122(38%)	74/122(61%)	6/122(5%)
emb CAC86258.1	lipid transfer protein [Fragaria x ananassa]	939745 bits(336)	3,09E-20	46/122(38%)	74/122(61%)	6/122(5%)
gb AAY83342.1	non-specific lipid transfer protein precursor [Fragaria x ananassa]	932041 bits(336)	6,41E-20	46/122(38%)	73/122(60%)	6/122(5%)
Contig54						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref NP_001237382.1	uncharacterized protein LOC100500139 [Glycine max]	140584 bits(400)	1,48E-36	82/95(86%)	89/95(94%)	1/95(1%)
ref XP_002298508.1	predicted protein [Populus trichocarpa]	158688 bits(400)	3,27E-43	82/96(85%)	88/96(92%)	1/96(1%)
ref XP_002532758.1	conserved hypothetical protein [Ricinus communis]	140198 bits(400)	1,96E-36	76/90(84%)	81/90(90%)	3/90(3%)
ref XP_003608407.1	hypothetical protein MTR_4g093770 [Medicago truncatula]	804925 bits(400)	3,92E-15	54/64(84%)	61/64(95%)	1/64(2%)
ref XP_002313962.1	predicted protein [Populus trichocarpa]	155221 bits(400)	6,39E-42	80/96(83%)	87/96(91%)	1/96(1%)
gb AFK47672.1	unknown [Medicago truncatula]	138272 bits(400)	1,13E-35	79/95(83%)	89/95(94%)	1/95(1%)
ref NP_001236099.1	uncharacterized protein LOC100499791 [Glycine max]	136732 bits(400)	4,00E-35	77/95(81%)	90/95(95%)	1/95(1%)
emb CAN75425.1	hypothetical protein VITISV_019911 [Vitis vinifera]	128257 bits(400)	4,67E-32	69/85(81%)	78/85(92%)	1/85(1%)

gb EMJ03052.1	hypothetical protein PRUPE_ppa012968mg [Prunus persica]	147132 bits(400)	4,62E-39	75/94(80%)	88/94(94%)	0/94(0%)
gb EMJ03053.1	hypothetical protein PRUPE_ppa012968mg [Prunus persica]	147132 bits(400)	6,96E-39	75/94(80%)	88/94(94%)	0/94(0%)
ref XP_002269289.1	PREDICTED: uncharacterized protein C6C3.02c [Vitis vinifera]	145976 bits(400)	1,39E-38	75/96(78%)	86/96(90%)	1/96(1%)
tpg DAA50710.1	TPA: hypothetical protein ZEAMMB73_985373 [Zea mays]	897373 bits(400)	2,42E-18	50/64(78%)	52/64(81%)	3/64(5%)
gb ACG40676.1	p8MTCP1 [Zea mays]	136346 bits(400)	1,36E-35	72/93(77%)	79/93(85%)	3/93(3%)
gb ACG37552.1	p8MTCP1 [Zea mays]	137887 bits(400)	1,47E-35	72/93(77%)	79/93(85%)	3/93(3%)
tpg DAA50711.1	TPA: p8MTCP1 [Zea mays]	137502 bits(400)	1,61E-35	72/93(77%)	79/93(85%)	3/93(3%)
gb ACG30030.1	p8MTCP1 [Zea mays]	137117 bits(400)	2,73E-35	72/93(77%)	79/93(85%)	3/93(3%)
tpg DAA50708.1	TPA: p8MTCP1, partial [Zea mays]	138658 bits(400)	4,98E-35	72/93(77%)	79/93(85%)	3/93(3%)
ref XP_004290403.1	PREDICTED: uncharacterized protein C6C3.02c-like [Fragaria vesca subsp. vesca]	135191 bits(400)	1,16E-34	70/91(77%)	79/91(87%)	5/91(5%)
ref NP_001150231.1	p8MTCP1 [Zea mays]	134806 bits(400)	1,48E-34	71/93(76%)	78/93(84%)	3/93(3%)
gb ACG31202.1	p8MTCP1 [Zea mays]	132494 bits(400)	1,10E-33	71/93(76%)	78/93(84%)	3/93(3%)
tpg DAA50712.1	TPA: hypothetical protein ZEAMMB73_985373 [Zea mays]	127102 bits(400)	2,20E-32	67/88(76%)	74/88(84%)	3/88(3%)
ref NP_001050926.1	Os03g0685500 [Oryza sativa Japonica Group]	139043 bits(400)	5,11E-36	70/93(75%)	80/93(86%)	2/93(2%)
gb EEC75957.1	hypothetical protein Osl_13062 [Oryza sativa Indica Group]	138272 bits(400)	7,04E-36	70/93(75%)	80/93(86%)	2/93(2%)
ref XP_002464069.1	hypothetical protein SORBIDRAFT_01g011740 [Sorghum bicolor]	13288 bits(400)	9,88E-34	70/93(75%)	77/93(83%)	3/93(3%)
ref XP_004248792.1	PREDICTED: uncharacterized protein C6C3.02c-like isoform 2 [Solanum lycopersicum]	119398 bits(400)	4,59E-29	67/90(74%)	80/90(89%)	2/90(2%)
ref XP_004248791.1	PREDICTED: uncharacterized protein C6C3.02c-like isoform 1 [Solanum lycopersicum]	119398 bits(400)	5,54E-29	67/90(74%)	80/90(89%)	2/90(2%)
gb ACF84231.1	unknown [Zea mays]	119013 bits(400)	3,06E-29	64/88(73%)	75/88(85%)	3/88(3%)
ref NP_001140638.1	uncharacterized protein LOC100272713 [Zea mays]	120168 bits(400)	3,27E-29	64/88(73%)	75/88(85%)	3/88(3%)
dbj BAJ97723.1	predicted protein [Hordeum vulgare subsp. vulgare]	808777 bits(400)	3,64E-15	47/64(73%)	49/64(77%)	3/64(5%)
ref XP_002285635.1	PREDICTED: uncharacterized protein C6C3.02c [Vitis vinifera]	148673 bits(400)	1,67E-39	69/96(72%)	85/96(89%)	1/96(1%)
emb CAN72156.1	hypothetical protein VITISV_019019 [Vitis vinifera]	148288 bits(400)	2,40E-39	69/96(72%)	84/96(88%)	1/96(1%)
ref XP_002864899.1	hypothetical protein ARALYDRAFT_496648 [Arabidopsis lyrata subsp. lyrata]	128257 bits(400)	4,69E-32	69/96(72%)	83/96(86%)	3/96(3%)

