

UNIVERSIDADE FEDERAL DE PELOTAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE SEMENTES



Tese

**Análise *in silico* de proteínas relacionadas a sementes e
identificação de microssatélites através da bioinformática**

GERI EDUARDO MENEGHELLO

Pelotas, 2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

GERI EDUARDO MENEGHELLO

Análise *in silico* de proteínas relacionadas a sementes e identificação de microssatélites através da bioinformática

Tese apresentada à Faculdade de Agronomia “Eliseu Maciel”, Universidade Federal de Pelotas, sob a orientação do Prof. Dr. Paulo Dejalma Zimmer, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia de Sementes, para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Paulo Dejalma Zimmer, FAEM/UFPeI

Co-Orientadores: Prof. Antonio Carlos Souza Albuquerque Barros, FAEM/UFPeI

Pesq. Dr. Eliseu Binneck, Dr. EMBRAPA SOJA

Pelotas, 2007

**Dados de catalogação na fonte:
(Marlene Cravo Castillo – CRB-10/744)**

M541c Meneghello, Geri Eduardo

Análise in silico de proteínas relacionadas a sementes e identificação de microssatélites através da bioinformática / Geri Eduardo Meneghello . - Pelotas, 2007.
58f. : tab.

Tese (Doutorado) –Programa de Pós-Graduação em Ciência e Tecnologia de Sementes. Faculdade de Agronomia Eliseu Maciel. Universidade Federal de Pelotas. - Pelotas, 2007, Paulo Dejalma Zimmer, Orientador.

1. Sementes 2. Marcador molecular 3. Proteínas 4. Microssatélites 5. Bioinformática I Zimmer, Paulo Dejalma (orientador) II .Título.

CDD 631.521

Banca Examinadora

Prof. Antonio Carlos Souza Albuquerque Barros, Dr.

Prof. Luis Osmar Braga Schuch, Dr

Prof. César Valmor Rombaldi, Dr.

Pesq. Élbio Treicha Cardoso, Dr.

Se não der frutos, valeu a beleza das flores.
Se não der flores, valeu a sombra das folhas.
Se não der folhas, valeu a intenção da semente.

Henfil

Ao Gabriel,

dedico

AGRADECIMENTOS

Uma tese é o produto coletivo, pelo qual colaboradores contribuíram para o resultado deste trabalho. A todas essas pessoas quero registrar minha gratidão.

Agradeço ao orientador deste trabalho Prof. Dr. Paulo Dejalma Zimmer pela orientação, estímulo, respeito e amizade.

Aos co-orientadores Prof. Dr. Antonio Carlos Souza Albuquerque Barro e Dr. Eliseu Binneck pela orientação, disposição, paciência e amizade.

Aos professores e funcionários da Área de Sementes da Universidade Federal de Pelotas pelas contribuições e ajuda prestada.

Aos colegas de Pós-Graduação pela bela convivência e sincera amizade, em especial ao Luciano C. Maia (Paulista) pelo auxílio no desenvolvimento dos *scripts*.

À Faculdade de Agronomia Eliseu Maciel e a Universidade Federal de Pelotas pela oportunidade de realizar a Pós-Graduação.

A todos aqueles que, direta ou indiretamente, contribuíram de alguma forma para realização deste trabalho.

À minha família, pelo apoio, estímulo e carinho.

A todas as pessoas que se furtaram em algum momento do meu convívio durante essa caminhada

A você pela atenção ...

SUMÁRIO

AGRADECIMENTOS.....	iv
LISTA DE TABELAS.....	vi
LISTA DE FIGURAS.....	vii
INTRODUÇÃO GERAL.....	01
CAPITULO I - Análise <i>in silico</i> de proteínas presentes em sementes depositadas no banco de dados Swiss-Prot	03
Resumo	04
Abstract	05
Introdução.....	06
Material e Métodos	10
Resultados e Discussão	11
Conclusões.....	22
Referências Bibliográficas	23
CAPITULO II - Localização <i>in silico</i> de microssatélites em espécies Fabaceae de interesse agrônomo.....	27
Resumo	28
Abstract	29
Introdução.....	30
Material e Métodos	35
Resultados e Discussão	38
Conclusões.....	45
Referências Bibliográficas	46
ANEXO I – RELAÇÃO DE PROTEÍNAS.....	51

LISTA DE TABELAS

Capítulo I

Tabela 1: Número de proteínas, relacionadas diretamente às sementes, com no mínimo cem aminoácidos, depositadas no banco de dados Swiss-Prot, em julho de 2006	11
Tabela 2: Número de proteínas depositadas no banco de dados Swiss-Prot, até julho de 2006, relacionadas a sementes em comparação com o total disponível na espécie	13
Tabela 3: Número de proteínas já depositadas no banco Swiss Prot em função do tecido em que se encontram: epicótilo, camada de aleurona, coleóptilo, semente imatura, mesocótilo, hipocótilo e endosperma, quantificado por espécie botânica	14
Tabela 4: Número de proteínas encontradas em cotilédones, quantificado por espécie botânica	15
Tabela 5: Número de proteínas encontradas em sementes e plântulas quantificado por espécie botânica	16

Capítulo II

Tabela 1: Número total de seqüências expressas analisadas (ESTs), microsátélites (SSRs) localizados, quantidade de <i>primers</i> desenhados, percentagens de SSRs existentes nas ESTs estudadas e percentagens de <i>primers</i> nos SSRs encontrados para cada espécie analisada nesse estudo	39
Tabela 2: Número de seqüências expressas (ESTs), microsátélites (SSRs) localizados e <i>primers</i> desenhados nas espécies estudadas e respectivo número de amplificações encontradas dentro da própria espécie e nas demais que constaram no estudo	41
Tabela 3: Freqüência das diferentes combinações de nucleotídeos encontradas nas espécies estudadas	41

LISTA DE FIGURAS

Capítulo I

- Figura 1:** Alinhamento de duas proteínas Glutathione S-transferase, encontradas nas espécies *Arabidopsis thaliana* e *Zea mays*. realizadas com o programa Clustal W (THOMPSON et al., 1994). A numeração dos resíduos se refere à posição do aminoácido na proteína. Aminoácidos idênticos estão indicados por asterisco (*), substituições conservadas, por dois pontos (:), substituições semi-conservadas por um ponto (.) 18
- Figura 2:** Alinhamento de duas proteínas Malato desidrogenase (EC 1.1.1.37), encontradas nas espécies *Cucumis sativus* e *Glycine max*. realizadas com o programa Clustal W (THOMPSON et al., 1994). A numeração dos resíduos refere-se à posição do aminoácido na proteína. Aminoácidos idênticos estão indicados por asterisco (*), substituições conservadas, por dois pontos (:), substituições semi-conservadas por um ponto (.) 18
- Figura 3:** Alinhamento de duas proteínas Malate sintase, (EC 2.3.3.9), encontradas nas espécies *Brassica napus* e *Glycine max*, realizado com o programa Clustal W (THOMPSON et al., 1994). A numeração dos resíduos refere-se à posição do aminoácido na proteína. Aminoácidos idênticos estão indicados por asterisco (*), substituições conservadas, por dois pontos (:), substituições semi-conservadas por um ponto (.) 19
- Figura 4:** Alinhamento de quatro proteínas catalase (EC 1.11.1.6), encontradas nas espécies *Gossypium hirsutum*, *Cucurbita pepo*, *Helianthus annuus* e *Zea mays*, realizado com o programa Clustal W (THOMPSON et al., 1994). A numeração dos resíduos refere-se à posição do aminoácido na proteína. Aminoácidos idênticos estão indicados por asterisco (*), substituições conservadas, por dois pontos (:), substituições semi-conservadas por um ponto (.) 20
- Figura 5:** Padrões de hidrofobicidade em duas proteínas encontradas em sementes. A: P15192 (*Chlorophyll a-b binding protein*) *Pinus sylvestris*. B: P07088 Proteína transportadora de Acil (ACP) - *Brassica campestris*..... 21
- Figura 6:** Padrões de hidrofobicidade em duas proteínas encontradas em sementes. A: P32765 proteína presente em sementes de *Theobroma cacao*. B:P46525: Proteína de choque frio (Cold-shock protein CS120) encontrada em sementes de *Triticum aestivum*. Valores positivos indicam regiões hidrofóbicas e valores negativos regiões hidrofílicas..... 22
- ### Capítulo II
- Figura 1:** Pipeline utilizado para a identificação *in silico* de microsátélites em leguminosas de interesse agrônomo 36

INTRODUÇÃO GERAL

Bioinformática é a aplicação da tecnologia de informação no gerenciamento de dados biológicos, uma disciplina científica de rápida evolução, impulsionado principalmente, pelo aumento verificado nos últimos anos da quantidade de dados biológicos armazenados em bancos públicos. Em paralelo, a literatura biológica também esta crescendo fortemente, tornando-se impossível, até mesmo para o pesquisador mais cuidadoso, manter-se atualizado com as informações necessárias da sua área de atuação sem o auxílio de ferramentas computacionais. Outra definição possível para bioinformática é a simples utilização de programas computacionais ou o desenvolvimento de algoritmos matemáticos para a análise de dados biológicos. Constituindo-se numa ciência integradora, que envolve a união de diversas áreas de conhecimento, iniciando pela engenharia de softwares, matemática, genômica comparativa, sequenciamento de genomas, ciência da computação, biologia molecular e outros, exigindo a participação de profissionais de diferentes áreas.

Tanto o conhecimento do funcionamento das ferramentas de bioinformática, quanto os problemas de biologia podem ser estudados ou elucidados por meio de "softwares". O desenvolvimento desses ou de algoritmos são uma nova área de pesquisa e trabalho para pesquisadores interessados em identificar novos genes ou novos padrões, e para matemáticos e cientistas, da área de computação, que desejam desenvolver e aplicar novos algoritmos computacionais a problemas da biologia. A expansão extraordinária da Internet na última década não só permitiu a disseminação deste conhecimento, mas também, funcionou como agente catalisador de cooperação para a maioria dos cientistas de ambas as áreas. Essa ampla disseminação das informações científicas faz com que haja uma democratização do conhecimento como jamais foi observada na história.

Essa gama de informação está disponível livremente em bancos de dados, sendo que os mais utilizados são: *Genbank/NCBI-National Center for Biotechnology Information* (<http://www.ncbi.nlm.nih.gov>), *EBI-European Bioinformatics Institute* (<http://www.ebi>.

ac.uk), *DDBJ - DNA Data Bank of Japan* (<http://www.ddbj.nig.ac.jp>), *PDB – Protein Data bank* (<http://www.rcsb.org/pdb>), *GDB – The Genome Database* (<http://www.gdb.org/>), *TIGR Databases - The Institute for Genomic Research* (<http://www.tigr.org/tdb>), *PIR-Protein Information Resource* (<http://www-nbrf.georgetown.edu>), *SWISS-PROT - Protein knowledge Base* (<http://us.expasy.org/sprot>), *KEGG-Kyoto Encyclopedia of Genes and Genomes* (<http://www.genome.ad.jp/kegg/>), *Soybean Genomics Initiative* (<http://soybean.ccg.umn.edu>), *Gramene: A Comparative Mapping Resource for Grains* (<http://www.gramene.org>) e *Soybase* (<http://soybase.agron.iastate.edu>). Os principais grupos de pesquisa que atuam na área genômica possuem profissionais que atuam com bioinformática, desenvolvendo a chamada pesquisa *in silico*, servindo de suporte para as atividades realizadas *in vitro* e *in vivo*.

Dentre os campos de atuação dessas promissoras ferramentas está o estudo das proteínas, incluindo a seqüência de aminoácidos, sua conformação tri dimensional, o órgão em que determinada proteína atua, a similaridade entre duas ou mais seqüências de aminoácidos e o comportamento dos diversos aminoácidos que a compõem quando expostos a presença de água.

Em paralelo, os marcadores moleculares tem sido largamente utilizados principalmente no melhoramento de plantas através da Seleção Assistida por Marcadores Moleculares e na Ciência e Tecnologia de Sementes, é possível, com a utilização de marcadores moleculares do tipo microssatélites, fazer a caracterização de cultivares visando identificar possíveis misturas varietais. Os microssatélites, são apenas um dos vários tipos de marcadores genéticos existentes, também chamados de SSRs. Uma limitação para o uso dessa tecnologia sempre foi o custo para o desenvolvimento dos primers. A utilização de ferramentas computacionais permite que sejam localizados de forma eficiente novos primers e que seja testado *in silico* a sua eficácia, gerando com isso redução nos custos e no tempo de obtenção.

A ciência e tecnologia de sementes ainda carece de informações detalhadas sobre as características que as muitas proteínas existentes possuem, bem como, a função e ação biológica. Muita informação existe, e muita coisa ainda há para ser descoberta e organizada. Buscou-se nesse trabalho utilizar algumas das ferramentas da bioinformática em prol da ciência e tecnologia de sementes. Visando atingir esse objetivo, são apresentados dois artigos, um primeiro que procura caracterizar as proteínas que atuam diretamente nas sementes das diversas espécies agrícolas; e, no segundo trabalho, são identificados *in silico* primers do tipo microssatélites presentes em seqüências expressas de espécies pertencentes a família fabaceae e consideradas de importância agrônômica.

CAPÍTULO I

ANÁLISE *IN SILICO* DE PROTEÍNAS DE SEMENTES
DESCRITAS NO BANCO DE DADOS SWISS-PROT

Resumo

MENEGHELLO, Geri Eduardo. **Análise *in silico* de proteínas presentes em sementes depositadas no banco de dados Swiss-Prot.** 2007. 61f. Tese (Doutorado) - Programa de Pós-Graduação em Ciência e Tecnologia de Sementes. Universidade Federal de Pelotas, Pelotas.

Dentre os principais componentes de reserva de uma semente destacam-se os carboidratos, os lipídeos e as proteínas. Além da função nutritiva, as proteínas têm diversas funções importantes nas sementes, sendo integrantes da biologia molecular da planta. O conhecimento detalhado das proteínas possibilita que sejam adotadas estratégias de melhoramento genético visando aumento de produção, resistência a patógenos, etc... Avanços na biologia molecular, genômica e proteômica impulsionaram a criação de um grande volume de dados sobre proteínas de diversas espécies vegetais, com informações sobre a sua funcionalidade nos diversos tecidos e órgãos. Para utilizar essas informações, é imprescindível a utilização de ferramentas computacionais. Essa necessidade impulsionou o surgimento e o desenvolvimento da bioinformática, que é a utilização de ferramentas computacionais para o estudo de dados biológicos. O objetivo deste trabalho foi quantificar as proteínas relacionadas às sementes já descritas e disponíveis no banco Swiss-Prot, e verificar a similaridade e padrão de hidrofobicidade das proteínas de mesma função em espécies diferentes. Para isso, foi feita uma consulta detalhada no banco de dados Swiss-Prot em busca de proteínas encontradas em sementes, plântulas e seus diversos tecidos. As seqüências encontradas foram agrupadas por tecido, sendo selecionadas e analisadas aquelas expressas no epicótilo, camada de aleurona, coleóptilo, cotilédones, endosperma, hipocótilo, mesocótilo, semente imatura, plântula e semente. Realizou-se alinhamentos entre as seqüências com função similar encontradas em diferentes órgãos e em diferentes espécies para verificar a similaridade existente entre as mesmas. Estudou-se a hidrofobicidade/hidrofilicidade nas proteínas encontradas, buscando identificar padrões. Existem 450 proteínas relacionadas a sementes depositadas no banco de dados Swiss-Prot. *Oryza sativa*, *Zea mays*, *Arabidopsis thaliana*, *Triticum aestivum*, *Hordeum vulgare*, *Glycine max* são espécies que possuem um maior número de proteínas de sementes estudadas. Não há um padrão de similaridade nas seqüências de aminoácidos das proteínas de mesma função.

Palavras chave: Bioinformática, sementes, estrutura de proteínas.

Abstract

MENEGHELLO, Geri Eduardo. ***In silico* analysis of proteins present in seeds stocked at the swiss-prot database**. 2007. 61f. Tese (Doutorado) - Programa de Pós-Graduação em Ciência e Tecnologia de Sementes. Universidade Federal de Pelotas, Pelotas.

The main reserve components in seeds are carbohydrates, lipids and proteins. Aside from their nutritious role, proteins have several important functions in seeds, being essential to the molecular biology of the plant. The detailed knowledge of proteins makes it possible to outline strategies of genetic improvement aimed at increasing production and tolerance to different pathologies, among other traits. Progress in the molecular biology, genomics and proteomics impelled the creation of a great volume of data on proteins of several vegetable species, with information about their functionality in the several tissues and organs they are part of. To use this information, it is necessary to rely on computational tools. This need impelled the appearance and development of bioinformatics, e.g. the use of computational tools for the study of biological data. The objective of this work was to quantify the proteins related to the seeds described already and available in the Swiss-Prot database, and to verify the similarity and hydrophobic pattern of the proteins with same function in different species. A detailed consultation was performed in the Swiss-Prot database, in search of proteins found in seeds, seedlings and their component tissues. The sequences found were grouped and analyzed according to the tissue in which they were expressed, epicotyl, aleurone, coleoptile, cotyledon, endosperm, hypocotyl, mesocotyl, immature seed, seedling and seeds. Alignments were performed among the sequences with similar function in different organs and in different species to verify similarities across tissues and species. The hydrophobic/ hydrophilic character in the proteins found was analyzed to identify patterns. Four hundred and fifty seed-related proteins are in stock at the Swiss Prot database. *Oryza sativa*, *Zea mays*, *Arabidopsis thaliana*, *Triticum aestivum*, *Hordeum vulgare*, *Glycine max*, which are the species that possess the largest number of seed proteins studied so far. There isn't a similarity in the sequences of amino acids of the proteins with same function.

Key-words: Bioinformatic, seeds, protein's structure.

INTRODUÇÃO

Nas próximas décadas a agricultura terá que produzir mais alimentos, rações e fibras do que em toda a história, será preciso desenvolver variedades com maior tolerância à seca, ao frio, ao calor e menos exigentes em nutrientes e pesticidas (HAMILTON et al., 2005). Além disso, o incremento da produção deverá ser acompanhado pelo aumento da responsabilidade ambiental. A produção de alimentos passa, inevitavelmente, pelo aumento da produção e produtividade de grãos e sementes. Ferramentas como a biotecnologia vegetal, genômica e proteômica propiciam uma oportunidade capaz de atingir esses objetivos (KUZMA, 2005).

Nas sementes, as proteínas não possuem apenas importância nutritiva, são partes ativas da maquinaria molecular (DOYLE et al., 1986). São originadas diretamente do gene, isto é, a sua seqüência de bases é traduzida em aminoácidos que serão unidos através de ligações peptídicas, na seqüência exata determinada pelo mRNA. Portanto, em uma proteína, tanto a seqüência como o número de aminoácidos, são especificamente determinados (NEDEL et al., 2003).