tpg DAA50718.1	TPA: hypothetical protein ZEAMMB73_419715 [Zea mays]	11439 bits(400)	1,40E-27	64/89(72%)	75/89(84%)	4/89(4%)
gb ACG31028.1	p8MTCP1 [Zea mays]	100908 bits(400)	9,46E-23	52/72(72%)	58/72(81%)	3/72(4%)
ref NP_001149273.1	p8MTCP1 [Zea mays]	11439 bits(400)	4,45E-27	59/83(71%)	69/83(83%)	3/83(4%)
ref NP_201245.1	uncharacterized protein [Arabidopsis thaliana]	126716 bits(400)	1,48E-31	67/96(70%)	78/96(81%)	3/96(3%)
ref XP_004248937.1	PREDICTED: uncharacterized protein C6C3.02c-like [Solanum lycopersicum]	12402 bits(400)	3,43E-30	63/90(70%)	72/90(80%)	2/90(2%)
gb AEJ88279.1	putative CHCH domain protein [Wolffia australiana]	120939 bits(400)	1,77E-29	67/96(70%)	78/96(81%)	5/96(5%)
gb AAS21010.1	unknown [Hyacinthus orientalis]	115931 bits(400)	1,39E-27	54/77(70%)	65/77(84%)	0/77(0%)
ref XP_003561192.1	PREDICTED: uncharacterized protein LOC100823022 [Brachypodium distachyon]	128257 bits(400)	4,31E-32	66/95(69%)	76/95(80%)	3/95(3%)
dbj BAJ96007.1	predicted protein [Hordeum vulgare subsp. vulgare]	125946 bits(400)	2,69E-31	66/95(69%)	74/95(78%)	3/95(3%)
ref XP_004152606.1	PREDICTED: uncharacterized protein C6C3.02c-like isoform 1 [Cucumis sativus]	134035 bits(400)	3,87E-34	64/96(67%)	83/96(86%)	2/96(2%)
gb ADB08696.1	hypothetical protein [Wolffia arrhiza]	120939 bits(400)	1,60E-29	64/96(67%)	78/96(81%)	5/96(5%)
ref XP_004152607.1	PREDICTED: uncharacterized protein C6C3.02c-like isoform 2 [Cucumis sativus]	828037 bits(400)	6,10E-16	43/64(67%)	54/64(84%)	2/64(3%)
ref NP_001032137.1	uncharacterized protein [Arabidopsis thaliana]	727886 bits(400)	4,71E-12	42/64(66%)	47/64(73%)	3/64(5%)
ref XP_002871376.1	hypothetical protein ARALYDRAFT_908906 [Arabidopsis lyrata subsp. lyrata]	125946 bits(400)	1,79E-31	62/96(65%)	78/96(81%)	5/96(5%)
ref XP_004292470.1	PREDICTED: uncharacterized protein LOC101293556 [Fragaria vesca subsp. vesca]	117087 bits(400)	5,38E-28	62/96(65%)	77/96(80%)	12/96(12%)
ref XP_002971196.1	hypothetical protein SELMODRAFT_147481 [Selaginella moellendorffii]	109768 bits(400)	1,53E-25	60/93(65%)	72/93(77%)	3/93(3%)
ref XP_002961605.1	hypothetical protein SELMODRAFT_76418 [Selaginella moellendorffii]	109383 bits(400)	4,11E-25	60/93(65%)	72/93(77%)	3/93(3%)
ref XP_002527365.1	conserved hypothetical protein [Ricinus communis]	114005 bits(400)	1,93E-27	56/89(63%)	66/89(74%)	4/89(4%)
ref NP_196519.1	Cox19-like CHCH family protein [Arabidopsis thaliana]	125561 bits(400)	3,93E-31	59/96(61%)	76/96(79%)	8/96(8%)
gb ABK23486.1	unknown [Picea sitchensis]	118242 bits(400)	1,71E-28	56/92(61%)	74/92(80%)	2/92(2%)
gb ABK22607.1	unknown [Picea sitchensis]	118242 bits(400)	1,86E-28	56/92(61%)	74/92(80%)	2/92(2%)
ref XP_001776040.1	predicted protein [Physcomitrella patens subsp. patens]	109383 bits(400)	2,26E-25	54/88(61%)	67/88(76%)	1/88(1%)
ref NP_001235122.1	uncharacterized protein LOC100305931 [Glycine max]	12479 bits(400)	5,99E-31	58/96(60%)	77/96(80%)	4/96(4%)
gb ABK21201.1	unknown [Picea sitchensis]	118242 bits(400)	2,00E-28	56/94(60%)	74/94(79%)	1/94(1%)
ref XP_002963063.1	hypothetical protein SELMODRAFT_404614 [Selaginella moellendorffii]	997525 bits(400)	5,59E-22	55/92(60%)	64/92(70%)	7/92(8%)

ref XP_003608406.1	hypothetical protein MTR_4g093770 [Medicago truncatula]	118627 bits(400)	3,18E-28	79/133(59%)	90/133(68%)	39/133(29%)
ref XP_001785264.1	predicted protein [Physcomitrella patens subsp. patens]	103219 bits(400)	3,21E-23	51/88(58%)	64/88(73%)	3/88(3%)
ref XP_002980272.1	hypothetical protein SELMODRAFT_419954 [Selaginella moellendorffii]	993673 bits(400)	8,04E-22	53/92(58%)	64/92(70%)	7/92(8%)
ref XP_001758905.1	predicted protein [Physcomitrella patens subsp. patens]	978265 bits(400)	1,27E-21	54/93(58%)	64/93(69%)	8/93(9%)
ref XP_003593279.1	hypothetical protein MTR_2g009710 [Medicago truncatula]	862705 bits(400)	8,47E-17	52/90(58%)	67/90(74%)	2/90(2%)
ref XP_002981832.1	hypothetical protein SELMODRAFT_17960 [Selaginella moellendorffii]	889669 bits(400)	1,53E-18	46/83(55%)	57/83(69%)	2/83(2%)
gb ABD28551.1	CHCH [Medicago truncatula]	843445 bits(400)	4,36E-16	52/94(55%)	66/94(70%)	6/94(6%)
gb ABD28550.1	hypothetical protein MtrDRAFT_AC148971g12v2 [Medicago truncatula]	739442 bits(400)	1,40E-12	45/86(52%)	55/86(64%)	7/86(8%)
ref XP_003593278.1	hypothetical protein MTR_2g009700 [Medicago truncatula]	75485 bits(400)	8,16E-13	45/88(51%)	56/88(64%)	7/88(8%)
ref XP_002901744.1	conserved hypothetical protein [Phytophthora infestans T30-4]	654698 bits(400)	6,07E-10	39/82(48%)	50/82(61%)	1/82(1%)
gb EKX36456.1	hypothetical protein GUIHDRAFT_155227 [Guillardia theta CCMP2712]	797221 bits(400)	7,31E-15	42/92(46%)	58/92(63%)	0/92(0%)
ref XP_001694892.1	predicted protein [Chlamydomonas reinhardtii]	700922 bits(400)	2,06E-11	41/89(46%)	54/89(61%)	3/89(3%)
ref XP_002896713.1	conserved hypothetical protein [Phytophthora infestans T30-4]	681662 bits(400)	1,48E-10	42/91(46%)	58/91(64%)	3/91(3%)
ref XP_002947810.1	hypothetical protein VOLCADRAFT_103588 [Volvox carteri f. nagariensis]	71633 bits(400)	1,52E-11	39/90(43%)	56/90(62%)	2/90(2%)
ref XP_003578628.1	PREDICTED: uncharacterized protein LOC100838236 [Brachypodium distachyon]	770258 bits(400)	1,21E-13	39/93(42%)	58/93(62%)	8/93(9%)
ref NP_001067375.1	Os12g0637900 [Oryza sativa Japonica Group]	670106 bits(400)	4,67E-10	35/85(41%)	51/85(60%)	0/85(0%)
gb EGZ23984.1	hypothetical protein PHYSODRAFT_556628 [Phytophthora sojae]	65855 bits(400)	4,09E-10	38/94(40%)	51/94(54%)	7/94(7%)
ref XP_002863107.1	hypothetical protein ARALYDRAFT_359144 [Arabidopsis lyrata subsp. lyrata]	770258 bits(400)	1,56E-13	37/97(38%)	53/97(55%)	5/97(5%)

Contig55

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
ref XP_002335751.1	predicted protein [Populus trichocarpa]	159073 bits(417)	1,70E-43	76/128(59%)	99/128(77%)	1/128(1%)
ref XP_002297701.1	predicted protein [Populus trichocarpa]	147517 bits(417)	6,83E-39	80/135(59%)	104/135(77%)	5/135(4%)
ref XP_002527846.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	162925 bits(417)	1,19E-44	77/134(57%)	103/134(77%)	1/134(1%)
ref XP_002264151.2	PREDICTED: sulfotransferase 16-like [Vitis vinifera]	156762 bits(417)	4,56E-42	75/135(56%)	105/135(78%)	5/135(4%)
ref XP_002334144.1	predicted protein [Populus trichocarpa]	140198 bits(417)	3,01E-36	74/133(56%)	99/133(74%)	1/133(1%)
emb CAN74617.1	hypothetical protein VITISV_002219 [Vitis vinifera]	157918 bits(417)	1,74E-43	75/136(55%)	103/136(76%)	1/136(1%)
ref XP_002264574.1	PREDICTED: flavonol sulfotransferase-like [Vitis vinifera]	157532 bits(417)	1,61E-42	75/136(55%)	103/136(76%)	1/136(1%)
emb CBI35080.3	unnamed protein product [Vitis vinifera]	157147 bits(417)	2,03E-42	75/136(55%)	103/136(76%)	1/136(1%)
ref XP_002263832.2	PREDICTED: LOW QUALITY PROTEIN: flavonol sulfotransferase-like [Vitis vinifera]	154836 bits(417)	2,71E-41	73/132(55%)	100/132(76%)	1/132(1%)
ref XP_002334686.1	predicted protein [Populus trichocarpa]	150984 bits(417)	4,55E-40	76/137(55%)	101/137(74%)	2/137(1%)
ref XP_002303966.1	predicted protein [Populus trichocarpa]	147132 bits(417)	1,18E-38	75/137(55%)	100/137(73%)	2/137(1%)
ref XP_002334685.1	predicted protein [Populus trichocarpa]	144436 bits(417)	2,04E-38	80/145(55%)	107/145(74%)	2/145(1%)
ref XP_002297700.1	predicted protein [Populus trichocarpa]	141739 bits(417)	7,13E-37	81/146(55%)	108/146(74%)	3/146(2%)
emb CAN77893.1	hypothetical protein VITISV_030448 [Vitis vinifera]	150984 bits(417)	3,77E-40	71/132(54%)	99/132(75%)	1/132(1%)
ref XP_002337045.1	predicted protein [Populus trichocarpa]	146747 bits(417)	1,63E-38	73/136(54%)	94/136(69%)	1/136(1%)
emb CAN83405.1	hypothetical protein VITISV_022677 [Vitis vinifera]	13442 bits(417)	2,76E-33	59/110(54%)	84/110(76%)	3/110(3%)
emb CAN62733.1	hypothetical protein VITISV_015318 [Vitis vinifera]	165236 bits(417)	3,70E-46	82/155(53%)	113/155(73%)	2/155(1%)
ref XP_002303947.1	predicted protein [Populus trichocarpa]	144821 bits(417)	8,21E-38	72/137(53%)	93/137(68%)	1/137(1%)
ref XP_003542828.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	140198 bits(417)	4,68E-36	72/137(53%)	93/137(68%)	2/137(1%)
ref XP_003541597.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	139813 bits(417)	6,04E-36	71/135(53%)	93/135(69%)	2/135(1%)
emb CBI15344.3	unnamed protein product [Vitis vinifera]	120168 bits(417)	8,93E-31	56/106(53%)	78/106(74%)	3/106(3%)
gb AFK33848.1	unknown [Lotus japonicus]	136732 bits(417)	9,10E-35	69/136(51%)	90/136(66%)	1/136(1%)
ref XP_004250849.1	PREDICTED: cytosolic sulfotransferase 12-like [Solanum lycopersicum]	132494 bits(417)	3,52E-33	67/132(51%)	89/132(67%)	1/132(1%)
ref XP_004253506.1	PREDICTED: cytosolic sulfotransferase 12-like [Solanum lycopersicum]	130954 bits(417)	1,06E-32	67/132(51%)	88/132(67%)	1/132(1%)

ref XP_003633588.1	PREDICTED: LOW QUALITY PROTEIN: sulfotransferase 17-like [Vitis vinifera]	142895 bits(417)	1,97E-38	67/135(50%)	95/135(70%)	4/135(3%)
gb ADX97329.1	sulfotransferase-like protein [Mangifera indica]	142895 bits(417)	5,73E-37	67/133(50%)	95/133(71%)	1/133(1%)
ref XP_002265863.1	PREDICTED: flavonol sulfotransferase-like [Vitis vinifera]	140584 bits(417)	3,16E-36	72/145(50%)	100/145(69%)	2/145(1%)
ref XP_002303948.1	predicted protein [Populus trichocarpa]	139428 bits(417)	9,35E-36	67/133(50%)	91/133(68%)	1/133(1%)
gb ABK94985.1	unknown [Populus trichocarpa]	139043 bits(417)	1,06E-35	67/133(50%)	91/133(68%)	1/133(1%)
ref XP_002267209.1	PREDICTED: flavonol sulfotransferase-like [Vitis vinifera]	13365 bits(417)	1,46E-33	67/133(50%)	89/133(67%)	1/133(1%)
gb EMJ19923.1	hypothetical protein PRUPE_ppa009048mg [Prunus persica]	13288 bits(417)	1,62E-33	66/133(50%)	87/133(65%)	1/133(1%)
ref XP_002276423.2	PREDICTED: sulfotransferase 17-like [Vitis vinifera]	132109 bits(417)	5,91E-33	67/135(50%)	91/135(67%)	4/135(3%)
ref XP_002538120.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	129413 bits(417)	3,57E-32	67/133(50%)	96/133(72%)	1/133(1%)
emb CAN78535.1	hypothetical protein VITISV_018319 [Vitis vinifera]	127487 bits(417)	1,64E-31	63/125(50%)	85/125(68%)	3/125(2%)
ref XP_002318367.1	predicted protein [Populus trichocarpa]	12402 bits(417)	4,24E-30	67/135(50%)	90/135(67%)	7/135(5%)
gb ABK95273.1	unknown [Populus trichocarpa]	12402 bits(417)	4,24E-30	67/135(50%)	90/135(67%)	7/135(5%)
ref XP_004306829.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	140969 bits(417)	2,34E-36	67/138(49%)	91/138(66%)	1/138(1%)
ref XP_002524372.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	13288 bits(417)	5,36E-33	67/136(49%)	89/136(65%)	5/136(4%)
ref XP_002276618.2	PREDICTED: sulfotransferase 17-like [Vitis vinifera]	131339 bits(417)	9,85E-33	66/135(49%)	91/135(67%)	4/135(3%)
ref XP_003539305.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	131339 bits(417)	1,13E-32	63/129(49%)	87/129(67%)	1/129(1%)
ref XP_002533884.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	126716 bits(417)	5,38E-31	66/136(49%)	84/136(62%)	5/136(4%)
ref XP_002524376.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	12325 bits(417)	1,71E-29	66/136(49%)	85/136(62%)	5/136(4%)
ref XP_002265783.1	PREDICTED: flavonol 4'-sulfotransferase [Vitis vinifera]	146747 bits(417)	2,24E-38	70/145(48%)	104/145(72%)	2/145(1%)
ref XP_004308399.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	136346 bits(417)	1,40E-34	66/137(48%)	92/137(67%)	1/137(1%)
gb AFK43741.1	unknown [Lotus japonicus]	135576 bits(417)	2,70E-34	67/139(48%)	93/139(67%)	1/139(1%)
ref XP_002873289.1	sulfotransferase family protein [Arabidopsis lyrata subsp. lyrata]	13288 bits(417)	3,93E-33	64/133(48%)	92/133(69%)	1/133(1%)
ref XP_003598424.1	Sulfotransferase [Medicago truncatula]	130568 bits(417)	1,39E-32	64/133(48%)	90/133(68%)	1/133(1%)
ref XP_004307199.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	128257 bits(417)	9,91E-32	63/132(48%)	86/132(65%)	1/132(1%)