A característica mais importante de uma proteína é sua função, sendo assim, o custo energético de sintetizar uma proteína assegura que somente proteínas com funções necessárias para um organismo sejam produzidas (RIDGEN & MELLO, 2002). Durante a formação da semente, ocorre acúmulo de proteínas de reserva, forma como as plantas armazenam elementos como nitrogênio e enxofre, que são utilizados durante o processo germinativo, fase em que há demanda energética (OKITA et al., 1989). O armazenamento de proteínas com função nutricional em plantas, especialmente na família Fabaceae tem sido o objeto de muitos estudos que objetivam determinar suas características físicas, desenvolvimento, regulação, processos pós traducional, transporte e armazenamento (SLIGHTOM et al., 1985)

Todos os eventos metabólicos têm participação direta de proteínas, através de suas funções enzimáticas e regulatórios. Como exemplo dessa atividade pode-se destacar o

processo germinativo, quando é necessária a produção e ativação de enzimas (amilases) para degradar o amido armazenado no endosperma, usado para o desenvolvimento do embrião (ROGERS & MILLIMAN, 1983). As enzimas hidrolíticas são secretadas predominantemente na camada de aleurona e no escutelo, possuem a função de hidrolisar os polissacarídeos e proteínas no endosperma, gerando sub produtos que são translocados para o embrião visando suprir o crescimento da plântula (DOAN & FINCHER, 1988).

Algumas enzimas atuam em tecidos específicos, por exemplo, em sementes de cereais, a camada de aleurona e o embrião compartilham várias funções específicas, como a síntese e acúmulo de lipídeos, tolerância à dessecação e dormência (AALEN et al. 1994). Na camada de aleurona e no escutelo são observados níveis altos de enzimas concomitantemente com o desencadeamento da germinação, enquanto que durante a formação são observados baixos níveis (TIBBOT & SKADSEN, 1996)

Outro evento em que há participação ativa de proteínas específicas na maquinaria celular e molecular nas sementes é a biossíntese de ácidos graxos, que em plantas ocorre no cloroplasto dos tecidos verdes e nos plastídeos dos tecidos não fotossintéticos, (SHANKLIN & SOMERVILLE, 1991). Por outro lado, em sementes de espécies oleaginosas, quando inicia-se o processo germinativo, ocorre a hidrólise do óleo armazenado para ser usado como fonte de energia para nutrir o embrião em desenvolvimento, processo este que requer a participação de diversas enzimas localizadas em várias estruturas sub-celulares (FROMAN et al., 2000)

As proteínas participam dos diversos mecanismos que as plantas utilizam para proteger-se de infecções por patógenos. Dentre os quais podem ser relacionadas as alterações na parede celular, a síntese de fitoalexinas tóxicas e a acumulação de proteínas relacionadas à patogenezidade como a inibição de proteínas de fungos e enzimas hidrolíticas, que são sintetizadas em diversos tecidos vegetais em resposta a invasões de fungos (LEAH et al., 1991). Durante a maturação e a germinação as sementes são muito sensíveis a vírus, fungos e bactérias, sendo que várias proteínas estão envolvidas nos mecanismos de defesa (BLOCHET et al., 1993).

A participação de proteínas está evidenciada nos mais diversos processos metabólicos, e o conhecimento detalhado desses peptídeos, incluindo seqüência primária, secundária, estrutura tri dimensional, sítios ativos, entre outros, possibilita que sejam adotadas estratégias de melhoramento genético e/ou melhoramento genômico, visando maximizar os diversos processos envolvidos no melhoramento das culturas (BORÉM et al., 2003)

Avanços na biologia molecular, genômica e proteômica impulsionaram a criação de um grande volume de dados, tornando difícil o armazenamento, análise e interpretação sem o uso de modernas e potentes ferramentas computacionais. Esta necessidade impulsionou o surgimento e o desenvolvimento da bioinformática (GIBAS e JAMBECK, 2001). Ainda neste sentido, procurando associar proteínas às funções, a bioinformática pode e deverá, proporcionar suas maiores contribuições à biologia, principalmente na modelagem estrutural das biomoléculas, simulando *in silico* o seu comportamento *in vivo*. O conhecimento da estrutura terciária de uma proteína constitui uma informação valiosa para determinação de sua possível função, pois pode permitir a identificação de domínios conhecidos, como sítios catalíticos, sítios de modificação alostérica e outros. Além disso, através da determinação da estrutura tridimensional das proteínas, podem ser realizadas pesquisas mais direcionadas no sentido de encontrar inibidores, ativadores enzimáticos e outros ligantes (PROSDOCIMI et al., 2001).

O surgimento de enorme volume de dados requer a organização de Bancos de Dados, que armazenam informações compostas, basicamente, por seqüências de nucleotídeos e aminoácidos (RUDD et al., 2005). Alguns bancos especializaram-se no armazenamento de determinado tipo de informação como seqüência de nucleotídeos, aminoácidos ou outro. Outra classificação existente dentro dos bancos de dados é a divisão em bancos primários que são aqueles que derivam diretamente dos dados obtidos a partir do sequenciamento de ácidos nucléicos ou proteínas. Os bancos de dados primários podem conter além da seqüência em si, outros dados como uma tradução de uma seqüência de um clone de DNA, seqüências padrão (como sítios de fosforilação), promotores e outras anotações semelhantes. Os principais bancos de dados primários são o Genbank, o EBI (EMBL), e o DDBJ (BAXEVANIS & OUELLETTE, 2001). Os bancos secundários por sua vez, são bancos cujos dados são derivados dos bancos primários, as seqüências geralmente apresentam-se sem falhas, alinhadas e contendo as regiões mais conservadas em proteínas. Como exemplo de bancos cita-se o PDB e o SWISS-PROT (ATWOOD & PARRY-SMITH, 1999; HIGGINS & TAYLOR, 2000).

Os bancos de dados oferecem aos usuários diversas opções de busca, desde mineração de dados por espécie até por gene/proteína expressas em tecido específico. Dentre os bancos merece destaque o Swiss-Prot, que caracteriza-se por ser um banco de dados secundários sendo exclusivo de proteínas (ATWOOD & PARRY-SMITH, BAIROCH & APWEILER, 2000, 1999; HIGGINS & TAYLOR, 2000; BAXEVANIS & OUELLETTE, 2001).

A quantidade de informação disponível para cada espécie depende basicamente da

complexidade do genoma e da importância econômica ou científica que determinada espécie possui. Levando-se em consideração tecidos específicos também ocorre uma diversidade na quantidade de informações existentes. Há uma demanda para estudos referentes a proteínas que atuam especificamente nas sementes, independente da fase de desenvolvimento, processo fisiológico envolvido ou espécie vegetal. O conhecimento das proteínas existentes relacionadas a sementes possibilitará um melhor entendimento da fisiologia das mesmas.

O objetivo deste trabalho foi quantificar as proteínas relacionadas às sementes já descritas e disponíveis no banco Swiss-Prot.

MATERIAL E MÉTODOS

Realizou-se uma busca no banco de dados público Swiss-Prot [<http://expasy.org/sprot/>] (BAIROCH & APWEILER, 1997) por proteínas encontradas em sementes, plântulas e seus diversos tecidos relacionados, em todas as espécies vegetais disponíveis. A escolha por este banco, em detrimento de outros, foi em função de ser um banco validado (bancos secundários) do que os depósitos encontrados nos bancos primários (ATWOOD & PARRY-SMITH, 1999; HIGGINS & TAYLOR, 2000).

As seqüências encontradas foram agrupadas por tecido específico, sendo selecionadas e analisadas aquelas expressas no epicótilo, camada de aleurona, coleóptilo, cotilédones, endosperma, hipocótilo, mesocótilo, semente imatura, plântula e semente. Foram escolhidas seqüências com, no mínimo, 100 aminoácidos, evitando-se com isso, a análise de fragmentos muito pequenos, o que poderia levar a alguma interpretação com baixa consistência. Em cada tecido específico/órgão selecionado para a mineração de dados, as seqüências foram agrupadas e analisadas por espécie botânica e então, organizadas por função protéica.

Realizaram-se alinhamentos entre as seqüências com função similar encontradas em diferentes órgãos e em diferentes espécies para verificar a similaridade existente entre elas e assim estimar o grau de conservação da seqüência de aminoácidos no decorrer da evolução das espécies. Para tanto, foi usada a versão *online* do programa Clustal-W (<http://www.ebi.ac.uk/clustalw/>), que utiliza algoritmos propostos por THOMPSON, et al. (1994). Também foram realizados estudos de hidrofobicidade/hidrofilicidade nas proteínas encontradas, buscando identificar padrões, através do programa ProtScale (<http://www.expasy.org/tools/protscale.html>), usando a escala proposta por KYTE & DOOLITTLE (1982).

Todo o trabalho foi realizado através de pesquisa em bancos de dados públicos, e com a utilização de softwares livres, disponíveis gratuitamente na internet.

RESULTADOS E DISCUSSÃO

Quantificação das proteínas

Na tabela 1 é apresentado o número de proteínas depositadas junto ao banco de dados público Swiss-Prot (<http://us.expasy.org/prot>), cuja descrição indicava estarem presentes nos diversos tecidos específicos integrantes de uma semente ou plântula. Considerou-se que, mesmo atuando na fase da plântula, a proteína ou seu precursor estavam presentes na semente. O gene que codifica as proteínas identificadas na fase de plântula pode ter sido expresso e codificado para a tradução da proteína somente na fase de início da germinação, razão pela qual se torna importante o conhecimento de quando e em que tecido/órgão está presente e atuando determinada proteína.

Tabela 1: Número de proteínas relacionadas diretamente às sementes, com no mínimo cem aminoácidos, depositadas no banco de dados público Swiss-Prot, em julho de 2006.

Tecido/Órgão	Número de proteínas
Epicótilo	15
Camada de Aleurona	15
Coleóptilo	26
Cotilédones	88
Endosperma	68
Hipocótilo	51
Semente imatura	17
Mesocótilo	2
Semente	71
Plântula	72
Total	425

Atualmente, nos depósitos do Swiss-Prot, existem proteínas relacionadas com sementes para mais de trinta espécies botânicas. Esse número, embora expressivo, ainda é insignificante frente à grande diversidade de espécies existentes e, considerando a magnitude do banco, que ultrapassa o número de 120 mil depósitos quando da realização

desse levantamento. Destacam-se os cotilédones, endosperma e hipocótilo como os tecidos em que se encontra um maior número de seqüências depositadas atualmente.

Os tecidos/órgãos em que se observou maior quantidade de proteínas caracterizadas foram os cotilédones, seguido da semente inteira, plântula e endosperma. Acredita-se que essa maior quantidade nesses tecidos, deve-se à abundância de tecido existente. Por outro lado, o mesocótilo foi o tecido em que houve o registro de menor número de proteínas, seguido da camada de aleurona e do epicótilo.

Na tabela 2 há uma relação comparativa, para as espécies com maior quantidade de proteínas depositadas, entre a quantidade de proteínas presentes nos órgão/tecidos selecionados, e a quantidade total de proteínas descritas para a espécie. A percentagem de proteínas presentes, descritas como atuantes diretamente em sementes, variou de 0,73% em *Arabidopsis* até 29,17% em abóbora, ficando o maior número de espécies em patamares de aproximadamente 10% das proteínas. Destaca-se a espécie *Arabidopsis thaliana* por apresentar 4410 proteínas depositadas, uma quantidade muito superior que as demais espécies em razão de ser a espécie considerada modelo genômico entre as dicotiledôneas, inclusive sendo a primeira espécie vegetal com seu genoma seqüenciado (THE ARABIDOPSIS GENOME INITIATIVE, 2000¹).

Nas tabelas 3, 4 e 5 estão relacionados o número de proteínas por espécie, dentro de cada tecido específico. Espécies mais estudadas e com maior importância agrícola, como o milho, trigo e arroz apresentaram proteínas estudadas numa maior diversidade de tecidos, com exceção das proteínas presentes nos cotilédones, por se tratarem de espécies monocotiledôneas. GALARDO et al. (2001) relatam que, do total de proteínas identificadas em sementes de *Arabidopsis*, 56% possuíam atividade em todas as etapas da germinação e 19% passavam a ser ativas durante o processo. A perda de atividade foi observada em 2% das proteínas, 10% iniciavam sua função somente na protusão da radícula, ao passo que, 12% perdiam atividade nessa fase, evidenciando, com isso, que as proteínas, das sementes, agem de forma diferente, nos diversos estádios de desenvolvimento.

¹A autoria deste trabalho deverá ser "The Arabidopsis Genome Initiative". Uma lista completa dos colaboradores aparece no final do artigo original (**Nature**, v.408, p. 796-815).

Tabela 2: Número de proteínas depositadas no banco de dados Swiss-Prot, até julho de 2006, relacionadas a sementes em comparação com o total disponível na espécie.

Espécie	Proteínas relacionadas a sementes	Total de proteínas da espécie no banco	% relacionada a semente
<i>Oryza sativa</i>	55	958	5,74
<i>Zea mays</i>	50	546	9,16
<i>Arabidopsis thaliana</i>	32	4410	0,73
<i>Triticum aestivum</i>	32	304	10,53
<i>Hordeum vulgare</i>	30	278	10,79
<i>Glycine max</i>	28	286	9,79
<i>Pisum sativum</i>	21	339	6,19
<i>Brassica napus</i>	12	142	8,45
<i>Phaseolus vulgaris</i>	10	95	10,53
<i>Cucurbita maxima</i>	7	24	29,17
<i>Lycopersicon esculentum</i>	7	326	2,15
<i>Vicia faba</i>	7	57	12,28
<i>Avena sativa</i>	6	25	24,00
<i>Cucumis sativus</i>	6	65	9,23
<i>Gossypium hirsutum</i>	6	83	7,23

Fonte: Swiss-Prot. <http://www.expasy.org/sprot>.

A “semente inteira” e o “cotilédone” foram os partes das sementes em que foram estudadas para um maior número de espécies (Tabelas 4 e 5). Provavelmente isso se deve ao fato de que houve extração e estudo dessas proteínas sem o isolamento de tecidos. Por outro lado o “mesocótilo” foi o órgão que apresentou proteínas descritas em menor número de espécies, apenas uma, seguido do coleóptilo, onde somente três espécies tiveram estudo de proteínas nessa região.

Similaridade entre proteínas de mesma função

Visando comparar a similaridade entre proteínas com mesma função na classificação pelo EC (*Enzyme Commission*), realizou-se um alinhamento entre seqüências de aminoácidos de diversas proteínas, sendo apresentado nas figuras 1 a 4 exemplos representativos. Na figura 1, comparando proteínas Glutathione S-transferase (EC 2.5.1.18) em *A. thaliana* (P46421), presente no cotilédone, endosperma e hipocótilo e em *Z. mays* (P04907), encontrada no coleóptilo. Verificou-se que não há um bom alinhamento entre as seqüências, provavelmente por serem proteínas presentes em tecidos diferentes e em espécies que evolutivamente não são muito próximas. Esse resultado é semelhante ao encontrado por MARTIN et al. (2001), que ao comparar duas

seqüências de aminoácidos de feijão da proteína Zeatina O-glicosiltransferase (*zox1* e *zog1*) e uma de milho (*cisZOG1*) observou alta conservação nas duas seqüências do feijão, porém baixa, quando esta foi comparada com o milho.

Tabela 3: Número de proteínas já depositadas no banco, em função do tecido em que se encontram: epicótilo, camada de aleurona, coleótilo, semente imatura, mesocótilo, hipocótilo e endosperma, quantificado por espécie botânica.

Orgão	Espécie	Proteínas
Epicótilo	<i>Cicer arietinum</i>	2
	<i>Glycine max</i>	4
	<i>Lupinus luteus</i>	3
	<i>Pisum sativum</i>	5
	<i>Zea mays</i>	1
Aleurona	<i>Avena fátua</i>	1
	<i>Cyamopsis tetragonoloba</i>	1
	<i>Hordeum vulgare</i>	10
	<i>Oryza sativa</i>	2
	<i>Zea mays</i>	1
Coleoptilo	<i>Avena sativa</i>	2
	<i>Oryza sativa</i>	9
	<i>Zea mays</i>	15
Semente Imatura	<i>Cucurbita máxima</i>	1
	<i>Fagopyrum esculentum</i>	2
	<i>Oryza sativa</i>	10
	<i>Phaseolus lunatus</i>	1
	<i>Prunus dulcis</i>	1
Mesocotilo	<i>Avena sativa</i>	2
Hipocótilo	<i>Arabidopsis thaliana</i>	14
	<i>Glycine max</i>	2
	<i>Helianthus annuus</i>	1
	<i>Lycopersicon esculentum</i>	5
	<i>Phaseolus aureus</i>	12
	<i>Phaseolus vulgaris</i>	3
	<i>Ricinus communis</i>	2
	<i>Solanum melongena</i>	10
	<i>Triticum aestivum</i>	1
	<i>Zea mays</i>	1
Endosperma	<i>Abrus precatorius</i>	1
	<i>Avena sativa</i>	3
	<i>Cocos nucifera</i>	1
	<i>Coriandrum sativum</i>	1
	<i>Cucurbita maxima</i>	1
	<i>Fagopyrum esculentum</i>	1
	<i>Hordeum vulgare</i>	16
	<i>Oryza sativa</i>	7
	<i>Ricinus communis</i>	8
	<i>Secale cereale</i>	2
	<i>Triticum aestivum</i>	15
<i>Zea mays</i>	12	

Fonte: Swiss-Prot <http://www.expasy.org/sprot>.

Tabela 4: Número de proteínas encontradas em cotilédones, quantificado por espécie botânica

Órgão	Espécie	Número de proteínas
Cotilédones	<i>Arabidopsis thaliana</i>	1
	<i>Bauhinia unguolata</i>	2
	<i>Bauhinia variegata</i>	1
	<i>Brassica campestris</i>	1
	<i>Brassica napus</i>	4
	<i>Canavalia ensiformis</i>	2
	<i>Carthamus tinctorius</i>	1
	<i>Castanea crenata</i>	1
	<i>Castanea sativa</i>	1
	<i>Citrullus lanatus</i>	1
	<i>Cucumis sativus</i>	6
	<i>Cucurbita maxima</i>	5
	<i>Cucurbita moschata</i>	1
	<i>Cucurbita pepo</i>	2
	<i>Glycine max</i>	13
	<i>Gossypium hirsutum</i>	6
	<i>Helianthus annuus</i>	1
	<i>Ipomoea nil</i>	5
	<i>Leucaena glauca</i>	1
	<i>Linum usitatissimum</i>	1
	<i>Lycopersicon esculentum</i>	1
	<i>Manihot esculenta</i>	1
	<i>Phaseolus lunatus</i>	3
	<i>Phaseolus vulgaris</i>	3
	<i>Picea abies</i>	1
	<i>Pinus sylvestris</i>	4
	<i>Pisum sativum</i>	6
	<i>Raphanus sativus</i>	1
	<i>Ricinus communis</i>	1
	<i>Sesamum indicum</i>	1
	<i>Sinapis alba</i>	3
	<i>Vicia faba</i>	7
<i>Vigna mungo</i>	1	

Fonte: Swiss-Prot <http://www.expasy.org/sprot>.

Por outro lado, como pode ser observado na figura 2, a proteína Malato desidrogenase (EC 1.1.1.37) encontrada nos cotilédones de *C. Sativus* (P46488) e de soja (P37228) apresentam uma alta similaridade, evidenciada pelo alinhamento praticamente perfeito entre ambas. Alta similaridade entre proteínas de espécies diferentes também foi observado por HEPPARD et al. (1996) ao comparar seqüências de aminoácidos da proteína FAD-2 de soja com *Arabidopsis*.