ref XP_002514612.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	125946 bits(417)	1,33E-30	64/134(48%)	90/134(67%)	4/134(3%)
ref XP_002524373.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	125176 bits(417)	2,06E-30	65/136(48%)	89/136(65%)	5/136(4%)
ref XP_002524377.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	122865 bits(417)	1,18E-29	65/136(48%)	85/136(62%)	5/136(4%)
ref XP_004291156.1	PREDICTED: cytosolic sulfotransferase 15-like isoform 2 [Fragaria vesca subsp. vesca]	134806 bits(417)	1,82E-34	63/133(47%)	90/133(68%)	1/133(1%)
ref XP_002268667.1	PREDICTED: sulfotransferase 16 [Vitis vinifera]	135191 bits(417)	3,44E-34	64/136(47%)	91/136(67%)	5/136(4%)
ref XP_004291155.1	PREDICTED: cytosolic sulfotransferase 15-like isoform 1 [Fragaria vesca subsp. vesca]	13442 bits(417)	6,12E-34	63/133(47%)	90/133(68%)	1/133(1%)
ref XP_002529671.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	13288 bits(417)	1,64E-33	64/136(47%)	93/136(68%)	5/136(4%)
emb CBI15343.3	unnamed protein product [Vitis vinifera]	127872 bits(417)	1,16E-32	66/140(47%)	92/140(66%)	5/140(4%)
ref NP_190094.1	sulfotransferase family protein [Arabidopsis thaliana]	129028 bits(417)	6,72E-32	62/133(47%)	91/133(68%)	2/133(2%)
ref XP_002264929.1	PREDICTED: sulfotransferase 17-like [Vitis vinifera]	128642 bits(417)	9,04E-32	66/140(47%)	92/140(66%)	5/140(4%)
ref XP_002264512.1	PREDICTED: sulfotransferase 17-like [Vitis vinifera]	128642 bits(417)	1,19E-31	67/144(47%)	93/144(65%)	5/144(3%)
ref XP_004239453.1	PREDICTED: cytosolic sulfotransferase 12-like [Solanum lycopersicum]	127102 bits(417)	2,91E-31	62/132(47%)	89/132(67%)	1/132(1%)
ref XP_003542799.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	126716 bits(417)	5,42E-31	66/139(47%)	90/139(65%)	3/139(2%)
gb EMJ10097.1	hypothetical protein PRUPE_ppa026895mg, partial [Prunus persica]	125176 bits(417)	1,63E-30	64/135(47%)	85/135(63%)	4/135(3%)
ref XP_004240284.1	PREDICTED: cytosolic sulfotransferase 5-like [Solanum lycopersicum]	123635 bits(417)	6,35E-30	62/132(47%)	89/132(67%)	1/132(1%)
ref XP_002519139.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	122479 bits(417)	1,86E-29	62/132(47%)	86/132(65%)	1/132(1%)
sp P52838.1 FSTL_FLABI	RecName: Full=Flavonol sulfotransferase-like	121709 bits(417)	2,22E-29	64/136(47%)	88/136(65%)	2/136(1%)
ref XP_002525538.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	134035 bits(417)	5,84E-34	61/133(46%)	96/133(72%)	1/133(1%)
ref XP_003627950.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	13365 bits(417)	1,88E-33	63/137(46%)	91/137(66%)	1/137(1%)
ref NP_190093.1	sulfotransferase family protein [Arabidopsis thaliana]	128642 bits(417)	7,40E-32	61/132(46%)	88/132(67%)	1/132(1%)
ref XP_003627955.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	129798 bits(417)	8,04E-32	63/137(46%)	90/137(66%)	1/137(1%)
ref XP_002304810.1	predicted protein [Populus trichocarpa]	126716 bits(417)	5,05E-31	74/160(46%)	104/160(65%)	9/160(6%)
gb ACU20971.1	unknown [Glycine max]	123635 bits(417)	9,02E-31	61/132(46%)	84/132(64%)	1/132(1%)
ref XP_002514611.1	Flavonol 3-sulfotransferase, putative [Ricinus communis]	125946 bits(417)	9,66E-31	63/136(46%)	88/136(65%)	5/136(4%)

dbj BAB11158.1	steroid sulfotransferase-like protein [Arabidopsis thaliana]	125946 bits(417)	1,05E-30	61/134(46%)	93/134(69%)	1/134(1%)
ref NP_196317.2	sulfotransferase 2B [Arabidopsis thaliana]	125946 bits(417)	1,13E-30	61/134(46%)	93/134(69%)	1/134(1%)
ref XP_002329757.1	predicted protein [Populus trichocarpa]	124405 bits(417)	2,36E-30	62/134(46%)	84/134(63%)	5/134(4%)
ref XP_004292840.1	PREDICTED: cytosolic sulfotransferase 15-like [Fragaria vesca subsp. vesca]	12402 bits(417)	4,54E-30	62/135(46%)	84/135(62%)	1/135(1%)
ref XP_003542798.1	PREDICTED: flavonol sulfotransferase-like [Glycine max]	12402 bits(417)	4,57E-30	61/132(46%)	84/132(64%)	1/132(1%)
ref XP_003628738.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	121709 bits(417)	3,18E-29	61/132(46%)	86/132(65%)	1/132(1%)
ref XP_003633587.1	PREDICTED: LOW QUALITY PROTEIN: flavonol 4'-sulfotransferase-like [Vitis vinifera]	140584 bits(417)	2,50E-37	69/155(45%)	103/155(66%)	9/155(6%)
emb CBI24935.3	unnamed protein product [Vitis vinifera]	140969 bits(417)	2,81E-36	69/155(45%)	103/155(66%)	9/155(6%)
ref XP_004239020.1	PREDICTED: cytosolic sulfotransferase 12-like [Solanum lycopersicum]	132494 bits(417)	3,71E-33	70/156(45%)	97/156(62%)	2/156(1%)
ref XP_004298346.1	PREDICTED: cytosolic sulfotransferase 5-like [Fragaria vesca subsp. vesca]	129798 bits(417)	1,82E-32	60/132(45%)	89/132(67%)	1/132(1%)
ref NP_568177.1	sulfotransferase 2A [Arabidopsis thaliana]	127872 bits(417)	2,39E-31	60/133(45%)	91/133(68%)	1/133(1%)
dbj BAB11159.1	steroid sulfotransferase-like protein [Arabidopsis thaliana]	127872 bits(417)	2,47E-31	60/133(45%)	91/133(68%)	1/133(1%)
ref XP_003545813.1	PREDICTED: LOW QUALITY PROTEIN: flavonol sulfotransferase-like [Glycine max]	122479 bits(417)	2,15E-29	63/140(45%)	90/140(64%)	3/140(2%)
ref XP_004239455.1	PREDICTED: cytosolic sulfotransferase 5-like [Solanum lycopersicum]	131724 bits(417)	7,05E-33	69/156(44%)	97/156(62%)	2/156(1%)
gb AAL06879.1	AT5g07010/MOJ9_18 [Arabidopsis thaliana]	125946 bits(417)	1,08E-30	59/133(44%)	90/133(68%)	1/133(1%)
ref XP_002329759.1	predicted protein [Populus trichocarpa]	125561 bits(417)	1,40E-30	61/138(44%)	87/138(63%)	5/138(4%)
ref XP_002529982.1	Flavonol 4'-sulfotransferase, putative [Ricinus communis]	123635 bits(417)	6,47E-30	58/133(44%)	93/133(70%)	1/133(1%)
ref XP_003576818.1	PREDICTED: flavonol 3-sulfotransferase-like [Brachypodium distachyon]	122865 bits(417)	1,54E-29	52/119(44%)	78/119(66%)	0/119(0%)
emb CAN75260.1	hypothetical protein VITISV_041143 [Vitis vinifera]	13442 bits(417)	6,42E-35	68/157(43%)	97/157(62%)	13/157(8%)
ref XP_003545845.1	PREDICTED: LOW QUALITY PROTEIN: flavonol 4'-sulfotransferase-like [Glycine max]	130183 bits(417)	1,10E-32	67/155(43%)	93/155(60%)	5/155(3%)
ref XP_003627983.1	Flavonol sulfotransferase-like protein [Medicago truncatula]	126331 bits(417)	8,07E-31	59/137(43%)	92/137(67%)	1/137(1%)
ref XP_002317222.1	predicted protein [Populus trichocarpa]	122479 bits(417)	1,88E-29	60/138(43%)	86/138(62%)	5/138(4%)
ref XP_002317223.1	predicted protein [Populus trichocarpa]	122094 bits(417)	2,23E-29	60/138(43%)	86/138(62%)	5/138(4%)
ref XP_002329758.1	predicted protein [Populus trichocarpa]	120939 bits(417)	2,88E-29	60/138(43%)	86/138(62%)	5/138(4%)

ref XP_004247253.1	PREDICTED: cytosolic sulfotransferase 5-like [Solanum lycopersicum]	128642 bits(417)	8,38E-32	65/156(42%)	97/156(62%)	2/156(1%)
ref XP_003580959.1	PREDICTED: sulfotransferase 17-like [Brachypodium distachyon]	121709 bits(417)	2,90E-29	55/132(42%)	86/132(65%)	2/132(2%)
ref XP_002262899.2	PREDICTED: flavonol 4'-sulfotransferase-like [Vitis vinifera]	12325 bits(417)	3,18E-29	64/151(42%)	97/151(64%)	9/151(6%)
ref XP_003627666.1	Steroid sulfotransferase-like protein [Medicago truncatula]	131339 bits(417)	2,15E-32	65/157(41%)	103/157(66%)	2/157(1%)
Contig56						
Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb EMJ16906.1	hypothetical protein PRUPE_ppa008901mg [Prunus persica]	989821 bits(253)	4,12E-22	56/76(74%)	63/76(83%)	2/76(3%)
ref XP_002524600.1	conserved hypothetical protein [Ricinus communis]	102064 bits(253)	3,29E-23	56/78(72%)	62/78(79%)	9/78(12%)
ref XP_004144556.1	PREDICTED: uncharacterized protein LOC101216086 [Cucumis sativus]	908929 bits(253)	2,66E-19	50/70(71%)	56/70(80%)	2/70(3%)
ref XP_002320535.1	predicted protein [Populus trichocarpa]	982117 bits(253)	7,00E-22	54/78(69%)	60/78(77%)	4/78(5%)
ref XP_004303585.1	PREDICTED: uncharacterized protein LOC101313819 [Fragaria vesca subsp. vesca]	935893 bits(253)	3,74E-20	52/78(67%)	59/78(76%)	3/78(4%)
emb CBI28475.3	unnamed protein product [Vitis vinifera]	889669 bits(253)	9,71E-19	58/88(66%)	66/88(75%)	5/88(6%)
ref XP_002264683.1	PREDICTED: uncharacterized protein LOC100249313 isoform 1 [Vitis vinifera]	893521 bits(253)	1,13E-18	58/88(66%)	66/88(75%)	5/88(6%)
ref XP_004238678.1	PREDICTED: uncharacterized protein LOC101255093 [Solanum lycopersicum]	820333 bits(253)	5,46E-16	49/75(65%)	53/75(71%)	1/75(1%)
gb AFK35768.1	unknown [Lotus japonicus]	885817 bits(253)	2,22E-18	50/78(64%)	59/78(76%)	0/78(0%)
ref XP_003521708.1	PREDICTED: RRP15-like protein-like [Glycine max]	851149 bits(253)	3,45E-17	49/79(62%)	56/79(71%)	2/79(3%)
emb CAN82067.1	hypothetical protein VITISV_037473 [Vitis vinifera]	916633 bits(253)	9,70E-20	54/88(61%)	63/88(72%)	5/88(6%)
ref NP_001242109.1	uncharacterized protein LOC100780091 [Glycine max]	839593 bits(253)	8,15E-17	48/79(61%)	56/79(71%)	2/79(3%)
ref XP_002863882.1	hypothetical protein ARALYDRAFT_494878 [Arabidopsis lyrata subsp. lyrata]	766406 bits(253)	2,63E-14	44/78(56%)	49/78(63%)	7/78(9%)
ref NP_199635.2	uncharacterized protein [Arabidopsis thaliana]	743294 bits(253)	2,18E-13	41/73(56%)	47/73(64%)	7/73(10%)

gb ACG33693.1	hypothetical protein [Zea mays]	63929 bits(253)	3,33E-10	38/74(51%)	45/74(61%)	4/74(5%)
Contig57						
Sequence Id	Name	Score	Expect	Identities	Positities	Gaps
gb ACD69680.1	metallothionein [Mangifera indica]	912781 bits(225)	2,59E-20	43/54(80%)	47/54(87%)	0/54(0%)
gb ADH04476.1	metallothionein [Mangifera indica]	912781 bits(225)	2,73E-20	44/56(79%)	48/56(86%)	0/56(0%)
gb ABD97258.1	metallothionin 2 [Camellia sinensis]	69707 bits(225)	1,64E-12	37/52(71%)	40/52(77%)	1/52(2%)
ref XP_002265058.1	PREDICTED: metallothionein isoform 1 [Vitis vinifera]	63929 bits(225)	1,98E-10	36/53(68%)	40/53(75%)	2/53(4%)
dbj BAA96449.1	metallothionein-like protein [Pyrus pyrifolia]	704774 bits(225)	7,41E-13	36/54(67%)	42/54(78%)	1/54(2%)
gb ABN46988.1	metallothionein-like protein 2b [Nelumbo nucifera]	700922 bits(225)	1,04E-12	36/55(65%)	43/55(78%)	3/55(5%)
gb EMJ06266.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	69707 bits(225)	1,21E-12	35/54(65%)	43/54(80%)	1/54(2%)
gb EMJ06265.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	685514 bits(225)	3,42E-12	35/54(65%)	43/54(80%)	1/54(2%)
gb AFP93964.1	metallothionein type 2 [Ilex paraguariensis]	673958 bits(225)	9,03E-12	34/52(65%)	40/52(77%)	1/52(2%)
ref XP_002285180.1	PREDICTED: metallothionein-like protein type 2 [Vitis vinifera]	627734 bits(225)	5,60E-10	34/53(64%)	39/53(74%)	2/53(4%)
gb ABD97257.1	metallothionin 1 [Camellia sinensis]	673958 bits(225)	9,77E-12	33/52(63%)	39/52(75%)	1/52(2%)
gb ACS14665.1	metallothionein [Camellia oleifera]	666254 bits(225)	2,02E-11	33/52(63%)	39/52(75%)	1/52(2%)
gb ABN46987.1	metallothionein-like protein 2a [Nelumbo nucifera]	643142 bits(225)	1,26E-10	34/54(63%)	41/54(76%)	3/54(6%)
ref XP_002314182.1	predicted protein [Populus trichocarpa]	685514 bits(225)	2,71E-12	32/52(62%)	40/52(77%)	1/52(2%)
gb ABK94874.1	unknown [Populus trichocarpa]	681662 bits(225)	5,29E-12	32/52(62%)	40/52(77%)	1/52(2%)
gb AAT02525.1	metallothionein 2b [Populus trichocarpa X Populus deltoides]	670106 bits(225)	1,38E-11	32/52(62%)	39/52(75%)	1/52(2%)
ref XP_002299873.1	predicted protein [Populus trichocarpa]	670106 bits(225)	1,54E-11	32/52(62%)	40/52(77%)	1/52(2%)
gb ABM21762.1	metallothionein-like protein MT2B [Salix matsudana]	662402 bits(225)	2,25E-11	32/52(62%)	39/52(75%)	1/52(2%)
ref XP_002329424.1	predicted protein [Populus trichocarpa]	65855 bits(225)	2,29E-11	32/52(62%)	38/52(73%)	1/52(2%)

emb CAA10232.1	metallothionein-like protein class II [Fagus sylvatica]	650846 bits(225)	5,87E-11	33/53(62%)	39/53(74%)	2/53(4%)
sp P43390.1 MT2_ACTDE	RecName: Full=Metallothionein-like protein type 2	646994 bits(225)	9,12E-11	32/52(62%)	38/52(73%)	1/52(2%)
gb AAY16439.1	metallothionein-like protein [Betula platyphylla]	646994 bits(225)	9,40E-11	33/53(62%)	41/53(77%)	1/53(2%)
gb AAT02522.1	metallothionein 1a [Populus trichocarpa x Populus deltoides]	643142 bits(225)	1,04E-10	32/52(62%)	38/52(73%)	1/52(2%)
ref XP_002330684.1	predicted protein [Populus trichocarpa]	635438 bits(225)	2,15E-10	31/52(60%)	38/52(73%)	1/52(2%)
gb AAT02523.1	metallothionein 1b [Populus trichocarpa x Populus deltoides]	635438 bits(225)	2,24E-10	31/52(60%)	38/52(73%)	1/52(2%)
gb AFK41306.1	unknown [Lotus japonicus]	635438 bits(225)	2,28E-10	32/54(59%)	42/54(78%)	2/54(4%)
emb CAE12162.1	metallothionein-like protein [Quercus robur]	631586 bits(225)	4,93E-10	30/51(59%)	39/51(76%)	2/51(4%)
gb ABS72197.1	metallothionein 2b, partial [Corchorus olitorius]	627734 bits(225)	2,97E-10	30/52(58%)	38/52(73%)	1/52(2%)
gb ADM86706.1	metallothionein-like protein [Cicer microphyllum]	62003 bits(225)	8,25E-10	30/53(57%)	39/53(74%)	2/53(4%)
dbj BAJ39944.1	metallothionein [Ziziphus jujuba]	650846 bits(225)	7,16E-11	30/54(56%)	40/54(74%)	2/54(4%)