Tabela 5: Número de proteínas encontradas em sementes e plântulas quantificado por espécie botânica.

Órgão	Espécies	Número de proteínas
Semente	<i>Abrus precatorius</i> (Jequitiri)	2
	<i>Allium cepa</i>	1
	<i>Arabidopsis thaliana</i>	3
	<i>Avena sativa</i>	2
	<i>Bauhinia bauhinioides</i>	2
	<i>Brassica campestris</i>	1
	<i>Brassica juncea</i>	1
	<i>Brassica napus</i>	6
	<i>Canavalia gladiata</i> (fava)	1
	<i>Capparis masaikai</i>	4
	<i>Coffea arábica</i>	1
	<i>Cyamopsis tetragonoloba</i>	1
	<i>Fagus sylvatica</i>	1
	<i>Glycine max</i>	4
	<i>Helianthus annuus</i>	1
	<i>Hordeum vulgare</i>	1
	<i>Impatiens balsamina</i> (Balsamo)	1
	<i>Lupinus angustifolius</i>	1
	<i>Lupinus arboreus</i>	1
	<i>Macadamia integrifolia</i>	1
	<i>Malva parviflora</i>	1
	<i>Oryza sativa</i>	5
	<i>Phaseolus vulgaris</i>	3
	<i>Pisum sativum</i>	7
	<i>Psophocarpus tetragonolobus</i>	1
	<i>Ricinus communis</i>	1
<i>Sesamum indicum</i>	2	
<i>Sinapis alba</i> (<i>Brassica hirta</i>)	1	
<i>Theobroma cação</i>	1	
<i>Triticum aestivum</i>	12	
<i>Zea mays</i>	12	
Plântula	<i>Allium cepa</i>	2
	<i>Allium sativum</i>	1
	<i>Arabidopsis thaliana</i>	14
	<i>Brassica napus</i>	2
	<i>Glycine max</i>	3
	<i>Helianthus annuus</i>	1
	<i>Hordeum vulgare</i>	3
	<i>Lycopersicon esculentum</i>	6
	<i>Medicago sativa</i>	1
	<i>Oryza sativa</i>	18
	<i>Phaseolus vulgaris</i>	1
	<i>Pinus taeda</i>	2
	<i>Pisum sativum</i>	3
<i>Triticum aestivum</i>	4	
<i>Vitis vinifera</i>	3	
<i>Zea mays</i>	8	

A proteína catalase (EC 1.11.1.6), encontrada em cotilédones e endosperma de *G. hirsutum* (P17598), *C. pepo* (P48350) e *H. annuus* (P45739) e no hepícolo de *Z. mays* (P19123), também apresentou um bom índice de similaridade entre as espécies (figura 4), mesmo sendo descrita como presente em órgãos diferentes e em quatro espécies pertencentes a famílias botânicas distintas, que evoluíram de forma diferenciada, porém mantendo o mesmo padrão para essa enzima.

Observa-se, ao comparar os alinhamentos das figuras 1 a 4, que as diversas proteínas apresentam grau de conservação diferenciado, pois algumas se mantiveram praticamente idênticas entre espécies distintas e outras alteram significativamente a composição de aminoácidos ao longo da sua cadeia peptídica.

P13244 <i>B. napus</i>	--MELETSVYRPNVAVYDSPDGEVVRGRYDQVFAKILTRDALGFVAELQREFRGHVRYAM	58
P45458 <i>G. max</i>	GTYGYPTPAVKKIESYDVPPEGVDIRGRYDAEFKILTKDALKFVADLQREFRNHIKYAL	60
	*. . : : ** *:*:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:* *:*:*:	
P13244 <i>B. napus</i>	ECRREVKRRYNSGAVPGFDPSTKFI RDGEWVCASVPPAVADRVEITGPVERKMIINALN	118
P45458 <i>G. max</i>	ECRREAKKKYNEGALPEFDPATYIREQEWVCAVPPAVADRKVEITGPVDRKMVINALN	120
	*****. *:.*.*.*.* *:*:*.*. :*:*: *****.*****.*:*:*:*:*:*	
P13244 <i>B. napus</i>	SGAKVFMADFEDALSPSWENLMRGQVNLKDAVDGSI TFNDKARNKVYKLNQVAKLFVRP	178
P45458 <i>G. max</i>	SGAKVFMADFEDALSPGWENLMRGQVNLKDAVAGTISLHDKARNRVYKLNQVAKLFVRP	180
	*****.*****.*:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:*	
P13244 <i>B. napus</i>	RGWHLPEAHILIDGEPATGCLVDFGLYFFHNYAKFRQTQGSFGFPFYL PKMEHSREAKI	238
P45458 <i>G. max</i>	RGWHLPEAHILIDGEPATGCLVDFGLYFYHSYSAFRRTQGAGFGFPFYL PKMEHSREAKI	240
	*****.*****.*:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:*	
P13244 <i>B. napus</i>	WNSVFERAEKMAGIERGSIRATVLIETLPVAVFQMNELLYELRDHSVGLNCGRWDIYFSYV	298
P45458 <i>G. max</i>	WNNVFEKAEKVAGIERGSIRATVLIETLPVAVFQMNELLYELKDHVGLNCGRWDIYFSYV	300
	**.*:*	
P13244 <i>B. napus</i>	KTFQAHPRLLPDRVLVGMGQHFMRYSDDLIRTCCHRGVHAMGGMAAQIPIRDPKANE	358
P45458 <i>G. max</i>	KTFQAHPRLLPDRVLVGMGQHFMRYSDDLIRTCCHRGVHAMGGMAAQIPIKEDPVANE	360
	*****.*****.*:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:	
P13244 <i>B. napus</i>	MALDLVKKDKLREVRAGHDGTWAAHPGLIPI CMDAFS-HMGNPNQIKSMKRDDASAIT	417
P45458 <i>G. max</i>	VALELVKDKLREVKAGHDGTWAAHPGLI PACMEIFNNMGNASNQIDTVKREDGANITE	420
	:**:*:*:*:*:*:*:*:*:*:*:* *:*:* * : * :*** .***. :*:*:*.*:	
P13244 <i>B. napus</i>	EDLLQIPRGVRTLEGLRLNTRVGIQYLAAWLTGSGSVPLYNLMEDAATAEISRVQNWQWI	477
P45458 <i>G. max</i>	QDLLQIPRGARTMEGLRLNTRVGIQYVAAWLTGSGSVPLYNLMEDAATAEISRVQNWQWL	480
	:*****.*:*:*:*:*:*:*:*:*:*:*:* *:*:*:*:*:* *:*:*:*:*:* *:*:*:	
P13244 <i>B. napus</i>	RYGVELDGDGLGVRVSKELFGRVVEEMERIEKEVKGDKFKRGMVKEACKMFTKQCTAAE	537
P45458 <i>G. max</i>	KYGVELNGDGLGVKNKELFGRVVEEMARIEKEVGTETKFKEGMYKEACKIFTRQCTSPM	540
	:*****.*:*:*:*:*.* *:*:*:*:* *:*:*:*:* *:*:*:*:*:* *:*:*:*:	
P13244 <i>B. napus</i>	LDDFLTLAVYDHIVAHYPINASRL	561
P45458 <i>G. max</i>	LDDFLTLDAYNYIVVHHPRETSKL	564
	*****.*:*:*:*:*.* *:*:*:	

Figura 3: Alinhamento de duas proteínas Malato sintase, (EC 2.3.3.9), encontradas nas espécies *Brassica napus* e *Glycine max*, realizado com o programa Clustal W (THOMPSON et al., 1994). A numeração dos resíduos refere-se à posição do aminoácido na proteína. Aminoácidos idênticos estão indicados por asterisco (*), substituições conservadas, por dois pontos (:), substituições semi-conservadas por um ponto (.)

P17598	<i>G. hirsutum</i>	--MDPYKHRPSSAFNSPFWTTNSGAPVWNNNSLTVGPRGQYLLEDYHLV	48
P48350	<i>C. pepo</i>	--MDPYRHRPSSAFNAPFWTTNSGAPVWNNNSMTVGPGRGPIILEDYHLV	48
P45739	<i>H. annus</i>	--MDPYKVRSSAYNAPFWTTNSGAPVWNNNSLTVGSRGPIILEDYHLV	48
P18123	<i>Z. mays</i>	MTMDPTKFRPSSSHDTTVTTTNAGAPVWNDNEALTVGPRGPIILEDYHLI	50
		*** :.*.*.*.:.:.***:****:*.:.:****.* *****:	
P17598	<i>G. hirsutum</i>	EKLANFDRERI PERVVHARGASAKGFFEVTHDI SHLTCADFLRAPGVQTP	98
P48350	<i>C. pepo</i>	EKLANFDRERI PERVVHARGASAKGFFEVTHDI TNLSCADFLRAPGVQTP	98
P45739	<i>H. annus</i>	EKLANFDRERI PERVVHARGASAKGFFEVTHDI TALTCADFLRAPGVQTP	98
P18123	<i>Z. mays</i>	EKVAHFARERI PERVVHARGASAKGFFECTHDVTSLTCADFLRAPGV RTP	100
		:*.* ***:****:*.:.:****.* *****:**	
P17598	<i>G. hirsutum</i>	VI VRFSTVIHERGSPETLRDPRGFVAVKFYTREGNFDLVGNNFPVFFIRDG	148
P48350	<i>C. pepo</i>	VI VRFSTVIHERGSPETLRDPRGFVAVKFYTREGNFDLVGNNFPVFFIRDG	148
P45739	<i>H. annus</i>	VI VRFSTVIHERGSPETLRDPRGFVAVKFYTREGNFDLVGNNFPVFFIRDG	148
P18123	<i>Z. mays</i>	VI VRFSTVIHERGSPETIRDPGRGFVAVKFYTREGNWDL LGNNFPVFFIRDG	150
		*****:*****:*.:.:****.* *****:	
P17598	<i>G. hirsutum</i>	MKFPDMVHALKPNPKSHIQENWRILDFFSHHPESLHMFTFLFDDLGVPQD	198
P48350	<i>C. pepo</i>	MKFTRHVHPLKPNPKSHIQENWRILDFFSHHPESLNMFSFLFDDIGIPQD	198
P45739	<i>H. annus</i>	MKFPDMVHSLKPNPKSHIQEDWRIMDFFSHHPESLHMFTFLFDDIGVPQD	198
P18123	<i>Z. mays</i>	IKFPDVIHAFKPNPRSHVQEYWRVDFLFLSHLPESLHTFFFLFDDVGVPSD	200
		:**.* :*.:.:****:*** ***:***:****.* *****: * ***:**:*.* *	
P17598	<i>G. hirsutum</i>	YRHMDSGVTNTYTLINKAGKAHYVKFHWKPTCGVKCLL-EDEAIKVGGAN	247
P48350	<i>C. pepo</i>	YRHMDSGVTNTYTLINKAGKAHYVKFHWKPTCGVKCLL-EEDAIRVGGSN	247
P45739	<i>H. annus</i>	YRHMDSGVTNTYTLINKAGKAHYVKFHWKPTCGVKCLL-EEEAIRKIGGAN	247
P18123	<i>Z. mays</i>	YRHMDSGVTNTYTFVSAAGKAQVVKFHWKPTCGVRCILTDEEAALVGGRN	250
		:* **:.. *****:****:*.:.:****.* *****:**	
P17598	<i>G. hirsutum</i>	HSHATQDLYDSIAAGNYPEWKLFIQTIDPDHEDKDFDPLDVTKTWPEDI	297
P48350	<i>C. pepo</i>	HSHATQDLYDSIAAGNYPEWKLFIQTIDPDHEDKYDFDPLDVTKTWPEDI	297
P45739	<i>H. annus</i>	HSHATQDLYDSIAAGNYPEWKLFIQTIDPDHEDRLDFDPLDVTKTWPEDI	297
P18123	<i>Z. mays</i>	HSHATQDLYDSIAAGSFPWTLVQVMDPDTTEEQYDFDPLDVTKTWPEDL	300
		*****:*****:*.:.:****.* *****:** ***** *****:	
P17598	<i>G. hirsutum</i>	LPLQPVGRLVVLNKNIDNFFAENEQLAFCPAIVVPGIYYSDDKLLQTRIFS	347
P48350	<i>C. pepo</i>	LPLQPVGRMVLNKNIDNFFAENEQLAFCPAIVVPGVYYSDDKLLQTRIFS	347
P45739	<i>H. annus</i>	FPLQPVGRLVVLNKNIDNFFAENEQLAFCPAIVVPGIYYSDDKLLQTRIFS	347
P18123	<i>Z. mays</i>	LPLRPVGRVLDNRVNDNFFNENEQLAFGPGLVVPGIYYSDDKMLQCRVFA	350
		:**:* ***:***:***:***:***:***:***:***:***:***:***:***:***:***	
P17598	<i>G. hirsutum</i>	YSDTQRHRLGPNYLQLPANAPKCAHNNHHEGFMNFMHRDEEINYFPSRY	397
P48350	<i>C. pepo</i>	YADTQRHRLGPNYLQLPANAPKCAHNNHHEGFMNFMHRDEEVNYPFSPR	397
P45739	<i>H. annus</i>	YSDTQRHRLGPNYLQLPANAPKCAHNNHYDGFNFMHRDEEIDYFPSRY	397
P18123	<i>Z. mays</i>	YADTQRYRLGPNYMLPLVNAAPKCAHNNHYDGMNFMHRDEEVYYPSPRH	400
		*.***** ***** **.* ***:*****:*. *****:*.*****:	
P17598	<i>G. hirsutum</i>	DPVRHAEM-FPIPPAVCTGRREKCIIEKENNFKQPGERYRSWAADRQERF	446
P48350	<i>C. pepo</i>	DPSRHAER-YPHPPAVCSGKRERCIIEKENNFKEPGERYRSWTPDRQERF	446
P45739	<i>H. annus</i>	DPARHAEQ-YPIPPVRLSGKRDKCVIEKENNFKQPGERYRSFSPDRQERF	446
P18123	<i>Z. mays</i>	APLRQAAPPPLPPRPVAGRREKATIRKPNDFKQPGERYRSWDADRQDRF	450
		* **:* * ** :*:*:.. *.* ***:*****: .****:**	
P17598	<i>G. hirsutum</i>	ICRWVDALSDPRVTHEIRSIWISYWSQADKSVGQKLASLLNVRPSI	492
P48350	<i>C. pepo</i>	VRRWVDALSDTRVTHEIRSIWISYWSQADRSLGQKLASHLNVRPSI	492
P45739	<i>H. annus</i>	INRVVGLSDPRVTHEIRSIWISYWSQADKSLGQKIASRLNVRPNY	492
P18123	<i>Z. mays</i>	VRRFADSLGHPKVSQELRSIWIDLLAKCDASLGMKIATRNLNMKANM	496
		: * ..*..:*.:.:****:.. :.:. * ***:***:***:***:***:..	

Figura 4: Alinhamento de quatro proteínas catalase (EC 1.11.1.6), encontradas nas espécies *Gossypium hirsutum*, *Cucurbita pepo*, *Helianthus annus* e *Zea mays*, realizado com o programa Clustal W (THOMPSON et al., 1994). A numeração dos resíduos refere-se à posição do aminoácido na proteína. Aminoácidos idênticos estão indicados por asterisco (*), substituições conservadas, por dois pontos (:), substituições semi-conservadas por um ponto (.)

Padrão de hidrofobicidade

O padrão de comportamento de uma proteína na presença de água depende da seqüência de aminoácidos que a compõe, podendo possuir regiões hidrofóbicas ou hidrofílicas. Na figura 5A é apresentado o gráfico de hidrofobicidade da proteína P15192 (Chlorophyll a-b binding protein), enzima atuante na síntese de clorofila e encontrada nos cotilédones e no endosperma da espécie *Pinus sylvestris*, onde se pode observar,

claramente, dois pontos entre os aminoácidos de número 40 – 60 e 100 – 120 com comportamento de hidrofobicidade oposto ao restante da molécula, padrão típico de proteínas que são sintetizadas em um local e transportadas. Provavelmente tratam-se de regiões de ancoragem ou regiões em que a proteína ultrapassa uma membrana. Com relação à Proteína Transportadora de Acil (P07088) (Figura 5b), encontrada em cotilédones e endosperma de *Brassica campestris*, há uma variação muito mais intensa de hidrofobicidade e hidrofiliidade ao longo da cadeia de aminoácidos, padrão característico de proteínas citosólicas. Comportamento similar foi observado por ZUREK & CLOUSE (1994) na proteína BRU1 em soja e por DOAN & FINCHER (1988) na proteína carboxipeptidase I de cevada.

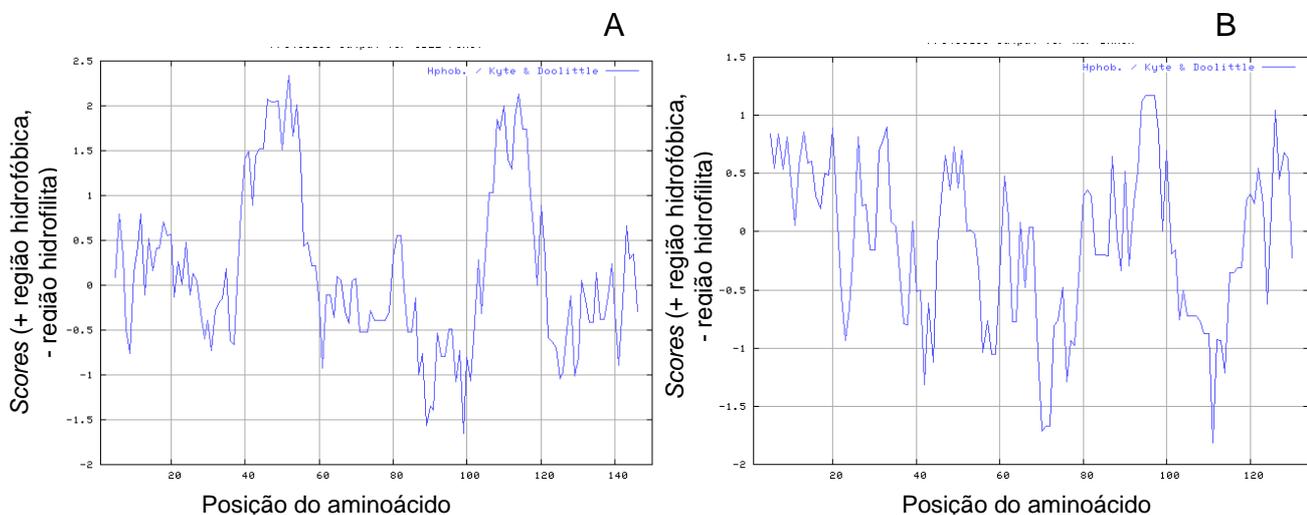


Figura 5: Padrões de hidrofobicidade em duas proteínas encontradas em sementes. A: P15192 (*Chlorophyll a-b binding protein*) *Pinus sylvestris*. B: P07088 Proteína transportadora de Acil (ACP) - *Brassica campestris*.