Contig59

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
emb CBI25971.3	unnamed protein product [Vitis vinifera]	905077 bits(335)	1,46E-26	43/51(84%)	46/51(90%)	0/51(0%)
gb ACJ11739.1	dihydrolipoamide dehydrogenase [Gossypium hirsutum]	13365 bits(335)	6,05E-33	74/117(63%)	79/117(68%)	2/117(2%)
gb AAF79529.1 AC023673_17	F21D18.28 [Arabidopsis thaliana]	128642 bits(335)	3,79E-31	71/121(59%)	80/121(66%)	8/121(7%)
ref NP_175237.1	dihydrolipoamide dehydrogenase 1 [Arabidopsis thaliana]	128642 bits(335)	3,86E-31	71/121(59%)	80/121(66%)	8/121(7%)
dbj BAD95420.1	lipoamide dehydrogenase precursor [Arabidopsis thaliana]	120553 bits(335)	5,15E-31	69/116(59%)	77/116(66%)	8/116(7%)
ref XP_002894091.1	F21D18.28 [Arabidopsis lyrata subsp. lyrata]	128257 bits(335)	5,22E-31	71/121(59%)	80/121(66%)	8/121(7%)
dbj BAJ34204.1	unnamed protein product [Thellungiella halophila]	128257 bits(335)	5,32E-31	71/120(59%)	80/120(67%)	8/120(7%)
ref XP_002519286.1	dihydrolipoamide dehydrogenase, putative [Ricinus communis]	126716 bits(335)	1,62E-30	71/120(59%)	80/120(67%)	8/120(7%)
gb AAG17888.1	dihydrolipoamide dehydrogenase precursor [Solanum tuberosum]	125946 bits(335)	3,19E-30	69/117(59%)	77/117(66%)	2/117(2%)

gb ACF17643.1	putative branched-chain alpha-keto acid dehydrogenase E3 subunit [<i>Capsicum annuum</i>]	125946 bits(335)	3,22E-30	69/117(59%)	77/117(66%)	2/117(2%)
ref XP_004252936.1	PREDICTED: dihydrolipoyl dehydrogenase, mitochondrial-like isoform 1 [<i>Solanum lycopersicum</i>]	125946 bits(335)	3,25E-30	69/117(59%)	77/117(66%)	2/117(2%)
dbj BAB44156.1	dihydrolipoamide dehydrogenase precursor [<i>Bruguiera gymnorhiza</i>]	127487 bits(335)	9,89E-31	70/120(58%)	81/120(68%)	8/120(7%)
ref XP_004252937.1	PREDICTED: dihydrolipoyl dehydrogenase, mitochondrial-like isoform 2 [<i>Solanum lycopersicum</i>]	125946 bits(335)	2,10E-30	69/118(58%)	77/118(65%)	2/118(2%)
gb AAS47493.1	lipoamide dehydrogenase [<i>Capsicum annuum</i>]	123635 bits(335)	1,94E-29	69/120(58%)	80/120(67%)	8/120(7%)
gb ABO61735.1	mitochondrial lipoamide dehydrogenase [<i>Populus tremuloides</i>]	123635 bits(335)	2,29E-29	70/121(58%)	78/121(64%)	10/121(8%)
ref XP_004157030.1	PREDICTED: dihydrolipoyl dehydrogenase 1, mitochondrial-like [<i>Cucumis sativus</i>]	119783 bits(335)	4,78E-28	68/117(58%)	75/117(64%)	2/117(2%)
ref XP_004138526.1	PREDICTED: dihydrolipoyl dehydrogenase 1, mitochondrial-like [<i>Cucumis sativus</i>]	119783 bits(335)	5,23E-28	68/117(58%)	75/117(64%)	2/117(2%)
ref XP_002263180.1	PREDICTED: dihydrolipoyl dehydrogenase, mitochondrial [<i>Vitis vinifera</i>]	114005 bits(335)	5,18E-26	68/117(58%)	75/117(64%)	2/117(2%)
ref XP_004147620.1	PREDICTED: dihydrolipoyl dehydrogenase 1, mitochondrial-like [<i>Cucumis sativus</i>]	125176 bits(335)	5,04E-30	68/120(57%)	79/120(66%)	8/120(7%)
gb EMJ24460.1	hypothetical protein PRUPE_ppa004441mg [<i>Prunus persica</i>]	125176 bits(335)	5,53E-30	68/120(57%)	80/120(67%)	8/120(7%)
ref NP_001237764.1	ferric leghemoglobin reductase-2 precursor [<i>Glycine max</i>]	121709 bits(335)	8,16E-29	68/120(57%)	78/120(65%)	8/120(7%)
gb ABK96408.1	unknown [<i>Populus trichocarpa</i> x <i>Populus deltoides</i>]	120168 bits(335)	3,62E-28	69/121(57%)	77/121(64%)	10/121(8%)
ref XP_002311367.1	precursor of dehydrogenase dihydrolipoamide dehydrogenase 1 [<i>Populus trichocarpa</i>]	119783 bits(335)	3,92E-28	69/121(57%)	77/121(64%)	10/121(8%)
ref XP_004295034.1	PREDICTED: dihydrolipoyl dehydrogenase 1, mitochondrial-like [<i>Fragaria vesca</i> subsp. <i>vesca</i>]	119398 bits(335)	7,10E-28	70/122(57%)	77/122(63%)	12/122(10%)
ref XP_003550821.1	PREDICTED: dihydrolipoyl dehydrogenase, mitochondrial-like [<i>Glycine max</i>]	120939 bits(335)	1,57E-28	67/120(56%)	78/120(65%)	8/120(7%)
emb CAA11554.1	2-oxoglutarate dehydrogenase, E3 subunit [<i>Arabidopsis thaliana</i>]	119783 bits(335)	3,24E-28	66/118(56%)	76/118(64%)	2/118(2%)
ref NP_851005.1	dihydrolipoyl dehydrogenase 2 [<i>Arabidopsis thaliana</i>]	119783 bits(335)	4,11E-28	66/117(56%)	76/117(65%)	2/117(2%)
ref XP_002885195.1	hypothetical protein ARALYDRAFT_898058 [<i>Arabidopsis lyrata</i> subsp. <i>lyrata</i>]	119783 bits(335)	4,96E-28	66/117(56%)	76/117(65%)	2/117(2%)
gb AAD53185.1 AF181096_1	ferric leghemoglobin reductase [<i>Vigna unguiculata</i>]	119783 bits(335)	5,60E-28	66/117(56%)	77/117(66%)	8/117(7%)
pdb 1DXL A	Chain A, Dihydrolipoamide Dehydrogenase Of Glycine Decarboxylase From <i>Pisum Sativum</i>	120168 bits(335)	2,86E-28	66/120(55%)	78/120(65%)	8/120(7%)
sp P31023.2 DLDH_PEA	RecName: Full=Dihydrolipoyl dehydrogenase, mitochondrial; AltName: Full=Dihydrolipoamide dehydrogenase; AltName: Full=Glycine cleavage system L protein; AltName: Full=Pyruvate	120168 bits(335)	2,90E-28	66/120(55%)	78/120(65%)	8/120(7%)