Na figura 6A é demonstrado o padrão de hidrofobicidade de uma proteína (P32765), presente em sementes de *Theobroma cacao*. Percebe-se uma região na parte inicial, até o aminoácido 30, com uma alta hidrofobicidade e no restante da proteína é formada por aminoácidos predominantemente de caráter hidrofílico. Esse padrão caracteriza a presença, nessa região C-terminal de um peptídeo sinal, concordando a descrição realizada por NARAYANAN et al. (2007). Já na proteína de choque frio (*Cold-shock protein* - CS120) (P46525) observada em plântulas de *Triticum aestivum*, figura 6B, há um predomínio, ao longo de toda a cadeia, de regiões altamente hidrofílicas, com apenas dois pontos hidrofóbicos.

A quantidade enorme de informação disponível e gerada continuamente não pode

ser analisada sem o auxílio da informática. Computadores, bancos de dados de grande capacidade, procura e ferramentas de análise e conexões de Internet de alta qualidade, são agora essenciais para qualquer grupo de pesquisa interessado no aumento da eficiência de suas ações. Além da grande quantidade de dados existentes, as possibilidades de combinações em qualquer análise que envolve seqüências primárias de aminoácidos que compõem uma proteína, gera combinações que se multiplicam exponencialmente, a cada adição de uma nova seqüência ou de um novo aminoácido.

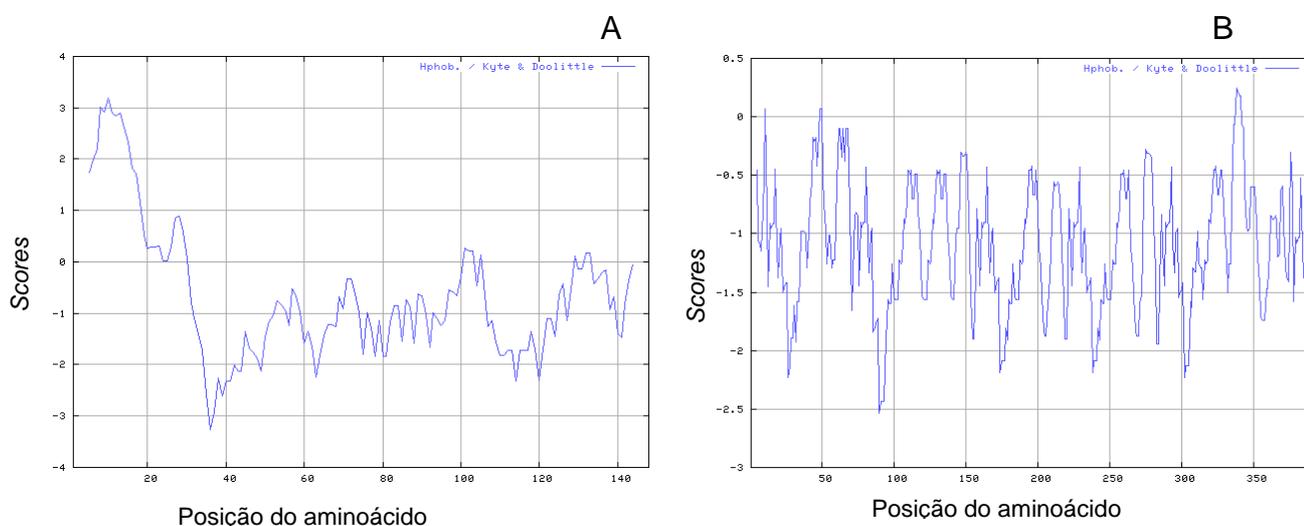


Figura 6: Padrões de hidrofobicidade em duas proteínas encontradas em sementes. A: P32765 proteína presente em sementes de *Theobroma cacao*. B:P46525: Proteína de choque frio (Cold-shock protein CS120) encontrada em sementes de *Triticum aestivum*. Valores positivos indicam regiões hidrofóbicas e valores negativos regiões hidrofílicas.

Essas significativas mudanças propiciadas pelo avanço da biologia molecular, associadas à bioinformática tenderão a impulsionar a comunidade científica a colocar maior ênfase na integração das ações voltadas a elucidação dos mecanismos de análise computacional de dados e simulações do comportamento *in silico* de biomoléculas.

CONCLUSÕES

Existem 450 proteínas relacionadas a sementes depositadas no banco de dados Swiss-Prot, com informações suficientes para determinar similaridades e padrões de hidrofobicidade;

Oryza sativa, *Zea mays*, *Arabidopsis thaliana*, *Triticum aestivum*, *Hordeum vulgare* e *Glycine max* são as espécies que possuem o maior número de proteínas de sementes depositadas no banco.

REFERÊNCIAS BIBLIOGRÁFICAS

AALEN, R.B.; OPSAHL-FERSTAD, H.G.; LINNESTAD, C.; OLSEN, O.A. Transcripts encoding an oleosin and a dormancy-related protein are present in both the aleurone layer and the embryo of developing barley (*Hordeum vulgare* L.) seeds. **Plant Journal**, Oxford v. 5, n.3, p.385-396. 1994.

ATWOOD, T.K.; PARRY-SMITH, D.J. **“Introduction to Bioinformatics”**, Pearson Education Ltd., Delhi, 1999.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acid Reserch**, Oxford, v.25, n.1, p.31-36. 1997.

BAXEVANIS, A.D.;OUELLETTE, B.F.F. **Bioinformatics - A practical guide to the analysis of genes and proteins**, 2nd Edition, Wiley, New York, 2001.

BLOCHET, J.; CHEVALIERB, C.; FORE, E.; PEBAY-PEYROULA, E.; GAUTIERD, M.F.; JOUDRIERD, P.; PCZOLET, M.; MARIONA, D. Complete amino acid sequence of puroindoline, a new basic andcystine-rich protein with a unique tryptophan-rich domain, isolated from wheat endosperm by Triton X-l 14 phase partitioning **FEBS LETTERS**, Amsterdam, v.329, n.3, p. 336-340. 1993.

BORÉM, A.; GIÚDICE, M.; SEDIYAMA, T. **Melhoramento Genômico** Viçosa: UFV, 2003. 224p.

DOAN, N.P.; FINCHERG, G.B. The A- and B-chains of Carboxypeptidase I from Germinated Barley Originate from a Single Precursor Polypeptide **The Journal of Biological Chemistry**, Bethesda, v.263, n.23, p.11106-11110, 1998.

DOYLE, J.J.; SCHULER, M.A.Q.; GODETTE, W.D.; ZENGER, V.; BEACHYN, R.N.; SLIGHTOM, J.L. The glycosylated seed storage proteins of *Glycine max* and *Phaseolus*

vulgaris **The Journal of Biological Chemistry**, Bethesda, v.261, n.20, p.9228-9238. 1986.

FROMAN, B.E.; EDWARDS, P.C.; BURSCH, A.G.; DEHESH, K. ACX3, a Novel Medium-Chain Acyl-Coenzyme A Oxidase from Arabidopsis. **Plant Physiology**, Washington, v. 123, p.733-741, 2000.

GALLARDO, K.; JOB, C.; GROOT, P.C.; PUYPE, M.; DEMOL, H.; VANDEKERCKHOVE, J.; JOB D. Proteomic Analysis of Arabidopsis Seed Germination and Priming. **Plant Physiology**, Washington, v.126, p.835–848, 2001.

GIBAS, C.; JAMBECK, P. **Desenvolvendo bioinformática: ferramenta de software para aplicações em biologia**, Rio de Janeiro, Editora Campus, 2001.

HAMILTON, R.; FLAVELL, R.B.; GOLDBERG, R.B. BIOTECNOLOGIA VEGETAL: Avanços em Alimentos, Energia e Saúde. **eJournal USA: Economic Perspectives**. Washington, v.10, n.4, p.17-20, 2005.

HEPPARD, E.P.; KINNEY, A.J.; STECCA, K.L.; MIAO, G. Developmental and growth temperature regulation of two different microsomal *w*-6 desaturase genes in soybeans. **Plant Physiology**, Washington, v.110, p.311-319. 1996.

HIGGINS, D.; TAYLOR, W. “Bioinformatics – sequence, structure and databanks”, **The Practical Approach Series**, Oxford University Press, New York, 2000.

KUZMA, J. Desafios globais e biotecnologia **eJournal USA: Economic Perspectives**. Washington, v.10, n.4. p.06-09. 2005

KYTE, J. & DOOLITTLE R.F. A simple method for displaying the hydropathic character of a protein. **Journal of Molecular Biology**. New York. v.157, p.105-132. 1982.

LEAH, R.; TOMMERUP, H.; SVENDSENS, I.; MUNDYS, J. Biochemical and molecular characterization of three barley seed proteins with antifungal properties. **The Journal of Biological Chemistry**, Bethesda, v.266, n.3, p.1564-1573,1991.

MARTIN, R.C.; MOK, M.C.; HABBEN, J.E.; MOK, D.W.S. A maize cytokinin gene encoding an O-glucosyltransferase specific to cis-zeatin **PNAS**, v.98, n.1, p.5922–5926, 2001.

NARAYANAN, S.; SATO, T.; WOLFE, M.S. A C-terminal region of signal peptide peptidase defines a functional domain for intramembrane aspartic protease catalysis. **Journal of Biological Chemistry**. Baltimore. v. 282, n. 28, p. 20172-20179. 2007.

NEDEL, J.L.; ASSIS, F.N. de; CARMONA, P.S. A planta de arroz: morfologia e Fisiologia. In: PESKE, S.T.; NEDEL, J.L.; BARROZ, A.C.S.A. **Produção de arroz**. Pelotas: EDITORA E GRÁFICA UNIVERSITÁRIA, 1996. cap. 1, p.11-65.

OKITA, S.Q.T.W.; HWANGS, Y.S.; HNILOG, J.; KIMS, W.T.; ARYAN, A.P.; LARSONLL, R.; KRISHNAN, H.B. Structure and Expression of the Rice Glutelin Multigene Family **The Journal of Biological Chemistry**, Bethesda, v. 264, n. 21, p.12573-12581, 1989.

PROSDOCIMI F.; CERQUEIRA, GUSTAVO C.; BINNECK, E. Bioinformática: Manual do usuário. **Biociência & Desenvolvimento**. n.29. p.12-25, 2001

RIDGEN, D.J.; MELLO, L.V. de Anotação Funcional Computacional de Proteínas. **Biociência & Desenvolvimento**. Brasília. v.25, n.1, p.64-70, 2002.

ROGERS, J.C.; MILLIMAN, C. Isolation and Sequence Analysis of a Barley α -Amylase cDNA Clone. **The Journal of Biological Chemistry**, Bethesda, v.258, n.13. p.8169-8174, 1983.

RUDD, S.; SCHOOF, H.; MAYER, K. PlantMarkers--a database of predicted molecular markers from plants. **Nucleic Acids Research**. Oxford v.33, n.1, p.628-632. 2005.

SHANKLIN, J. & SOMERVILLE C. Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs **Proceedings of the National Academy of Sciences – USA**, Washington, v.88, p.2510-2514, 1991.

SLIGHTOM, J.L.; DRONG, R.F.; KLASSY; R.C.; HOFFMAN, L.M. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. **Nucleic Acids Research**, Oxford, v.13, n.18, p.6483-6498. 1985.

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, Oxford, v.11, n.22, p.4673-4680, 1994.

TIBBOT, B.K.; SKADSEN R.W. Molecular cloning and characterization of a gibberellin-inducible, putative α -glucosidase gene from barley. **Plant Molecular Biology**, Dordrecht, v.30, n.2, p.229-241. 1996.

ZUREK, D.M. & CLOUSE, S.D. Molecular cloning and characterization of a brassinosteroid-regulated gene from elongating soybean (*Glycine max* L.) Epicotyls **Plant Physiology**, Washington, v.104, p.161-170. 1994.

CAPÍTULO II

LOCALIZAÇÃO *IN SILICO* DE MICROSSATÉLITES EM ESPÉCIES FABACEAE DE INTERESSE AGRONÔMICO

Resumo

MENEGHELLO, Geri Eduardo. **Localização *in silico* de microssatélites em espécies fabaceae de interesse agrônomo**. 2007. 61f. Tese (Doutorado) - Programa de Pós-Graduação em Ciência e Tecnologia de Sementes. Universidade Federal de Pelotas, Pelotas.

Marcadores moleculares são biomoléculas que podem ser relacionadas com uma característica genética ou fenotípica, sendo utilizados na construção de mapas de ligação em plantas, seleção assistida por marcadores moleculares, retrocruzamentos, proteção de cultivares e verificação da pureza genética. Recentemente tem merecido destaque marcadores do tipo microssatélites. A identificação dos mesmos é uma tarefa laboriosa e possui custos elevados quando realizada pelos métodos tradicionais. Atualmente com o aumento da disponibilidade pública das seqüências e com o desenvolvimento da bioinformática é possível a localização, identificação e caracterização de microssatélites, através de ferramentas computacionais. O presente trabalho teve por objetivo localizar *in silico* microssatélites em seqüências expressas (ESTs) de espécies Fabaceae de interesse agrônomo, desenhar *primers* e verificar a amplificação via e-PCR – *Eletronic Polymerase Reaction Chain* eletrônico em outras regiões dentro e entre espécies. Seqüências de ESTs foram mineradas junto ao NCBI - *National Center for Biotechnology Information*. Os dados provenientes dessa mineração foram analisados *in silico* através de um *pipeline*, inicialmente buscando identificar microssatélites, na seqüência, com o software Primer3 foram desenhados *primers*. Em seguida, realizou-se e-PCR contra todo o banco de ESTs de todas as espécies selecionadas. Os resultados evidenciaram que a análise *in silico* é eficiente na busca de Microssatélites em espécies pertencentes à família *Fabaceae*, que as espécies da família *Fabaceae* têm predomínio de microssatélites do tipo di e trinucleotídeos nos ESTs depositados no NCBI. Os microssatélites existentes nas seqüências estudadas são altamente conservados dentro de cada espécie.

Palavras Chave: Bioinformática, ESTs, SSRs, primer.

Abstract

MENEGHELLO, Geri Eduardo. ***In silico* location of simple sequence repeat in Fabaceae species of agronomic interest**. 2007. 61f. Tese (Doutorado) - Programa de Pós-Graduação em Ciência e Tecnologia de Sementes. Universidade Federal de Pelotas, Pelotas.

Molecular markers are biomolecules that can be related to a genetic or phenotypic characteristic; they are used in the construction of connection maps in plants, selection assessment by molecular markers, backcrossing, cultivar protection of and verification of the genetic purity. In the last years interest has been placed on markers of the type Single Sequence Repeats (SSRs). Their identification at the laboratory is a laborious and expensive task. At present, the increase in public access to the sequences together with the development of bioinformatics have made possible the location, identification and characterization of molecular markers through computational tools. The present work had for objective the identification of *in silico* SSRs in expressed sequence tags (ESTs) in legume species of agronomic interest, to design primers and to verify the amplification through e-PCR - Eletronic Polymerase Reaction Chain in other areas within and between species. Sequences of ESTs were prospected at NCBI - National Center for Biotechnology Information. The collected data were analyzed *in silico* through a pipeline, initially looking to identify SSRs, in the sequence with the Primer3 software used to design primers. Next, an e-PCR was run across the whole bank of ESTs for all the selected species. The results showed that the analysis *in silico* is efficient in the search of SSRs for legume species belonging to the Fabaceae family and that in these species there is a prevalence of dinucleotide and trinucleotide-type SSRs in the ESTs stocked at NCBI. The existent microsatellites in the studied sequences are highly conserved within each species.

Key-words: Bioinformatic, ESTs, SSRs, primer.

INTRODUÇÃO

Nas últimas décadas houve um grande progresso científico e tecnológico. A utilização de marcadores moleculares é um produto desse avanço e há uma tendência de ampliação de sua utilização, principalmente na identificação da pureza genética de cultivares e na seleção assistida por marcadores. Em paralelo, ocorreu um aumento na disponibilidade de informações biológicas (seqüências de nucleotídeos e aminoácidos) impulsionando o surgimento da bioinformática. Essa nova ciência, em contrapartida, permitiu a ampliação da quantidade de marcadores moleculares, pois possibilita a identificação de novos padrões através de análises *in silico*.

Marcadores moleculares são biomoléculas que podem estar relacionadas com uma característica genética fenotípica. As biomoléculas consideradas marcadores moleculares são partes de proteínas ou do DNA (FERREIRA e GRATAPAGLIA, 1998). YAMANE, (2006) define marcadores moleculares como características do DNA que diferem entre dois ou mais indivíduos e são herdadas geneticamente, diferenciam-se pelas tecnologias utilizadas para revelar variabilidade em nível de DNA e assim, podem variar quanto à habilidade de detectar diferenças entre indivíduos, custo, facilidade de uso, consistência e repetibilidade. Atualmente os marcadores vêm sendo utilizados de forma crescente como ferramenta útil em programas de melhoramento de plantas, pois, representam um recurso experimental promissor nas iniciativas genômicas envolvendo mapeamento genético e na ancoragem de mapas genéticos nos mapas físicos, bem como na seleção assistida por Marcadores Moleculares.

Marcadores moleculares possuem várias aplicações MALONE & ZIMMER et al. (2005) destacam como principais: a construção de mapas de

ligação em plantas, seleção assistida por marcadores moleculares, retrocruzamentos, distância genética e heterose, proteção de cultivares e pureza genética. Os autores enfatizam que a caracterização de uma variedade, linhagem ou híbrido, ao nível de DNA, tem se tornado uma ferramenta importante na proteção dos direitos dos obtentores vegetais. Marcadores moleculares também podem ser utilizados para a quantificação da taxa de polinização cruzada na produção de sementes híbridas. Outra aplicação ligada diretamente a indústria de sementes híbridas é a utilização dos mesmos no estabelecimento de grupos heteróticos, os quais otimizam a escolha de genitores com elevada capacidade de combinação.

Estudando a pureza genética em sementes de milho híbrido por meio de marcadores moleculares do tipo microssatélite, SALGADO et al. (2006) e RAMOS et al. (2006), concluíram que esse tipo de marcadores são eficientes na diferenciação dos híbridos e das linhagens parentais.

Várias técnicas mediadas por PCR - *Polymerase Reaction Chain* ou reação em cadeia da polimerase - são utilizadas em estudos com marcadores moleculares, entre elas as técnicas de RAPD - *Random Amplified Polymorphic DNA* e AFLP - *Amplified Fragment Length Polymorphism* são frequentemente descritas em trabalhos publicados. Entretanto, recentemente, diversos autores têm mostrado que o uso de marcadores microssatélites é uma alternativa eficiente para o mapeamento genético (DIETRICH et al. 1996; DIB et al. 1996). Os marcadores microssatélites são constituídos de seqüências de poucos pares de bases repetidas no genoma, constituindo marcas que podem ser detectadas por meio de hibridização com uma sonda específica ou mesmo pela construção de "primers" específicos para amplificação por meio de PCR (MALONE & ZIMMER, 2005). A eficiência dos microssatélites está diretamente relacionada ao fato de que o acesso a esses locos, caracteristicamente hipervariáveis dentro de uma espécie, tem grande poder de mostrar as taxas de recombinações ocorridas naquele loco.

Atualmente alguns grupos de pesquisa tem concentrado esforços no sequenciamento de RNA, isto é, EST (*Express Sequence Tag*), ou ainda regiões de seqüências expressas, correspondentes diretamente a genes expressos em determinadas situações (temporal e espacial). O uso desses dados tem derivado um novo modelo de marcadores chamados "EST

derivated SSR" (seqüências de regiões repetidas derivadas de sequencias de regiões expressas) WANG et al. 2006; AGGARWAL., et al. 2007.