	dehydrogenase complex E3 subunit; Short=E3; Short=PDC-E3; Flags: Precursor					
gb AFK33342.1	unknown [Medicago truncatula]	119398 bits(335)	5,32E-28	66/120(55%)	78/120(65%)	8/120(7%)
ref NP_001234770.1	dihydrolipoamide dehydrogenase precursor [Solanum lycopersicum]	119013 bits(335)	7,85E-28	70/128(55%)	78/128(61%)	24/128(19%)
ref NP_001238628.1	ferric leghemoglobin reductase [Glycine max]	113235 bits(335)	1,13E-25	64/117(55%)	75/117(64%)	8/117(7%)
ref XP_001767995.1	predicted protein [Physcomitrella patens subsp. patens]	111309 bits(335)	4,15E-25	62/117(53%)	72/117(62%)	2/117(2%)
gb ABR16524.1	unknown [Picea sitchensis]	114005 bits(335)	4,54E-26	62/121(51%)	78/121(64%)	10/121(8%)
gb EAY73818.1	hypothetical protein Osl_01694 [Oryza sativa Indica Group]	108997 bits(335)	2,53E-24	62/122(51%)	74/122(61%)	10/122(8%)
ref NP_001183318.1	uncharacterized protein LOC100501719 [Zea mays]	108227 bits(335)	2,87E-24	60/118(51%)	71/118(60%)	2/118(2%)
ref NP_001042918.1	Os01g0328700 [Oryza sativa Japonica Group]	108997 bits(335)	2,95E-24	62/122(51%)	74/122(61%)	10/122(8%)
gb AFW80669.1	hypothetical protein ZEAMMB73_669487 [Zea mays]	107842 bits(335)	7,53E-24	60/118(51%)	71/118(60%)	2/118(2%)
gb EKM52919.1	hypothetical protein PHACADRAFT_261612 [Phanerochaete carnosae HHB-10118-sp]	107071 bits(335)	1,41E-23	59/116(51%)	73/116(63%)	2/116(2%)
gb EIN03538.1	dihydrolipoamide dehydrogenase [Punctularia strigosozonata HHB-11173 SS5]	10476 bits(335)	8,85E-23	60/117(51%)	74/117(63%)	4/117(3%)
gb EIW56650.1	dihydrolipoamide dehydrogenase [Trametes versicolor FP-101664 SS1]	107457 bits(335)	1,26E-23	58/117(50%)	71/117(61%)	2/117(2%)
gb EJF60116.1	dihydrolipoamide dehydrogenase [Dichomitus squalens LYAD-421 SS1]	105916 bits(335)	4,42E-23	59/118(50%)	74/118(63%)	4/118(3%)
tpg DAA54462.1	TPA: hypothetical protein ZEAMMB73_412429 [Zea mays]	104375 bits(335)	5,32E-23	59/118(50%)	70/118(59%)	2/118(2%)
ref XP_002457769.1	hypothetical protein SORBIDRAFT_03g013290 [Sorghum bicolor]	10399 bits(335)	1,99E-22	59/118(50%)	70/118(59%)	2/118(2%)
ref XP_003568861.1	PREDICTED: dihydrolipoamide dehydrogenase 1, mitochondrial-like [Brachypodium distachyon]	101679 bits(335)	1,48E-21	60/121(50%)	74/121(61%)	10/121(8%)
dbj BAK03695.1	predicted protein [Hordeum vulgare subsp. vulgare]	100908 bits(335)	2,17E-21	60/121(50%)	74/121(61%)	10/121(8%)
gb EGN94475.1	hypothetical protein SERLA73DRAFT_188396 [Serpula lacrymans var. lacrymans S7.3]	105916 bits(335)	3,96E-23	59/120(49%)	78/120(65%)	8/120(7%)
dbj GAC74752.1	dihydrolipoamide dehydrogenase [Pseudozyma antarctica T-34]	105145 bits(335)	7,54E-23	57/117(49%)	72/117(62%)	2/117(2%)
ref XP_001875195.1	dihydrolipoamide dehydrogenase [Laccaria bicolor S238N-H82]	103605 bits(335)	2,81E-22	59/121(49%)	76/121(63%)	8/121(7%)
ref XP_001214883.1	dihydrolipoamide dehydrogenase, mitochondrial precursor [Aspergillus terreus NIH2624]	993673 bits(335)	1,07E-20	58/118(49%)	69/118(58%)	4/118(3%)
gb EEC78570.1	hypothetical protein Osl_18555 [Oryza sativa Indica Group]	978265 bits(335)	3,03E-20	59/121(49%)	73/121(60%)	10/121(8%)

dbj BAK02719.1	predicted protein [Hordeum vulgare subsp. vulgare]	105916 bits(335)	3,62E-23	60/125(48%)	77/125(62%)	18/125(14%)
dbj BAK01655.1	predicted protein [Hordeum vulgare subsp. vulgare]	105916 bits(335)	3,70E-23	60/125(48%)	77/125(62%)	18/125(14%)
gb EJT97247.1	dihydrolipoyl dehydrogenase [Dacryopinax sp. DJM-731 SS1]	105145 bits(335)	6,25E-23	56/117(48%)	73/117(62%)	2/117(2%)
ref XP_503086.1	YALI0D20768p [Yarrowia lipolytica]	105145 bits(335)	6,91E-23	56/117(48%)	70/117(60%)	2/117(2%)
emb CCO29553.1	dihydrolipoamide dehydrogenase [Rhizoctonia solani AG-1 IB]	103219 bits(335)	1,52E-22	58/120(48%)	73/120(61%)	8/120(7%)
emb CBQ71205.1	probable LPD1-dihydrolipoamide dehydrogenase precursor [Sporisorium reilianum SRZ2]	104375 bits(335)	1,63E-22	56/117(48%)	72/117(62%)	2/117(2%)
ref XP_758608.1	hypothetical protein UM02461.1 [Ustilago maydis 521]	103605 bits(335)	2,86E-22	56/117(48%)	72/117(62%)	2/117(2%)
emb CCF48013.1	probable LPD1-dihydrolipoamide dehydrogenase precursor [Ustilago hordei]	102449 bits(335)	7,97E-22	56/117(48%)	71/117(61%)	2/117(2%)
gb EFX00956.1	dihydrolipoamide dehydrogenase [Grosmania clavigera kw1407]	101293 bits(335)	1,74E-21	57/118(48%)	72/118(61%)	4/118(3%)
ref XP_001014580.2	dihydrolipoamide dehydrogenase family protein [Tetrahymena thermophila]	100523 bits(335)	3,31E-21	59/123(48%)	72/123(59%)	14/123(11%)
gb AFP20527.1	dihydrolipoamide dehydrogenase E3 subunit [Rhizopertha dominica]	100523 bits(335)	3,99E-21	58/122(48%)	74/122(61%)	11/122(9%)
gb EGF82242.1	hypothetical protein BATDEDRAFT_87007 [Batrachochytrium dendrobatidis JAM81]	100138 bits(335)	4,23E-21	59/123(48%)	72/123(59%)	14/123(11%)
emb CCA74849.1	probable LPD1-dihydrolipoamide dehydrogenase precursor [Piriformospora indica DSM 11827]	107071 bits(335)	1,60E-23	57/121(47%)	74/121(61%)	10/121(8%)
ref XP_003285611.1	dihydrolipoamide:NAD oxidoreductase [Dictyostelium purpureum]	106301 bits(335)	2,43E-23	56/120(47%)	79/120(66%)	8/120(7%)
gb EIW75583.1	dihydrolipoyl dehydrogenase [Coniophora puteana RWD-64-598 SS2]	102449 bits(335)	6,87E-22	56/120(47%)	76/120(63%)	8/120(7%)
ref XP_001246110.1	hypothetical protein CIMG_05551 [Coccidioides immitis RS]	955153 bits(335)	1,72E-21	55/118(47%)	68/118(58%)	4/118(3%)
ref XP_001835168.1	dihydrolipoamide dehydrogenase [Coprinospis cinerea okayama7#130]	100908 bits(335)	2,64E-21	57/121(47%)	75/121(62%)	8/121(7%)
emb CCM06592.1	predicted protein [Fibroporia radiculosa]	100523 bits(335)	3,22E-21	55/117(47%)	73/117(62%)	4/117(3%)
dbj GAA85350.1	dihydrolipoamide dehydrogenase [Aspergillus kawachii IFO 4308]	989821 bits(335)	1,19E-20	56/118(47%)	70/118(59%)	4/118(3%)
ref XP_001391760.1	dihydrolipoyl dehydrogenase [Aspergillus niger CBS 513.88]	989821 bits(335)	1,22E-20	56/118(47%)	70/118(59%)	4/118(3%)
ref XP_001273001.1	dihydrolipoamide dehydrogenase [Aspergillus clavatus NRRL 1]	989821 bits(335)	1,22E-20	55/118(47%)	71/118(60%)	4/118(3%)
gb EIE20004.1	dihydrolipoamide dehydrogenase of glycine decarboxylase from pisum Sativum [Coccomyxa subellipsoidea C-169]	989821 bits(335)	1,29E-20	56/119(47%)	72/119(61%)	8/119(7%)