Utilizando ESTs de cevada (*Hordeum vulgare* L.) THIEL et al. (2003), construíram um banco de dados para a localização e caracterização de microssatélites dessa espécie. Posteriormente, 2.019 seqüências contendo microssatélites foram utilizadas na busca por seqüências homólogas no banco de dados de ESTs de trigo e do arroz. A partir dos dados obtidos pelo *software BlastN* (ALTSCHUL et al. 1990) que efetuou a busca pelas homologias, apenas 311 seqüências foram consideradas promissoras, pois esse protocolo *in silico* demonstrou que esses marcadores estavam presentes nas 3 espécies. Posteriormente esses marcadores foram transpostos para outras espécies, sendo que, para a especie *Hordeum bulbosum* L. (cinco acessos), 80% dos *primers* mostraram sucesso nas amplificações, em centeio (*Secale cereale* L. – 2 acessos) e trigo (*Triticum aestivum* L. cult. Chinese Spring) ambos resultados foram de 60% e em arroz (*Oryza sativa* L.) 40% dos *primers* mostraram amplificações.

Segundo BRONDANI et al. (2002), foram desenvolvidos 400 marcadores microssatélites para *Eucalyptus* e a seleção de mais 1.000 locos microssatélites é prevista para a saturação do mapa. O autor afirma ainda que, o objetivo da seleção desses marcadores é o mapeamento de grande parte dos QTLs - *Quantitative Trait Loci* para esta espécie e o ancoramento do mapa genético ao mapa físico. BRONDANI et al. 1998, cita que numa etapa inicial de 180 microssatélites, 68% desses apresentaram polimorfismo.

RODER et al. (1998), realizou a seleção de 279 microssatélites, analisando uma população em geração F₈, conduzida pelo método SSD (*single seed descent*) oriunda do cruzamento entre a variedade sintética W7984 (*Aegilops tauschii* L. (DD) x ALTAR 84 (*Triticum durum* L. (AABB)) e OPATA 85 (CYMMYT), concluindo que 80% dos *primers* construídos para os 279 locos microssatélites mostraram polimorfismo entre os genótipos OPATA 85, W7984 e os descendentes F₈.

GUYOMARC et al. (2002), usando bibliotecas enriquecidas de DNA genômico extraído de *Aegilops tauchii* L., sequenciaram 2.500 ESTs. Desta biblioteca foram extraídos 566 microssatélites dos quais foram desenhados *primers* para 378 locos. Posteriormente, 293 foram testados, e, destes 270

mostraram ampliações desejadas em trigo (*Triticum aestivum* L.), sendo que, 132 *primers* não mostraram polimorfismo e dos 138 remanescentes 84 foram polimorficos entre as variedade de trigo Courtot e Chinese Spring.

A descoberta de marcadores microssatélites em ESTs, com o auxilio de ferramentas computacionais está crescendo. Entretanto, a maioria dos esforços tem se concentrado em espécies monocotiledôneas. KUMPATLA & MUKHOPADHYAY (2005) afirmaram que também há potencial para a descoberta de microssatélites em espécies dicotiledôneas.

A identificação de microssatélites é uma tarefa laboriosa e possui custos elevados quando realizada em laboratório. Atualmente, com o aumento da disponibilidade pública das seqüências do genoma do arroz e de outras espécies é possível o desenvolvimento de trabalhos de análise ou inferências sobre qualquer tipo de informação existente nesses genomas, incluindo a localização, identificação e caracterização de marcadores moleculares, em especial para as espécies que possuem seu genoma completamente seqüenciado ou com grande parte do seu genoma mapeado em mapas genéticos, ancorado com ESTs e QTLs.

TEMNYKH et al. (2000), utilizando dados publicados pelo IRGSP (*International Rice Genome Sequencing Project*) realizou, no Centro de Bioinformática para Agricultura da Universidade de Cornell, um experimento computacional para análise de freqüência, variação nos tamanhos e potencial como marcadores moleculares do genoma do arroz, num total de 57.8 Mb (megabases) de DNA da variedade Nipponbare. O resultado do referido experimento mostrou a existência de 1.178 microssatélites referentes a classe 1 (com mais de 20 pares de base) e 12.811 microssatélites pertencentes a classe 2 (com mais de 12 e igual ou menor que 20 pares de base), sendo 8,4% e 91,6%, respectivamente, os percentuais de cada classe sobre o número total encontrado.

A busca *in silico* de microssatélites através de ferramentas computacionais foi realizada com sucesso por MAIA et al, (2005) e MAIA (2007) em arroz (*Oryza sativa*). Estudos confirmam marcadores encontrados nessa espécie, mostram sucesso na amplificação dos mesmos *locos* na espécie africana *Oryza glaberrima* (PANAUD et al., 1996; LORIEUX et al., 2000) e outras cinco espécies (WU & TANKSLEY, 1993; HARRINGTON,

2000), e que esse sucesso aponta para o alto potencial de que estes marcadores possam ser usados na integração de mapas genéticos com o mapa físico, permitindo a geração de marcadores para a saturação das regiões de interesse para cada grupo de pesquisa.

Em citrus também foi logrado êxito na localização, *in silico*, de microssatélites (CHEN et al., 2006). A equipe do projeto CitEST (*Integration of Genetic Breeding, Functional and Comparative Genomics of Citrus*) publicou dados referentes à localização de microssatélites em 100.000 EST de diferentes bibliotecas de *Citrus* com a utilização de ferramentas computacionais (PALMIERI et al., 2005).

Entretanto, é importante reforçar que a grande limitação do uso em larga escala de marcadores microssatélites é a obtenção dos *primers* que serão usados na reação de PCR para amplificar alelos em cada loco (CHAVES, 2006).

O presente trabalho teve por objetivo localizar *in silico* microssatélites em ESTs de espécies fabaceae de interesse agrônômico, desenhar *primers* e verificar a amplificação via e-PCR em outras regiões, dentro e entre espécies.

MATERIAL E MÉTODOS

Seqüências de ESTs (*Expressed Sequence Tags*) das principais espécies pertencentes a família Fabaceae consideradas de interesse agrônômico, incluindo sojas (*Glycine max*, *G. soja*, *G. clandestina*), alfafas (*Medicago sativa* e *Medicago truncatula*), amendoim (*Arachis hypogea*), ervilha (*Pisum sativum*), guandu (*Cajanus cajan*), fava (*Vicia faba*), feijão miúdo (*Vigna unguiculata*), grão de bico (*Cicer arietinum*), feijão (*Phaseolus vulgaris*) e lótus (*Lotus japonicus*), foram mineradas junto ao NCBI - *National Center for Biotechnology Information* (www.ncbi.nlm.nih.gov). As seqüências (*contigs*) foram salvas no formato *fasta* (padrão de arquivos contendo seqüências contíguas de nucleotídeos.).

Os dados provenientes dessa mineração foram analisados *in silico* através do *pipeline*, composto por diversos algoritmos escritos em linguagem PERL - *Practical Extraction and Report Language* rodando na plataforma LINUX (Figura 1).

Inicialmente, através do script *file2.pl* as seqüências foram renomeadas passando a serem denominadas em cada espécie de *contig 1*, *contig 2*, *contig n*. Esta etapa fez-se necessária para tornar funcional os demais passos do programa e para reduzir o tamanho dos arquivos, tornando, com isso, o *pipeline* mais ágil e veloz.

Na seqüência, com os nomes dos *contigs* ajustados utilizou-se uma adaptação do algoritmo MISA (não publicado) escrito em *Perl* desenvolvido por Thomas Thiel do *Plant Genome Resource Center* - Alemanha, para a localização de microssatélites nos *contigs* dentro de cada uma das espécies analisadas. Considerou-se microssatélite quando houve uma repetição mínima de dez vezes para dinucleotídeos, sete para trinucleotídeos e cinco vezes para tetranucleotídeos, pentanucleotídeos e hexanucleotídeos.

Com o software Primer3 (ROZEN & SKALETSKY, 2000) foram desenhados *primers*, utilizando-se como principais características tamanhos mínimos, máximos e ótimos de 18, 35 e 20 pb, respectivamente; Tm mínima, máxima e ótima 58, 60 e 62°C; conteúdo de CG mínimo e ótimo de 20 e 50%, respectivamente. Por se tratar de uma análise *in silico* foi necessário que todos os *primers* tivessem as mesmas características uma vez que seriam usados massivamente nas análises posteriores. Gerou-se um arquivo com as seqüências *forwad* e *reverse* dos *primers* em cada espécie, nas seqüências em que foi possível desenhar *primers* nos parâmetros pré-determinados.

```
perl file2.pl sequences_especie_alvo.fasta especie_alvo.fasta;
echo "alterando nome dos contigs...OK";

perl misa.pl especie_alvo.fasta;
echo "MISA ...OK";

perl p3_in.pl especie_alvo.fasta.misa especie_alvo.fasta
especie_alvo.fasta.p3in;
echo "P3IN ...OK";

./primer3_core < especie_alvo.fasta.p3in> especie_alvo.fasta.p3out;
echo "PRIMER3 ...OK";

perl p3_out.pl especie_alvo.fasta.p3out especie_alvo.fasta.misa;
echo "P3OUT ...OK";

cut -f1,8,11,14 -s especie_alvo.fasta.results> especie_alvo_primers.txt;
echo "CUT...OK";

./e-PCR -w9 -f1 -m10 especie_alvo_primers.txt D=100-400 especie_alvo.fasta N=1
G=1 T=3 O=saida_especie_alvo_x_especie_alvo.txt;
echo "e-PCR ...OK";

perl contgs.pl especie_alvo_x_especie_alvo.txt arquivo_saida_contig_renomeado.txt
arquivo_contigs_homologos.txt arquivo_contgs_primer.txt
echo "ronomeando contigs.....OK. PIPELINE CONCLUÍDO";
```

Figura 1: Pipeline utilizado para a identificação *in silico* de microssatélites em leguminosas de interesse agrônômico.

Com os *primers* desenhados para cada uma das espécies, utilizou-se o programa e-PCR (SCHULER et al. 1997) para identificar possíveis ampliações em regiões distintas daquelas em que o microssatélite foi localizado e o *primer* desenhado. O PCR eletrônico foi realizado com todos os

primers de uma determinada espécie contra todas as seqüências da mesma espécie e contra cada uma das demais espécies analisadas.

Por fim utilizando um algoritmo denominado CONTIGS.PL, os contigs voltaram a ser nomeados com a nomenclatura que estavam depositados no NCBI.

RESULTADOS E DISCUSSÃO

Na tabela 1 estão relacionados o número de ESTs presentes nas espécies selecionadas, o número de microssatélites identificados e a quantidade de primers desenhados. Destacam-se as espécies *G. max*, *M. truncatula* e *L. japonicus* por possuírem uma grande quantidade de ESTs depositadas (mais de 150 mil em cada uma delas). Essa maior quantidade de informação disponível se deve ao fato da importância econômica que a soja possui e as outras duas espécies podem ser consideradas modelos genômicos para a família Fabaceae (COOK, 1999; YOUNG et al., 2005; BATAILLON & RONFORT, 2006; RONFORT et al., 2006).

A percentagem de microssatélites localizados em relação ao número de ESTs depositadas variou de 0,57 % em *V. fava* até 4,73 em *C. arietinum*, porém na grande maioria das espécies estudadas foram encontrados microssatélites numa faixa de 1 a 2% das ESTs. Estes resultados são semelhantes aos encontrados por KANTETY et al. (2002), que observaram a ocorrência de microssatélites em 1,5% de ESTs de milho e 4,7% para as seqüências expressas de arroz. YI et al. (2003), analisando *in silico* 10.232 ESTs de pimenta encontraram 1.201 microssatélites, ou seja, em mais de 11% das seqüências. Já PONCET et al. (2006) identificaram 6,8% de microssatélites em 5.534 ESTs do gênero *Coffea* estudados. No algodão (*Gossypium hirsutum*), HAN et al. (2006), observaram que a taxa de ocorrência de microssatélites nas 13.505 ESTs analisadas ficou em 7,15%.

As seqüências que apresentaram microssatélites foram submetidas ao programa PRIMER3 para serem desenhados *primers*. Chama a atenção espécie *G. clandestina* por ter sido a única em que não foi possível o desenho de *primers* nas condições estabelecidas em nenhuma das seqüências em que

foi encontrado microssatélite. As demais espécies apresentaram desempenho bastante similar entre si, pois a taxa de *primers* gerados foi acima de 60%, chegando a 100% em *V. faba*. É possível que, alterando os padrões dos *primers* essa taxa pudesse ser aumentada em algumas espécies, porém torna-se inviável para um *pipeline* que analisa um número grande de seqüências, trabalhar com padrões diferenciados.

Tabela 1: Número total de seqüências expressas analisadas (ESTs), microssatélites (SSRs) localizados, quantidade de *primers* desenhados, percentagens de SSRs existentes nas ESTs estudadas e percentagens de *primers* nos SSRs encontrados para cada espécie analisada nesse estudo.

Espécie*	ESTs	SSR	Primers	% SSR	% de primer no SSR
Mas	7.681	123	104	1,60	84,55
Ahy	9.732	251	200	2,58	79,68
Psa	6.022	81	70	1,35	86,42
Vfa	704	4	4	0,57	100,00
Pvu	25.717	421	301	1,64	71,50
Vun	562	8	5	1,42	62,50
Gcl	931	11	0	1,18	0,00
Gso	18.511	275	179	1,49	65,09
Car	1.290	61	56	4,73	91,80
Cca	506	11	8	2,17	72,73
Lja	150.813	2.818	1.984	1,87	70,40
Mtr	225.129	4.309	2.652	1,91	61,55
Gma	358.905	5.936	3.814	1,65	64,25

* Mas – *Medicago sativa*; Ahy - *Arachis hypogea*; Psa - *Pisum sativum*; Vfa – *Vicia faba*; Pvu – *Phaseolus vulgaris*; Vun - *Vigna unguiculata*; Gcl - *Glycine Clandestina*; Gso - *Glycine soja*; Car - *Cicer arietinum*; Cca - *Cajanus cajan*; Lja - *Lotus japonicus*; Mtr – *Medicago truncatua*; Gma – *Glycine max*.

Na tabela 2 estão apresentados o número de amplificações encontradas quando as seqüências foram submetidas ao PCR eletrônico utilizando os *primers* desenhados. Em todas as espécies ocorreram amplificações em outros ESTs que não somente naquela em que foi localizado o microssatélite e desenhado o *primer* em si. Isso provavelmente aconteceu porque houve mais de um depósito da mesma ESTs que por não serem exatamente idênticas o

algoritmo não conseguiu identificar microsatélite porém na região de ancoragem do *primer* há uma similaridade capaz de gerar amplificação no PCR eletrônico. Outra possibilidade é a existência de mesmo microstélite em outra região do genoma.

Entre espécies diferentes houve um pequeno número de amplificações, e estas quando ocorreram foram predominantemente entre espécies do mesmo gênero, como é o caso das amplificações observadas em *Medicago truncatula* (25) quando as seqüências dessa espécie foram submetidas ao e-PCR com os *primers* obtidos nos microsatélites de *M. sativa*. Por outro lado, ocorreram 57 amplificações nas seqüências de *M. sativa* quando se tentou amplifica-la usando *primers* gerados nas seqüências que apresentaram microsatélites de *M. truncatula*. O mesmo foi observado em soja, uma vez que quando utilizaram-se *primers* de *G. max* houve 4.173 amplificações em *G. soja* e 75 em *G. clandestina*. E também foram encontradas 2.494 amplificações nas seqüências de *G. soja* quando foram usados *primers* de *G. max*.

As únicas amplificações entre espécies de gêneros diferentes foram observadas em *L. japonicus* que amplificou quatro seqüências com *primers* de *P. vulgaris*, e quando foram utilizados *primers* gerados a partir dos microsatélites de *G. max* observaram-se 10 amplificações em *P. vulgaris*, uma em *L. japonicus* e outra em *M. truncatula*. Esse comportamento evidencia o caráter de alta especificidade que os microsatélites apresentam dentro de cada espécie. Optou-se por não apresentar os alinhamentos das amplificações encontradas razão da grande quantidade gerada.

Em estudos que visam identificar a presença de microsatélites é fundamental saber o número de bases repetidas. Na tabela 3 são apresentados o número de microsatélites encontrados estratificados de acordo com o número de nucleotídeos repetidos (di, tri, tetra, penta e hexanucleotídeos) em cada espécie. A grande concentração de microsatélites é formada por repetição de dois ou três nucleotídeos, a espécie que apresentou menor quantidade desse tipo de microsatélites foi *L. japonicus* onde foram encontrados 77,40%, por outro lado, em *V. faba*, *G. clandestina* e *C. arietinum*, 100% dos microsatélites observados possuíam repetições de no máximo 3 pares de bases.

Tabela 2: Número de seqüências expressas (ESTs), microssatélites (SSRs) localizados e *primers* desenhados nas espécies estudadas e respectivo número de ampliações encontradas dentro da própria espécie e nas demais que constaram no estudo.

Espécie*	Mas	Ahy	Psa	Vfa	Pvu	Vun	Gcl	Gso	Car	Cca	Lja	Mtr	Gma
ESTs→	7.681	9.732	6.022	7.074	25.717	562	931	18.511	1.290	506	150.813	225.129	358.905
SSRs→	122	251	81	4	421	8	11	275	61	11	2.818	4.309	5.936
Primer↓													
Msa	104	123	0	0	0	0	0	0	0	0	0	25	0
Ahy	200	0	5.459	0	0	0	0	0	0	0	0	0	0
Psa	70	0	0	208	2	0	0	0	0	0	0	0	0
Vfa	4	0	0	0	4	0	0	0	0	0	0	0	0
Pvu	301	0	0	0	0	456	0	0	0	0	4	0	6
Vun	5	0	0	0	0	0	11	0	0	0	0	0	0
Gcl	0	0	0	0	0	0	0	0	0	0	0	0	0
Gso	179	0	0	0	0	0	0	469	0	0	0	0	2.494
Car	56	0	0	0	0	0	0	0	294	0	0	0	0
Cca	8	0	0	0	0	0	0	0	0	10	0	0	0
Lja	1.984	0	0	0	0	0	0	0	0	0	27.112	0	0
Mtr	2.652	57	0	0	0	0	0	0	0	0	0	37.897	8
Gma	3.814	0	0	0	0	10	0	75	4.173	0	0	1	84.505

* Msa – *Medicago sativa*; Ahy - *Arachis hypogea*; Psa - *Pisum sativum*; Vfa – *Vicia faba*; Pvu – *Phaseolus vulgaris*; Vun - *Vigna unguiculata*; Gcl - *Glycine Clandestina*; Gso - *Glycine soja*; Car - *Cicer arietinum*; Cca - *Cajanus cajan*; Lja - *Lotus japonicus*; Mtr – *Medicago truncatua*; Gma – *Glycine max*.

Tabela 3: Frequência das diferentes combinações de nucleotídeos repetidos encontradas nas ESTs das espécies estudadas.

Espécie*	Repetições	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Mais	Total	%
Msa	Dinucleotídeos						9	4	7	5	4	2	5	3	0	1	2	5	47	38,52
	Trinucleotídeos			25	10	9	2	1	0	1	0	0	0	0	0	0	0	0	48	39,34
	Tetranucleotídeos	12	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	15	12,30
	Pentanucleotídeos	7	3		0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	8,20
	Hexanucleotídeos	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	1,64
Ahy	Dinucleotídeos						17	13	4	10	4	5	4	3	3	1	1	7	72	28,69
	Trinucleotídeos			100	21	10	5	0	0	0	1	0	2	3	0	0	0	2	144	57,37
	Tetranucleotídeos	12	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	6,37
	Pentanucleotídeos	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	5,18
	Hexanucleotídeos	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	6	2,39
Psa	Dinucleotídeos						2	4	3	1	2	0	1	2	2	0	0	8	25	30,86
	Trinucleotídeos			28	8	2	0	0	0	2	1	5	0	1	0	0	0	1	48	59,26
	Tetranucleotídeos	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	6,17
	Pentanucleotídeos	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1,23
	Hexanucleotídeos	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2,47
Vfa	Dinucleotídeos						1	0	0	0	0	0	0	1	0	0	0	0	2	50,00
	Trinucleotídeos			0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	2	50,00
	Tetranucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Pentanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Hexanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
Pvu	Dinucleotídeos						41	29	15	16	12	17	12	12	8	20	8	88	278	66,03
	Trinucleotídeos			61	21	8	9	0	0	0	0	0	0	0	0	0	0	0	99	23,52
	Tetranucleotídeos	23	2	5	1	0	0	0	1	0	0	0	0	0	0	0	0	0	32	7,60
	Pentanucleotídeos	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0,95
	Hexanucleotídeos	4	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	1,90
Vun	Dinucleotídeos						3	0	0	0	0	0	0	0	0	0	0	0	3	37,50
	Trinucleotídeos			1	0	4	0	0	0	0	0	0	0	0	0	0	0	0	5	62,50
	Tetranucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Pentanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Hexanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
Gcl	Dinucleotídeos						2	0	2	0	1	1	0	0	0	1	0	1	8	80,00
	Trinucleotídeos			0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	20,00
	Tetranucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Pentanucleotídeos	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Hexanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00

Continua....

...continuação

Espécie	Repetições	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Mais	Total	%
Gso	Dinucleotídeos						39	14	27	10	0	2	7	3	6	7	5	37	157	57,09
	Trinucleotídeos			47	31	7	6	4	1	2	1	1	0	1	0	0	1	0	102	37,09
	Tetranucleotídeos	7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	3,27
	Pentanucleotídeos	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0,36
	Hexanucleotídeos	5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	2,18
Car	Dinucleotídeos						3	3	4	5	3	3	5	14	2	6	7	1	56	90,32
	Trinucleotídeos			5	0	0	1	0	0	0	0	0	0	0	0	0	0	0	6	9,68
	Tetranucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Pentanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Hexanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
Cca	Dinucleotídeos						2	0	1	0	0	1	0	1	0	2	1	2	10	90,91
	Trinucleotídeos			0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	9,09
	Tetranucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Pentanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
	Hexanucleotídeos	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,00
Lja	Dinucleotídeos						273	202	157	53	96	74	32	26	16	13	11	71	1024	33,77
	Trinucleotídeos			722	349	106	47	31	25	13	8	2	2	11	0	2	2	3	1323	43,63
	Tetranucleotídeos	177	22	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	205	6,76
	Pentanucleotídeos	227	38	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	271	8,94
	Hexanucleotídeos	165	29	11	2	2	0	0	0	0	0	0	0	0	0	0	0	0	209	6,89
Mtr	Dinucleotídeos						401	239	153	88	71	102	76	70	98	67	54	574	1993	46,25
	Trinucleotídeos			787	388	152	54	54	36	28	5	3	5	3	0	0	0	0	1516	35,18
	Tetranucleotídeos	327	79	19	5	101	0	0	0	0	0	0	0	0	0	0	0	0	531	12,32
	Pentanucleotídeos	153	9	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	177	4,11
	Hexanucleotídeos	74	14	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	92	2,14
Gma	Dinucleotídeos						681	435	195	162	141	110	99	117	94	299	94	902	3329	56,08
	Trinucleotídeos			1260	383	198	91	25	23	19	14	5	5	5	4	1	2	9	2044	34,43
	Tetranucleotídeos	227	88	29	3	2	0	0	0	1	0	1	0	0	0	1	0	0	352	5,93
	Pentanucleotídeos	74	1	1	3	0	1	0	0	0	0	0	0	0	0	0	0	0	80	1,35
	Hexanucleotídeos	109	15	6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	131	2,21

* Mas – *Medicago sativa*; Ahy - *Arachis hypogea*; Psa - *Pisum sativum*; Vfa – *Vicia faba*; Pvu – *Phaseolus vulgaris*; Vun - *Vigna unguiculata*; Gcl - *Glycine Clandestina*; Gso - *Glycine soja*; Car - *Cicer arietinum*; Cca - *Cajanus cajan*; Lja - *Lotus japonicus*; Mtr – *Medicago truncatua*; Gma – *Glycine max*

Destaca-se a espécie *C. arietinum* e *G. clandestina* que apresentaram respectivamente 90,01 e 80,00% de microssatélites formado por repetições de dinucleotídeos. Por outro lado, em *A. hypogea*, *P. sativum* e *V. unguiculata* foram encontrados 59,37, 59,26 e 62,50% de microssatélites formados por trinucleotídeos. Padrão similar é apresentado pelo fungo *Neurospora crassa*, que segundo estudos desenvolvidos por SHANKER et al. (2006) tem 88,8% de microssatélites do tipo trinucleotídeos. *M. sativa* e *M. truncatula* destacam-se por terem apresentado 12,30 e 12,32% dos microssatélites formados por tetranucleotídeos. Essa característica foi observada apenas nessas duas espécies, pertencentes ao mesmo gênero, sugerindo ser esta uma característica do gênero. Grandes taxas de ocorrências de tetranucleotídeos também foram observadas por FOLTA et al. (2005) em morango (*Fragaria x ananassa*), com 27,9% de microssatélites formados por repetições de 4 nucleotídeos.

O predomínio de microssatélites formado por di e trinucleotídeos também foi observado em café (AGGARWAL., et al. 2007). Em algodão, HAN et al. (2006) observaram que 59,1% dos microssatélites em ESTs eram do tipo trinucleotídeos, seguido por dinucleotídeos que representavam 30%, resultados semelhantes foram obtidos por YI et al., (2006) que encontraram 66% dos microssatélites em pimenta formado pela repetição de três nucleotídeos. Na espécie *Citrus unigenes* JIANG et al. (2006), encontraram predomínio de microssatélites formados pela repetição de um único nucleotídeo, seguido pelos trinucleotídeos.

A identificação de microssatélites *in silico* possibilita que uma grande quantidade de dados seja analisada em pouco tempo, inclusive testando “n” microssatélites em outras espécies, identificando regiões que permanecem conservadas em diferentes espécies ao longo da evolução. Estudos dessa natureza apresentam uma grande potencialidade no estudo da sintonia em regiões hipervariáveis entre diferentes espécies. Sem o uso de ferramentas computacionais, tais estudos seriam inviabilizados, tanto pelo tempo que seria demandado, quanto pelo custo que apresentam.

CONCLUSÕES

A análise *in silico* é eficiente na localização de microssatélites em espécies pertencentes a família Fabaceae;

As espécies da família Fabaceae tem predomínio de microssatélites do tipo di e trinucleotídeos nos ESTs depositados no NCBI;

Os microssatélites existentes nas seqüências estudadas são altamente conservados dentro de cada espécie.

REFERÊNCIAS BIBLIOGRÁFICAS

AGGARWAL., R.K.; HENDRE, P.S.; VARSHNEY, R.K.; BHAT, P.R.; KRISHNAKUMAR, V.; SINGH, L. Identification characterization and utilization of EST derived genic microsatellite markers for genome analyses of coffee and related species. **Theoretical Applied Genetics**, Berlin v.114, n.2, p.359-372. 2007

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of Molecular Biology**. London, v.215, p.403-410, out. 1990.

BATAILLON, T.; RONFORT J. Evolutionary and Ecological Genetics of *Medicago truncatula*. In.: **Medicago truncatula Handbook**. Disponível na Internet: <<http://www.noble.org/medicagohandbook/pdf/EvolutionaryEcologicalGenetics.pdf>> Acesso em 11 nov. 2006.

BRONDANI, R.; BRONDANI, C.; GRATTAPAGLIA, D. Towards a genus-wide reference linkage map for *Eucalyptus* based exclusively on highly informative microsatellite markers. **Molecular Genetics and Genomics**. Berlin, v.267, p.338–347, 2002.

BRONDANI, R.P.V.; BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical Applied Genetics**, Berlin v.97, p.816-827, 1998.

CHAVES, R.A. **Marcadores moleculares**. Disponível em <<http://www.ufv.br/dbg/trab2002/GMOL/GMOL008.htm>>. Acesso em 28 nov. 2006.

COOK, D.R. *Medicago truncatula*--a model in the making! **Current Opinion in Plant Biology**. London, v. 2, n.4, p.301-304. 1999.

DIB, C.; FAURE, S.; FIZAMES, C.; SAMSON, D.; DROUOT, N.; VIGNAL, A.; MILLASSEAU, P.; MARC, S.; HAZAN, J.; SEBOUN, E.; LATHROP, M.; GYAPAY, G.; MORISSETE, J.; WEISSENBACH, J. A comprehensive genetic map of the human genome based on 5.264 microsatellites. **Nature**, London, v.380, p.152-154, 1996.

DIETRICH, W.F.; MILLER, J.; STEEN, R.; MERCHANT, M.A.; DAMRON-BOLES, D.; HUSAIN, Z.; DREDGE, R.; DALY, M.J.; INGALLS, K.A.; O'CONNOR, T.J.; EVANS, C.A.; DEANGELIS, M.M.; LEVINSON, D.M.; KRUGLYAK, L.; GOODMAN, N.; COPELAND, N.G.; JENKINS, N.A.; HAWKINS, T.L.; STEIN, L.; PAGE, D.C.; LANDER, E.S. A comprehensive genetic map of the mouse genome. **Nature**, London, v. 380, p. 149-152. 1996.

FOLTA, K.M.; STATON, M.; STEWART, P.J.; JUNG, S.; BIES, D.H.; JESDURAI, C.; MAIN D. Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (*Fragaria × ananassa*) **BMC Plant Biology**, Stuttgart, v.5, n. 12. p. 1-11. 2005.

FERREIRA, M.E. e D. GRATTAPAGLIA. 1998. **Introdução ao uso de marcadores moleculares em análise genética**. 3a ed. Brasília: EMBRAPA-CENARGEN. 220p.

GUYOMARC, H.; SOURDILLE, P.; CHARMET, G.; EDWARDS, K.J.; BERNARD, M.; Characterisation of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. **Theoretical and Applied Genetics**, Berlin, v. 104, p.1164–1172, 2002.

HAN, Z.; WANG, C.; SONG, X.; GUO, W.; GOU, J.; LI, C.; CHEN, X.; ZHANG, T. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. **Theoretical Applied Genetics**, Berlin v. 112, n.3, p. 430-439. 2006

HARRINGTON, S. A survey of genetic diversity of eight AA genome species of *Oryza* using microsatellite markers. MS thesis, Cornell University, Ithaca, NY. **Genome Research**. Cold Spring Harbor, v.7, p.471–482, 2000.

JIANG, D.; ZHONG, G.Y.; HONG, Q.B. Analysis of microsatellites in citrus unigenes. **Acta genetica Sinica**, Peiping, v. 33, n. 4, p. 345-353. 2006

KANTETY, R.V.; LA ROSA, M.; MATTHEWS, D.E.; SORRELLS, M.E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. **Plant Mol Biology**, Dordrecht v.48, n. 5-6, p. 501-510. 2002

KUMPATLA, S.P.; MUKHOPADHYAY, S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. **Genome**, Ottawa, v. 48, n.6, p. 985-998. 2005.

LORIEUX, M.; NDJIONDJOP, M.N., GHESQUIERE, A., A first interspecific *Oryza sativa* x *Oryza glaberrima* microsatellite-based genetic linkage map. **Theoretical and Applied Genetics**, Berlin, v.100, p.593-601, 2000.

MAIA, L.C.; OLIVEIRA, A.C.; CARVALHO, F.I.F.; SILVA, J.A.G.; KOPP, M.M. Localização in silico de microssatélites no cromossomo 1 do arroz (*Oryza sativa* L.). Congresso Brasileiro de Arroz Irrigado, 4., Reunião da Cultura do Arroz Irrigado, 26., 2005, Santa Maria. **Anais ...**, 2005. v. 1. p. 184-186.

MALONE, G.; ZIMMER, P.D.; Marcadores Moleculares. In: ZIMMER,P.D.; OLIVEIRA,A.C.; MALONE,G., **Ferramentas da Biotecnologia Vegetal**. Pelotas, editora Gráfica Universitária UFPEL, 157p. 2005.

PALMIERI, D.; ASTUÁ-MONGE, G.; MACHADO, M. A.; BASILIO, A. C.; LIMA, V.; BACOCINA, G.; RONCOLETTA, J. G. T.X. **In silico identification and characterization of SSRs from citrus ESTs**; XIII Plant and Animal Genome: XIII Plant and Animal Genome, SanDiego, p. 275. Disponível em: <http://www.intl-pag.org/13/abstracts/PAG13_P825.html>. Acesso em 20 dez. 2005.

PANAUD, O.; CHEN, X.;MCCOUCH, S.R.; Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). **Molecular and General Genetics**. New York, v. 252, n.5, p.597-607, 1996.

PONCET, V.; RONDEAU, M.; TRANCHANT, C.; CAYREL, A.; HAMON, S; DE KOCHKO, A.; HAMON, P. SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. **Molecular genetics and genomics**, Berlin, v.276, n.5, p.436-449. 2006.

RAMOS, N.P.; BRUNELLI, K.R.; CAMARGO, L.E.A.; MARCOS FILHO, J. Sensibilidade dos microssatélites para determinar a pureza varietal em sementes de milho. **Revista Brasileira de Sementes**, Pelotas, v.28, n.1, p.99-105, 2006

RONFORT, J.; BATAILLON, T.; SANTONI, S.; DELALANDE, M.; DAVID, J.L.; PROSPERI, J.M. Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. **BMC Plant Biology**, London, v. 13, n.6, p. 28-40. 2006.

ROZEN, S.; SKALETSKY, H. Primer3 on the WWW for general users and for biologist programmers. **Methods in molecular biology**, Clifton, v.132, p.365–386. 2000.

SALGADO, K.C.P.C.; VIEIRA, M.G.G.C.; PINHO, E.V.R.; GUIMARÃES, C.T.; VON PINHO, R. G.; SOUSA, L.V. Genetic purity certificate in seeds of hybrid maize using molecular markers **Revista Brasileira de Sementes**, Pelotas, v. 28, n.1, p.169-175, 2006.

SHANKER, A.; SINGH, A.; SHARMA V. *In silico* mining in expressed sequences of *Neurospora crassa* for identification and abundance of microsatellites. **Microbiological Research**. Jena, v.161, p.1-6. 2006.

SCHULER, G.D. Sequence mapping by electronic PCR. **Genome Research**, Cold Spring Harbor, v.7, n.5, p. 541-550. 1997.

TEMNYKH, S.; DECLERCK, G.; LUKASHOVA, A.; LIPOVICH, L.; CARTINHO, S.; MCCOUCH, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. **Genome Research**. Cold Spring Harbor, v.11, p.1441-1452, 2000.

THIEL, T.; MICHALEK, W.; VARSHNEY, W.; GRANER, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, Berlin, v.106, n.3, p.411-422, fev. 2003.

WANG, M.L.; MOSJIDIS, J.A.; MORRIS, J.B.; DEAN, R.E.; JENKINS, T.M.; PEDERSON, G.A. Genetic diversity of *Crotalaria* germplasm assessed through phylogenetic analysis of EST SSR markers. **Genome**. Ottawa, v.49, n.6, p.707-715. 2006

WU, K.S.; TANKSLEY, S.D. Abundance, polymorphism and genetic mapping of microsatellites in rice. **Molecular Genetics and Genomics**. Berlin, v.241, p.225-235, 1993.

YAMANE, R.S.. **Marcadores Moleculares**. Disponível em <<http://www.ufv.br/dbg/trab2002/GMOL/GMOL008.htm>>. Acesso em 28 nov. 2006.

YI, G.; LEE, J.M.; LEE, S.; CHOI, D.; KIM, B.D. Exploitation of pepper EST-SSRs and an SSR-based linkage map. **Theoretical and Applied Genetics**, Berlin, v.106, n.3, p.411-422. 2003

YOUNG, N.D.; CANNON, S.B.; SATO, S.; KIM, D.; COOK, D.R.; TOWN, C.D.; ROE, B.A.; TABATA, S. Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. **Plant Physiology**, Washington, v.137, n. 4, p.1174-1181. 2005.