ref XP_002843244.1	dihydrolipoamide dehydrogenase [Arthroderma otae CBS 113480]	974413 bits(335)	1,85E-20	56/118(47%)	70/118(59%)	4/118(3%)
gb EGD94449.1	Dihydrolipoyl dehydrogenase [Trichophyton tonsurans CBS 112818]	978265 bits(335)	2,19E-20	56/118(47%)	70/118(59%)	4/118(3%)
gb AFP20529.1	dihydrolipoamide dehydrogenase E3 subunit [Rhizopertha dominica]	982117 bits(335)	2,40E-20	57/122(47%)	74/122(61%)	11/122(9%)
ref XP_003237891.1	dihydrolipoamide dehydrogenase [Trichophyton rubrum CBS 118892]	978265 bits(335)	3,42E-20	56/118(47%)	70/118(59%)	4/118(3%)
ref XP_004352811.1	dihydrolipoyl dehydrogenase [Acanthamoeba castellanii str. Neff]	101293 bits(335)	1,67E-21	56/123(46%)	73/123(59%)	12/123(10%)
ref XP_001457469.1	hypothetical protein [Paramecium tetraurelia strain d4-2]	989821 bits(335)	9,30E-21	56/121(46%)	73/121(60%)	10/121(8%)
ref XP_001460760.1	hypothetical protein [Paramecium tetraurelia strain d4-2]	989821 bits(335)	1,17E-20	56/121(46%)	73/121(60%)	10/121(8%)
ref XP_002982240.1	hypothetical protein SELMODRAFT_445122 [Selaginella moellendorffii]	989821 bits(335)	1,17E-20	58/128(45%)	71/128(55%)	24/128(19%)
ref XP_002993435.1	hypothetical protein SELMODRAFT_270051 [Selaginella moellendorffii]	989821 bits(335)	1,22E-20	58/128(45%)	71/128(55%)	24/128(19%)
gb EGA62521.1	Lpd1p [Saccharomyces cerevisiae FostersO]	100523 bits(335)	8,80E-22	52/119(44%)	75/119(63%)	8/119(7%)
pdb 1JEH A	Chain A, Crystal Structure Of Yeast E3, Lipoamide Dehydrogenase	100523 bits(335)	2,86E-21	52/119(44%)	75/119(63%)	8/119(7%)

Contig61

Sequence Id	Name	Score	Expect	Identities	Positives	Gaps
gb ACD69680.1	metallothionein [Mangifera indica]	912781 bits(225)	3,85E-20	43/54(80%)	47/54(87%)	0/54(0%)
gb ADH04476.1	metallothionein [Mangifera indica]	912781 bits(225)	4,05E-20	44/56(79%)	48/56(86%)	0/56(0%)
gb ABD97258.1	metallothionin 2 [Camellia sinensis]	69707 bits(225)	2,23E-12	37/52(71%)	40/52(77%)	1/52(2%)
ref XP_002265058.1	PREDICTED: metallothionein isoform 1 [Vitis vinifera]	63929 bits(225)	3,10E-10	36/53(68%)	40/53(75%)	2/53(4%)
dbj BAA96449.1	metallothionein-like protein [Pyrus pyrifolia]	704774 bits(225)	9,92E-13	36/54(67%)	42/54(78%)	1/54(2%)
gb ABN46988.1	metallothionein-like protein 2b [Nelumbo nucifera]	700922 bits(225)	1,40E-12	36/55(65%)	43/55(78%)	3/55(5%)
gb EMJ06266.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	69707 bits(225)	1,64E-12	35/54(65%)	43/54(80%)	1/54(2%)

gb EMJ06265.1	hypothetical protein PRUPE_ppa014268mg [Prunus persica]	685514 bits(225)	4,77E-12	35/54(65%)	43/54(80%)	1/54(2%)
gb AFP93964.1	metallothionein type 2 [Ilex paraguariensis]	673958 bits(225)	1,30E-11	34/52(65%)	40/52(77%)	1/52(2%)
ref XP_002285180.1	PREDICTED: metallothionein-like protein type 2 [Vitis vinifera]	627734 bits(225)	8,97E-10	34/53(64%)	39/53(74%)	2/53(4%)
gb ABD97257.1	metallothionin 1 [Camellia sinensis]	673958 bits(225)	1,41E-11	33/52(63%)	39/52(75%)	1/52(2%)
gb ACS14665.1	metallothionein [Camellia oleifera]	666254 bits(225)	2,97E-11	33/52(63%)	39/52(75%)	1/52(2%)
gb ABN46987.1	metallothionein-like protein 2a [Nelumbo nucifera]	643142 bits(225)	1,95E-10	34/54(63%)	41/54(76%)	3/54(6%)
ref XP_002314182.1	predicted protein [Populus trichocarpa]	685514 bits(225)	3,80E-12	32/52(62%)	40/52(77%)	1/52(2%)
gb ABK94874.1	unknown [Populus trichocarpa]	681662 bits(225)	7,47E-12	32/52(62%)	40/52(77%)	1/52(2%)
gb AAT02525.1	metallothionein 2b [Populus trichocarpa X Populus deltoides]	670106 bits(225)	2,01E-11	32/52(62%)	39/52(75%)	1/52(2%)
ref XP_002299873.1	predicted protein [Populus trichocarpa]	670106 bits(225)	2,25E-11	32/52(62%)	40/52(77%)	1/52(2%)
gb ABM21762.1	metallothionein-like protein MT2B [Salix matsudana]	662402 bits(225)	3,32E-11	32/52(62%)	39/52(75%)	1/52(2%)
ref XP_002329424.1	predicted protein [Populus trichocarpa]	65855 bits(225)	3,43E-11	32/52(62%)	38/52(73%)	1/52(2%)
emb CAA10232.1	metallothionein-like protein class II [Fagus sylvatica]	650846 bits(225)	8,90E-11	33/53(62%)	39/53(74%)	2/53(4%)
sp P43390.1 MT2_ACTDE	RecName: Full=Metallothionein-like protein type 2	646994 bits(225)	1,40E-10	32/52(62%)	38/52(73%)	1/52(2%)
gb AA16439.1	metallothionein-like protein [Betula platyphylla]	646994 bits(225)	1,45E-10	33/53(62%)	41/53(77%)	1/53(2%)
gb AAT02522.1	metallothionein 1a [Populus trichocarpa x Populus deltoides]	643142 bits(225)	1,61E-10	32/52(62%)	38/52(73%)	1/52(2%)
ref XP_002330684.1	predicted protein [Populus trichocarpa]	635438 bits(225)	3,38E-10	31/52(60%)	38/52(73%)	1/52(2%)
gb AAT02523.1	metallothionein 1b [Populus trichocarpa x Populus deltoides]	635438 bits(225)	3,53E-10	31/52(60%)	38/52(73%)	1/52(2%)
gb AFK41306.1	unknown [Lotus japonicus]	635438 bits(225)	3,60E-10	32/54(59%)	42/54(78%)	2/54(4%)
emb CAE12162.1	metallothionein-like protein [Quercus robur]	631586 bits(225)	7,82E-10	30/51(59%)	39/51(76%)	2/51(4%)
gb ABS72197.1	metallothionein 2b, partial [Corchorus olitorius]	627734 bits(225)	4,74E-10	30/52(58%)	38/52(73%)	1/52(2%)
dbj BAJ39944.1	metallothionein [Ziziphus jujuba]	650846 bits(225)	1,10E-10	30/54(56%)	40/54(74%)	2/54(4%)