ANEXO I

Relação de Proteínas

Local	Espécie	AC	Proteína
Aleurona	<i>Avena fátua</i>	P69309	Ubiquitin
Aleurona	<i>Cyamopsis tetragonoloba</i>	P14749	Alpha-galactosidase precursor (EC 3.2.1.22) (Melibiase)
Aleurona	<i>Hordeum vulgare</i>	Q43763	Alpha-glucosidase precursor (EC 3.2.1.20) (Maltase)
Aleurona	<i>Hordeum vulgare</i>	P00693	Alpha-amylase type A isozyme precursor (EC 3.2.1.1)
Aleurona	<i>Hordeum vulgare</i>	P07519	Serine carboxypeptidase 1 precursor (EC 3.4.16.5)
Aleurona	<i>Hordeum vulgare</i>	P21529	Serine carboxypeptidase 3 precursor (EC 3.4.16.5)
Aleurona	<i>Hordeum vulgare</i>	P25249	Cysteine proteinase EP-B 1 precursor (EC 3.4.22.-)
Aleurona	<i>Hordeum vulgare</i>	P25250	Cysteine proteinase EP-B 2 precursor (EC 3.4.22.-)
Aleurona	<i>Hordeum vulgare</i>	Q42831	Non-symbiotic hemoglobin (HORvu GLB1)
Aleurona	<i>Hordeum vulgare</i>	Q07764	Protein HVA22
Aleurona	<i>Hordeum vulgare</i>	P14928	ABA-inducible protein PHV A1
Aleurona	<i>Hordeum vulgare</i>	P52572	Probable peroxiredoxin (EC 1.11.1.15) (Thioredoxin peroxidase)
Aleurona	<i>Oryza sativa</i>	P93417	Transcription factor GAMYB (OsGAMyB)
Aleurona	<i>Oryza sativa</i>	P55142	Glutaredoxin
Aleurona	<i>Zea mays</i>	P55005	Beta-amylase (EC 3.2.1.2) (1,4-alpha-D-glucan maltohydrolase)
Coleóptilo	<i>Avena sativa</i>	P40412	T-complex protein 1 subunit epsilon (TCP-1-epsilon)
Coleóptilo	<i>Avena sativa</i>	P54411	T-complex protein 1 subunit epsilon (TCP-1-epsilon)
Coleóptilo	<i>Oryza sativa</i>	P53683	Calcium-dependent protein kinase, isoform 2 (EC 2.7.11.1)
Coleóptilo	<i>Oryza sativa</i>	P53684	Calcium-dependent protein kinase, isoform 11 (EC 2.7.11.1)
Coleóptilo	<i>Oryza sativa</i>	P07730	Glutelin type-A 2 precursor (Glutelin type II)
Coleóptilo	<i>Oryza sativa</i>	P49083	Guanine nucleotide-binding protein alpha-1 subunit
Coleóptilo	<i>Oryza sativa</i>	Q7XBH4	Myb-related protein Myb4 (OsMyb4) (Transcription factor RLTR1)
Coleóptilo	<i>Oryza sativa</i>	P28752	Tubulin alpha-1 chain
Coleóptilo	<i>Oryza sativa</i>	Q43594	Tubulin beta-1 chain (Beta-1 tubulin)
Coleóptilo	<i>Oryza sativa</i>	P45960	Tubulin beta-4 chain (Beta-4 tubulin)
Coleóptilo	<i>Oryza sativa</i>	P46265	Tubulin beta-5 chain (Beta-5 tubulin)
Coleóptilo	<i>Zea mays</i>	P13689	Auxin-binding protein 1 precursor (ABP) (ERABP1)
Coleóptilo	<i>Zea mays</i>	P19023	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14)
Coleóptilo	<i>Zea mays</i>	P49235	Beta-glucosidase, chloroplast precursor (EC 3.2.1.21)
Coleóptilo	<i>Zea mays</i>	P80608	Cysteine synthase (EC 2.5.1.47) (O-acetylserine sulfhydrylase)
Coleóptilo	<i>Zea mays</i>	Q9ZT66	Endo-1,3;1,4-beta-D-glucanase precursor (EC 3.2.1.-)
Coleóptilo	<i>Zea mays</i>	Q09054	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 2 (EC 1.2.1.12)
Coleóptilo	<i>Zea mays</i>	Q43247	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3 (EC 1.2.1.12)
Coleóptilo	<i>Zea mays</i>	P04907	Glutathione S-transferase III (EC 2.5.1.18) (GST-III) (GST class-phi)
Coleóptilo	<i>Zea mays</i>	P80639	Eukaryotic translation initiation factor 5A (eIF-5A) (eIF-4D)
Coleóptilo	<i>Zea mays</i>	P30792	2,3-bisphosphoglycerate-independent phosphoglycerate mutase
Coleóptilo	<i>Zea mays</i>	P80607	Alpha-1,4-glucan-protein synthase [UDP-forming] (EC 2.4.1.112)
Coleóptilo	<i>Zea mays</i>	F49087	Vacuolar ATP synthase catalytic subunit A (EC 3.6.3.14)
Coleóptilo	<i>Zea mays</i>	Q41773	Vacuolar ATP synthase 16 kDa proteolipid subunit (EC 3.6.3.14)
Coleóptilo	<i>Zea mays</i>	P16976	GTP-binding protein YPTM1
Coleóptilo	<i>Zea mays</i>	Q05737	GTP-binding protein YPTM2
Cotilédone	<i>Arabidopsis thaliana</i>	P46421	Glutathione S-transferase 103-1A (EC 2.5.1.18)
Cotilédone	<i>Bauhinia unguolata</i>	P83594	Factor Xa inhibitor BuXI
Cotilédone	<i>Bauhinia variegata</i>	P83595	Trypsin inhibitor BvTI
Cotilédone	<i>Brassica campestris</i>	P07088	Acyl carrier protein SF2, chloroplast precursor (ACP)
Cotilédone	<i>Brassica napus</i>	P25248	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)
Cotilédone	<i>Brassica napus</i>	Q06209	Basic endochitinase CHB4 precursor (EC 3.2.1.14)
Cotilédone	<i>Brassica napus</i>	P46267	Fructose-1,6-bisphosphatase, cytosolic (EC 3.1.3.11)
Cotilédone	<i>Brassica napus</i>	P13244	Malate synthase, glyoxysomal (EC 2.3.3.9)
Cotilédone	<i>Canavalia ensiformis</i>	P50477	Canavalin precursor
Cotilédone	<i>Canavalia ensiformis</i>	P49347	Concanavalin B precursor (Con B)
Cotilédone	<i>Carthamus tinctorius</i>	Q42713	Glycerol-3-phosphate acyltransferase, chloroplast precursor
Cotilédone	<i>Castanea crenata</i>	P82859	Agglutinin (CCA)
Cotilédone	<i>Castanea sativa</i>	P29137	Basic endochitinase CH1 (EC 3.2.1.14) (Fragment)
Cotilédone	<i>Citrullus lanatus</i>	P17783	Malate dehydrogenase, mitochondrial precursor (EC 1.1.1.37)
Cotilédone	<i>Cucumis sativus</i>	P13443	Glycerate dehydrogenase (EC 1.1.1.29)
Cotilédone	<i>Cucumis sativus</i>	P93111	Glutamyl-tRNA reductase 1, chloroplast precursor (EC 1.2.1.70)
Cotilédone	<i>Cucumis sativus</i>	P49295	Glutamyl-tRNA reductase 2, chloroplast precursor (EC 1.2.1.70)
Cotilédone	<i>Cucumis sativus</i>	P46488	Malate dehydrogenase, glyoxysomal precursor (EC 1.1.1.37)
Cotilédone	<i>Cucumis sativus</i>	P42066	Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49)
Cotilédone	<i>Cucumis sativus</i>	Q41249	Protochlorophyllide reductase, chloroplast precursor (EC 1.3.1.33)
Cotilédone	<i>Cucurbita máxima</i>	Q39649	2S albumin precursor
Cotilédone	<i>Cucurbita máxima</i>	P93110	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)

Local	Espécie	AC	Proteína
Cotilédone	<i>Cucurbita máxima</i>	Q64894	Acyl-coenzyme A oxidase, peroxisomal precursor (EC 1.3.3.6)
Cotilédone	<i>Cucurbita máxima</i>	Q05045	Chaperonin CPN60-1, mitochondrial precursor (HSP60-1)
Cotilédone	<i>Cucurbita máxima</i>	Q05046	Chaperonin CPN60-2, mitochondrial precursor (HSP60-2)
Cotilédone	<i>Cucurbita moschata</i>	P10349	Glycerol-3-phosphate acyltransferase, chloroplast precursor
Cotilédone	<i>Cucurbita pepo</i>	P48350	Catalase isozyme 1 (EC 1.11.1.6)
Cotilédone	<i>Cucurbita pepo</i>	P48351	Catalase isozyme 2 (EC 1.11.1.6)
Cotilédone	<i>Cucurbita pepo</i>	P48352	Catalase isozyme 3 (EC 1.11.1.6)
Cotilédone	<i>Glycine max</i>	P19594	2S albumin precursor (GM2S-1)
Cotilédone	<i>Glycine max</i>	P13917	Basic 7S globulin precursor (Bg) (SBg7S)
Cotilédone	<i>Glycine max</i>	P45456	Isocitrate lyase 1 (EC 4.1.3.1) (Isocitrase 1) (Isocitratase 1)
Cotilédone	<i>Glycine max</i>	P45457	Isocitrate lyase 2 (EC 4.1.3.1) (Isocitrase 2) (Isocitratase 2)
Cotilédone	<i>Glycine max</i>	Q03773	Glucan endo-1,3-beta-glucosidase precursor (EC 3.2.1.39)
Cotilédone	<i>Glycine max</i>	P13916	Beta-conglycinin, alpha chain precursor
Cotilédone	<i>Glycine max</i>	P08170	Seed lipoxygenase-1 (EC 1.13.11.12) (L-1)
Cotilédone	<i>Glycine max</i>	P38417	Lipoxygenase-4 (EC 1.13.11.12) (L-4) (VSP94)
Cotilédone	<i>Glycine max</i>	P24095	Seed lipoxygenase (EC 1.13.11.12)
Cotilédone	<i>Glycine max</i>	P45458	Malate synthase, glyoxysomal (EC 2.3.3.9) (MS) (Fragment)
Cotilédone	<i>Glycine max</i>	Q42777	Methylcrotonoyl-CoA carboxylase subunit alpha, (EC 6.4.1.4)
Cotilédone	<i>Glycine max</i>	P37228	Malate dehydrogenase, glyoxysomal precursor (EC 1.1.1.37)
Cotilédone	<i>Gossypium hirsutum</i>	P17069	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)
Cotilédone	<i>Gossypium hirsutum</i>	P17598	Catalase isozyme 1 (EC 1.11.1.6)
Cotilédone	<i>Gossypium hirsutum</i>	P30567	Catalase isozyme 2 (EC 1.11.1.6)
Cotilédone	<i>Gossypium hirsutum</i>	P46518	Late embryogenesis abundant protein Lea14-A
Cotilédone	<i>Gossypium hirsutum</i>	P46521	Late embryogenesis abundant protein Lea5-A
Cotilédone	<i>Gossypium hirsutum</i>	P46522	Late embryogenesis abundant protein Lea5-D
Cotilédone	<i>Helianthus annuus</i>	P45739	Catalase (EC 1.11.1.6)
Cotilédone	<i>Ipomoea nil</i>	P45853	Germin-like protein precursor
Cotilédone	<i>Ipomoea nil</i>	P40619	HMG1/2-like protein
Cotilédone	<i>Ipomoea nil</i>	Q01544	17.2 kDa class II heat shock protein
Cotilédone	<i>Ipomoea nil</i>	Q01545	18.8 kDa class II heat shock protein
Cotilédone	<i>Ipomoea nil</i>	P51819	Heat shock protein 83
Cotilédone	<i>Leucaena glauca</i>	P83036	Kunitz-type trypsin inhibitor LITI, alpha chain
Cotilédone	<i>Linum usitatissimum</i>	P81785	MLO-like protein (Fragment)
Cotilédone	<i>Lycopersicon esculentum</i>	P23418	Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1)
Cotilédone	<i>Lycopersicon esculentum</i>	P23419	Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2)
Cotilédone	<i>Manihot esculenta</i>	P52705	(S)-acetone-cyanohydrin lyase (EC 4.1.2.39)
Cotilédone	<i>Phaseolus lunatus</i>	P16300	Lectin precursor (LBL)
Cotilédone	<i>Phaseolus lunatus</i>	P80463	Phaseolin precursor
Cotilédone	<i>Phaseolus lunatus</i>	Q43617	Phaseolin precursor (Fragment)
Cotilédone	<i>Phaseolus vulgaris</i>	Q41114	Alpha-amylase inhibitor 2 precursor (Alpha-AI-2)
Cotilédone	<i>Phaseolus vulgaris</i>	P07219	Phaseolin, alpha-type precursor
Cotilédone	<i>Phaseolus vulgaris</i>	P02853	Phaseolin, beta-type precursor
Cotilédone	<i>Picea abies</i>	P51063	Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC)
Cotilédone	<i>Pinus sylvestris</i>	P15192	Chlorophyll a-b binding protein type 2 member 2
Cotilédone	<i>Pinus sylvestris</i>	P15193	Chlorophyll a-b binding protein type 2 member 1A,
Cotilédone	<i>Pinus sylvestris</i>	P15194	Chlorophyll a-b binding protein type 2 member 1B,
Cotilédone	<i>Pinus sylvestris</i>	Q36664	NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3)
Cotilédone	<i>Pisum sativum</i>	Q06931	ABA-responsive protein ABR17
Cotilédone	<i>Pisum sativum</i>	Q06930	ABA-responsive protein ABR18
Cotilédone	<i>Pisum sativum</i>	P62927	Albumin 1 B precursor (PA1 B)
Cotilédone	<i>Pisum sativum</i>	P28639	Dehydrin DHN1
Cotilédone	<i>Pisum sativum</i>	Q9SM60	Phosphoglucomutase, cytoplasmic (EC 5.4.2.2)
Cotilédone	<i>Pisum sativum</i>	Q9SM59	Phosphoglucomutase, chloroplast precursor (EC 5.4.2.2)
Cotilédone	<i>Raphanus sativus</i>	P27626	Senescence-associated protein DIN1
Cotilédone	<i>Ricinus communis</i>	P55143	Glutaredoxin
Cotilédone	<i>Ricinus communis</i>	Q07423	Hexose carrier protein HEX6
Cotilédone	<i>Sesamum indicum</i>	P48620	Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-)
Cotilédone	<i>Sinapis alba</i>	P13851	Chlorophyll a-b binding protein 1, chloroplast precursor
Cotilédone	<i>Sinapis alba</i>	P13416	Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1)
Cotilédone	<i>Sinapis alba</i>	P13417	Chalcone synthase 3 (EC 2.3.1.74) (Naringenin-chalcone synthase 3)
Cotilédone	<i>Vicia faba</i>	P42653	14-3-3-like protein A (VFA-1433A)
Cotilédone	<i>Vicia faba</i>	P42654	14-3-3-like protein B (VFA-1433B)
Cotilédone	<i>Vicia faba</i>	P52416	Glucose-1-phosphate adenyltransferase small subunit 1.
Cotilédone	<i>Vicia faba</i>	P52417	Glucose-1-phosphate adenyltransferase small subunit 2.
Cotilédone	<i>Vicia faba</i>	P40620	HMG1/2-like protein
Cotilédone	<i>Vicia faba</i>	P53537	Alpha-glucan phosphorylase, H isozyme (EC 2.4.1.1)
Cotilédone	<i>Vicia faba</i>	P53536	Alpha-1,4 glucan phosphorylase, L isozyme, (EC 2.4.1.1)
Cotilédone	<i>Vigna mungo</i>	P17859	Alpha-amylase precursor (EC 3.2.1.1)

Local	Espécie	AC	Proteína
Endosperma	<i>Arabidopsis thaliana</i>	P46421	Glutathione S-transferase 103-1A (EC 2.5.1.18)
Endosperma	<i>Bauhinia unguolata</i>	P83594	Factor Xa inhibitor BuXI
Endosperma	<i>Bauhinia variegata</i>	P83595	Trypsin inhibitor BvTI
Endosperma	<i>Brassica campestris</i>	P07088	Acyl carrier protein SF2, chloroplast precursor (ACP)
Endosperma	<i>Brassica napus</i>	P25248	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)
Endosperma	<i>Brassica napus</i>	Q06209	Basic endochitinase CHB4 precursor (EC 3.2.1.14)
Endosperma	<i>Brassica napus</i>	P46267	Fructose-1,6-bisphosphatase, cytosolic (EC 3.1.3.11)
Endosperma	<i>Brassica napus</i>	P13244	Malate synthase, glyoxysomal (EC 2.3.3.9)
Endosperma	<i>Canavalia ensiformis</i>	P50477	Canavalin precursor
Endosperma	<i>Canavalia ensiformis</i>	P49347	Concanavalin B precursor (Con B)
Endosperma	<i>Carthamus tinctorius</i>	Q42713	Glycerol-3-phosphate acyltransferase, (EC 2.3.1.15) (GPAT)
Endosperma	<i>Castanea tinctoria</i>	P82859	Agglutinin (CCA)
Endosperma	<i>Castanea sativa</i>	P29137	Basic endochitinase CH1 (EC 3.2.1.14) (Fragment)
Endosperma	<i>Citrullus lanatus</i>	P17783	Malate dehydrogenase, mitochondrial precursor (EC 1.1.1.37)
Endosperma	<i>Cucumis sativus</i>	P13443	Glycerate dehydrogenase (EC 1.1.1.29)
Endosperma	<i>Cucumis sativus</i>	P93111	Glutamyl-tRNA reductase 1, chloroplast precursor (EC 1.2.1.70)
Endosperma	<i>Cucumis sativus</i>	P49295	Glutamyl-tRNA reductase 2, chloroplast precursor (EC 1.2.1.70)
Endosperma	<i>Cucumis sativus</i>	P46488	Malate dehydrogenase, glyoxysomal precursor (EC 1.1.1.37)
Endosperma	<i>Cucumis sativus</i>	P42066	Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49)
Endosperma	<i>Cucumis sativus</i>	Q41249	Protochlorophyllide reductase, chloroplast precursor (EC 1.3.1.33)
Endosperma	<i>Cucurbita máxima</i>	Q39649	2S albumin precursor
Endosperma	<i>Cucurbita máxima</i>	P93110	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)
Endosperma	<i>Cucurbita máxima</i>	Q64894	Acyl-coenzyme A oxidase, peroxisomal precursor (EC 1.3.3.6)
Endosperma	<i>Cucurbita máxima</i>	Q05045	Chaperonin CPN60-1, mitochondrial precursor (HSP60-1)
Endosperma	<i>Cucurbita máxima</i>	Q05046	Chaperonin CPN60-2, mitochondrial precursor (HSP60-2)
Endosperma	<i>Cucurbita moschata</i>	P10349	Glycerol-3-phosphate acyltransferase, chloroplast precursor
Endosperma	<i>Cucurbita pepo</i>	P48350	Catalase isozyme 1 (EC 1.11.1.6)
Endosperma	<i>Cucurbita pepo</i>	P48351	Catalase isozyme 2 (EC 1.11.1.6)
Endosperma	<i>Cucurbita pepo</i>	P48352	Catalase isozyme 3 (EC 1.11.1.6)
Endosperma	<i>Glycine max</i>	P19594	2S albumin precursor (GM2S-1)
Endosperma	<i>Glycine max</i>	P13917	Basic 7S globulin precursor (Bg) (SBg7S)
Endosperma	<i>Glycine max</i>	P45456	Isocitrate lyase 1 (EC 4.1.3.1) (Isocitrase 1) (Fragment)
Endosperma	<i>Glycine max</i>	P45457	Isocitrate lyase 2 (EC 4.1.3.1) (Isocitrase 2) (Fragment)
Endosperma	<i>Glycine max</i>	Q03773	Glucan endo-1,3-beta-glucosidase precursor (EC 3.2.1.39)
Endosperma	<i>Glycine max</i>	P13916	Beta-conglycinin, alpha chain precursor
Endosperma	<i>Glycine max</i>	P08170	Seed lipoxygenase-1 (EC 1.13.11.12) (L-1)
Endosperma	<i>Glycine max</i>	P38417	Lipoxygenase-4 (EC 1.13.11.12) (L-4) (VSP94)
Endosperma	<i>Glycine max</i>	P24095	Seed lipoxygenase (EC 1.13.11.12)
Endosperma	<i>Glycine max</i>	P45458	Malate synthase, glyoxysomal (EC 2.3.3.9) (MS) (Fragment)
Endosperma	<i>Glycine max</i>	Q42777	Methylcrotonoyl-CoA carboxylase subunit alpha, (EC 6.4.1.4)
Endosperma	<i>Glycine max</i>	P37228	Malate dehydrogenase, glyoxysomal precursor (EC 1.1.1.37)
Endosperma	<i>Gossypium hirsutum</i>	P17069	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL)
Endosperma	<i>Gossypium hirsutum</i>	P17598	Catalase isozyme 1 (EC 1.11.1.6)
Endosperma	<i>Gossypium hirsutum</i>	P30567	Catalase isozyme 2 (EC 1.11.1.6)
Endosperma	<i>Gossypium hirsutum</i>	P46518	Late embryogenesis abundant protein Lea14-A
Endosperma	<i>Gossypium hirsutum</i>	P46521	Late embryogenesis abundant protein Lea5-A
Endosperma	<i>Gossypium hirsutum</i>	P46522	Late embryogenesis abundant protein Lea5-D
Endosperma	<i>Helianthus annuus</i>	P45739	Catalase (EC 1.11.1.6)
Endosperma	<i>Ipomoea nil</i>	P45853	Germin-like protein precursor
Endosperma	<i>Ipomoea nil</i>	P40619	HMG1/2-like protein
Endosperma	<i>Ipomoea nil</i>	Q01544	17.2 kDa class II heat shock protein
Endosperma	<i>Ipomoea nil</i>	Q01545	18.8 kDa class II heat shock protein
Endosperma	<i>Ipomoea nil</i>	P51819	Heat shock protein 83
Endosperma	<i>Leucaena glauca</i>	P83036	Kunitz-type trypsin inhibitor LITI, alpha chain
Endosperma	<i>Linum usitatissimum</i>	P81785	MLO-like protein (Fragment)
Endosperma	<i>Lycopersicon esculentum</i>	P23418	Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1)
Endosperma	<i>Lycopersicon esculentum</i>	P23419	Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2)
Endosperma	<i>Manihot esculenta</i>	P52705	(S)-acetone-cyanohydrin lyase (EC 4.1.2.39)
Endosperma	<i>Phaseolus lunatus</i>	P16300	Lectin precursor (LBL)
Endosperma	<i>Phaseolus lunatus</i>	P80463	Phaseolin precursor
Endosperma	<i>Phaseolus lunatus</i>	Q43617	Phaseolin precursor (Fragment)
Endosperma	<i>Phaseolus vulgaris</i>	Q41114	Alpha-amylase inhibitor 2 precursor (Alpha-AI-2) (Alpha-AI2)
Endosperma	<i>Phaseolus vulgaris</i>	P07219	Phaseolin, alpha-type precursor
Endosperma	<i>Phaseolus vulgaris</i>	P02853	Phaseolin, beta-type precursor
Endosperma	<i>Picea abies</i>	P51063	Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC)
Endosperma	<i>Pinus sylvestris</i>	P15192	Chlorophyll a-b binding protein type 2 member 2
Endosperma	<i>Pinus sylvestris</i>	P15193	Chlorophyll a-b binding protein type 2 member 1A,
Endosperma	<i>Pinus sylvestris</i>	P15194	Chlorophyll a-b binding protein type 2 member 1B

Local	Espécie	AC	Proteína
Endosperma	<i>Pinus sylvestris</i>	Q36664	NADH-ubiquinone oxidoreductase chain 3 (EC 1.6.5.3)
Endosperma	<i>Pisum sativum</i>	Q06931	ABA-responsive protein ABR17
Endosperma	<i>Pisum sativum</i>	Q06930	ABA-responsive protein ABR18
Endosperma	<i>Pisum sativum</i>	P62927	Albumin 1 B precursor (PA1 B)
Endosperma	<i>Pisum sativum</i>	P28639	Dehydrin DHN1
Endosperma	<i>Pisum sativum</i>	Q9SM60	Phosphoglucosmutase, cytoplasmic (EC 5.4.2.2)
Endosperma	<i>Pisum sativum</i>	Q9SM59	Phosphoglucosmutase, chloroplast precursor (EC 5.4.2.2)
Endosperma	<i>Raphanus sativus</i>	P27626	Senescence-associated protein DIN1
Endosperma	<i>Ricinus communis</i>	P55143	Glutaredoxin
Endosperma	<i>Ricinus communis</i>	Q07423	Hexose carrier protein HEX6
Endosperma	<i>Sesamum indicum</i>	P48620	Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-)
Endosperma	<i>Sinapis alba</i>	P13851	Chlorophyll a-b binding protein 1, chloroplast precursor
Endosperma	<i>Sinapis alba</i>	P13416	Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1)
Endosperma	<i>Sinapis alba</i>	P13417	Chalcone synthase 3 (EC 2.3.1.74) (Naringenin-chalcone synthase 3)
Endosperma	<i>Vicia faba</i>	P42653	14-3-3-like protein A (VFA-1433A)
Endosperma	<i>Vicia faba</i>	P42654	14-3-3-like protein B (VFA-1433B)
Endosperma	<i>Vicia faba</i>	P52416	Glucose-1-phosphate adenylyltransferase subunit 1, (EC 2.7.7.27)
Endosperma	<i>Vicia faba</i>	P52417	Glucose-1-phosphate adenylyltransferase subunit 2 (EC 2.7.7.27)
Endosperma	<i>Vicia faba</i>	P40620	HMG1/2-like protein
Endosperma	<i>Vicia faba</i>	P53537	Alpha-glucan phosphorylase, H isozyme (EC 2.4.1.1)
Endosperma	<i>Vicia faba</i>	P53536	Alpha-1,4 glucan phosphorylase, L isozyme, (EC 2.4.1.1)
Endosperma	<i>Vigna mungo</i>	P17859	Alpha-amylase precursor (EC 3.2.1.1)
Epicótilo	<i>Cicer arietinum</i>	Q49816	Late embryogenesis abundant protein 1 (CapLEA-1)
Epicótilo	<i>Cicer arietinum</i>	Q49817	Late embryogenesis abundant protein 2 (CapLEA-2)
Epicótilo	<i>Glycine max</i>	P35694	Brassinosteroid-regulated protein BRU1 precursor
Epicótilo	<i>Glycine max</i>	P48631	Omega-6 fatty acid desaturase, endoplasmic reticulum isozyme 2
Epicótilo	<i>Glycine max</i>	Q22518	40S ribosomal protein SA (p40)
Epicótilo	<i>Glycine max</i>	Q43468	Heat shock protein STI (Stress-inducible protein)
Epicótilo	<i>Lupinus luteus</i>	Q49886	Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)
Epicótilo	<i>Lupinus luteus</i>	Q49885	60S ribosomal protein L13a
Epicótilo	<i>Lupinus luteus</i>	Q49884	60S ribosomal protein L30
Epicótilo	<i>Pisum sativum</i>	Q01286	Chalcone synthase 1 (EC 2.3.1.74) (Naregenin-chalcone synthase 1)
Epicótilo	<i>Pisum sativum</i>	Q01287	Chalcone synthase 2 (EC 2.3.1.74) (Naregenin-chalcone synthase 2)
Epicótilo	<i>Pisum sativum</i>	Q01288	Chalcone synthase 6 (EC 2.3.1.74) (Naregenin-chalcone synthase 6)
Epicótilo	<i>Pisum sativum</i>	Q01861	Phenylalanine ammonia-lyase 1 (EC 4.3.1.5)
Epicótilo	<i>Pisum sativum</i>	Q04593	Phenylalanine ammonia-lyase 2 (EC 4.3.1.5)
Epicótilo	<i>Zea mays</i>	P18123	Catalase isozyme 3 (EC 1.11.1.6)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q9LLH9	Acyl-coenzyme A oxidase 3, peroxisomal precursor (EC 1.3.3.6)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q9SCX5	Probable potassium channel AKT5
Hipocótilo	<i>Arabidopsis thaliana</i>	Q9FXD6	Two-component response regulator ARR11 (Receiver-like protein 3)
Hipocótilo	<i>Arabidopsis thaliana</i>	P46310	Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-)
Hipocótilo	<i>Arabidopsis thaliana</i>	P48622	Temperature-sensitive omega-3 fatty acid desaturase, (EC 1.14.19.-)
Hipocótilo	<i>Arabidopsis thaliana</i>	P48623	Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.19.-)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q9SPK5	Formate-tetrahydrofolate ligase (EC 6.3.4.3)
Hipocótilo	<i>Arabidopsis thaliana</i>	P46421	Glutathione S-transferase 103-1A (EC 2.5.1.18)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q9FE22	Long hypocotyl in far-red 1 (bHLH-like protein HFR1)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q9FUR2	Type I inositol-1,4,5-trisphosphate 5-phosphatase 2 (EC 3.1.3.56)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q7GB25	Multidrug resistance-associated protein 5 (EC 3.6.3.44)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q48963	Phototropin-1 (EC 2.7.11.1)
Hipocótilo	<i>Arabidopsis thaliana</i>	P93733	Phospholipase D beta 1 (EC 3.1.4.4)
Hipocótilo	<i>Arabidopsis thaliana</i>	Q39189	DEAD-box ATP-dependent RNA helicase 7 (EC 3.6.1.-)
Hipocótilo	<i>Glycine max</i>	Q01915	ATP synthase alpha chain, mitochondrial (EC 3.6.3.14)
Hipocótilo	<i>Glycine max</i>	P19976	Ferritin-1, chloroplast precursor (EC 1.16.3.1)
Hipocótilo	<i>Helianthus annuus</i>	P49198	40S ribosomal protein S3a
Hipocótilo	<i>Lycopersicon esculentum</i>	P23418	Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1)
Hipocótilo	<i>Lycopersicon esculentum</i>	P23419	Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2)
Hipocótilo	<i>Lycopersicon esculentum</i>	P51107	Dihydroflavonol-4-reductase (EC 1.1.1.219)
Hipocótilo	<i>Lycopersicon esculentum</i>	P80022	Extracellular ribonuclease LE precursor (EC 3.1.27.1)
Hipocótilo	<i>Lycopersicon esculentum</i>	P80196	Intracellular ribonuclease LX precursor (EC 3.1.27.1)
Hipocótilo	<i>Phaseolus aureus</i>	Q01912	1-aminocyclopropane-1-carboxylate synthase (EC 4.4.1.14)
Hipocótilo	<i>Phaseolus aureus</i>	P32292	Indole-3-acetic acid-induced protein ARG2
Hipocótilo	<i>Phaseolus aureus</i>	P32295	Indole-3-acetic acid-induced protein ARG7
Hipocótilo	<i>Phaseolus aureus</i>	P21616	Pyrophosphate-energized vacuolar membrane proton pump
Hipocótilo	<i>Phaseolus aureus</i>	P32293	Auxin-induced protein 22A
Hipocótilo	<i>Phaseolus aureus</i>	P32294	Auxin-induced protein 22B
Hipocótilo	<i>Phaseolus aureus</i>	O24541	Auxin-induced protein 22C
Hipocótilo	<i>Phaseolus aureus</i>	O24542	Auxin-induced protein 22D
Hipocótilo	<i>Phaseolus aureus</i>	O24543	Auxin-induced protein 22E

Local	Espécie	AC	Proteína
Hipocótilo	<i>Phaseolus aureus</i>	P32291	Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.19.-)
Hipocótilo	<i>Phaseolus aureus</i>	P29001	Acid beta-fructofuranosidase precursor (EC 3.2.1.26)
Hipocótilo	<i>Phaseolus aureus</i>	P37115	Trans-cinnamate 4-monooxygenase (EC 1.14.13.11)
Hipocótilo	<i>Phaseolus aureus</i>	O22552	Vacuolar ATP synthase 16 kDa proteolipid subunit (EC 3.6.3.14)
Hipocótilo	<i>Phaseolus vulgaris</i>	P35334	Polygalacturonase inhibitor 1 precursor
Hipocótilo	<i>Phaseolus vulgaris</i>	P58822	Polygalacturonase inhibitor 2 precursor
Hipocótilo	<i>Phaseolus vulgaris</i>	P58823	Polygalacturonase inhibitor 3 precursor
Hipocótilo	<i>Ricinus communis</i>	Q01297	Catalase isozyme 1 (EC 1.11.1.6)
Hipocótilo	<i>Ricinus communis</i>	P49318	Catalase isozyme 2 (EC 1.11.1.6)
Hipocótilo	<i>Solanum melongena</i>	P37118	Cytochrome P450 71A2 (EC 1.14.-.-) (CYPLXXIA2) (P-450EG4)
Hipocótilo	<i>Solanum melongena</i>	P37119	Cytochrome P450 71A3 (EC 1.14.-.-) (CYPLXXIA3) (P-450EG3)
Hipocótilo	<i>Solanum melongena</i>	P37117	Cytochrome P450 71A4 (EC 1.14.-.-) (CYPLXXIA4) (P-450EG2)
Hipocótilo	<i>Solanum melongena</i>	P37120	Flavonoid 3',5'-hydroxylase (EC 1.14.13.88) (F3'5'H)
Hipocótilo	<i>Solanum melongena</i>	P37121	Cytochrome P450 76A1 (EC 1.14.-.-) (CYPLXXVIA1)
Hipocótilo	<i>Solanum melongena</i>	P37122	Cytochrome P450 76A2 (EC 1.14.-.-) (CYPLXXVIA2) (P-450EG7)
Hipocótilo	<i>Solanum melongena</i>	P37123	Cytochrome P450 77A1 (EC 1.14.-.-) (CYPLXXVIA1)
Hipocótilo	<i>Solanum melongena</i>	P37124	Cytochrome P450 77A2 (EC 1.14.-.-) (CYPLXXVIA2) (P-450EG5)
Hipocótilo	<i>Solanum melongena</i>	P55311	Catalase (EC 1.11.1.6)
Hipocótilo	<i>Solanum melongena</i>	Q43641	Anthocyanidin 3-O-glucosyltransferase (EC 2.4.1.115)
Hipocótilo	<i>Triticum aestivum</i>	P60098	Mitochondrial ribosomal protein S12
Hipocótilo	<i>Zea mays</i>	P60099	Mitochondrial ribosomal protein S12
Mesocótilo	<i>Avena sativa</i>	P40412	T-complex protein 1 subunit epsilon (TCP-1-epsilon)
Mesocótilo	<i>Avena sativa</i>	P54411	T-complex protein 1 subunit epsilon (TCP-1-epsilon)
Plântula	<i>Allium ascalonicum</i>	P31756	Alliin lyase precursor (EC 4.4.1.4) (Alliinase)
Plântula	<i>Allium cepa</i>	Q96008	Mitochondrial ribosomal protein S12
Plântula	<i>Allium sativum</i>	Q01594	Alliin lyase 1 precursor (EC 4.4.1.4) (Alliinase 1)
Plântula	<i>Arabidopsis thaliana</i>	Q38953	Pre-mRNA-splicing factor ATP-dependent RNA helicase (EC 3.6.1.-)
Plântula	<i>Arabidopsis thaliana</i>	Q82804	Protein EARLY FLOWERING 3 (Nematode-responsive protein)
Plântula	<i>Arabidopsis thaliana</i>	P35614	Eukaryotic peptide chain release factor subunit 1-3
Plântula	<i>Arabidopsis thaliana</i>	P25856	Glyceraldehyde-3-phosphate dehydrogenase A, (EC 1.2.1.13)
Plântula	<i>Arabidopsis thaliana</i>	P43288	Shaggy-related protein kinase alpha (EC 2.7.11.1) (ASK-alpha)
Plântula	<i>Arabidopsis thaliana</i>	P43289	Shaggy-related protein kinase gamma (EC 2.7.11.1) (ASK-gamma)
Plântula	<i>Arabidopsis thaliana</i>	Q42589	Nonspecific lipid-transfer protein 1 precursor (LTP 1)
Plântula	<i>Arabidopsis thaliana</i>	Q42191	Inner membrane protein OXA1, mitochondrial precursor
Plântula	<i>Arabidopsis thaliana</i>	P42731	Polyadenylate-binding protein 2 (Poly(A)-binding protein 2)
Plântula	<i>Arabidopsis thaliana</i>	Q43867	Pectinesterase-1 precursor (EC 3.1.1.11) (Pectin methylesterase 1)
Plântula	<i>Arabidopsis thaliana</i>	Q944S1	DEAD-box ATP-dependent RNA helicase 22 (EC 3.6.1.-)
Plântula	<i>Arabidopsis thaliana</i>	P42794	60S ribosomal protein L11-2 (L16)
Plântula	<i>Arabidopsis thaliana</i>	Q38961	DNA cross-link repair protein SNM1 (AtSNM1)
Plântula	<i>Arabidopsis thaliana</i>	Q41951	Aquaporin TIP2.1 (Tonoplast intrinsic protein 2.1)
Plântula	<i>Brassica napus</i>	O65727	Squalene monooxygenase 1,1 (EC 1.14.99.7)
Plântula	<i>Brassica napus</i>	O65726	Squalene monooxygenase 1,2 (EC 1.14.99.7)
Plântula	<i>Glycine max</i>	Q07185	Alternative oxidase 1, mitochondrial precursor
Plântula	<i>Glycine max</i>	Q02226	Cytochrome c oxidase subunit 2, mitochondrial precursor (EC 1.9.3.1)
Plântula	<i>Glycine max</i>	Q02243	Wound-induced protein (Fragment)
Plântula	<i>Helianthus annuus</i>	Q96033	Mitochondrial ribosomal protein S12
Plântula	<i>Hordeum vulgare</i>	P49968	Signal recognition particle 54 kDa protein 1 (SRP54)
Plântula	<i>Hordeum vulgare</i>	P49969	Signal recognition particle 54 kDa protein 2 (SRP54)
Plântula	<i>Hordeum vulgare</i>	P49970	Signal recognition particle 54 kDa protein 3 (SRP54)
Plântula	<i>Lycopersicon esculentum</i>	P27057	GAST1 protein precursor
Plântula	<i>Lycopersicon esculentum</i>	Q41330	Homeotic protein knotted-1 (TKN1)
Plântula	<i>Lycopersicon esculentum</i>	Q9XG54	12-oxophytodienoate reductase 1 (EC 1.3.1.42)
Plântula	<i>Lycopersicon esculentum</i>	Q9FEW9	12-oxophytodienoate reductase 3 (EC 1.3.1.42)
Plântula	<i>Lycopersicon esculentum</i>	Q9FEX0	12-oxophytodienoate reductase-like protein (EC 1.3.1.-) (LeOPR2)
Plântula	<i>Lycopersicon esculentum</i>	Q8RVB2	UDP-N-acetylglucosamine--peptide N- acetylglucosaminyltransferase
Plântula	<i>Medicago sativa</i>	Q9ZT05	Photosystem I reaction center subunit psaK,
Plântula	<i>Oryza sativa</i>	P15998	ATP synthase alpha chain, mitochondrial (EC 3.6.3.14)
Plântula	<i>Oryza sativa</i>	P14578	Cytochrome c oxidase subunit 1 (EC 1.9.3.1)
Plântula	<i>Oryza sativa</i>	P14833	Cytochrome b
Plântula	<i>Oryza sativa</i>	O22567	1-deoxy-D-xylulose-5-phosphate synthase, chloroplast precursor
Plântula	<i>Oryza sativa</i>	Q42977	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic
Plântula	<i>Oryza sativa</i>	P14656	Glutamine synthetase cytosolic isozyme 1-1 (EC 6.3.1.2)
Plântula	<i>Oryza sativa</i>	P14655	Glutamine synthetase, chloroplast precursor (EC 6.3.1.2)
Plântula	<i>Oryza sativa</i>	Q8L481	Probable cation transporter HKT3 (OsHKT3)
Plântula	<i>Oryza sativa</i>	Q59AF4	Auxin-responsive protein IAA4 (Indoleacetic acid-induced protein 4)
Plântula	<i>Oryza sativa</i>	Q59AF3	Auxin-responsive protein IAA5 (Indoleacetic acid-induced protein 5)
Plântula	<i>Oryza sativa</i>	Q9XJ60	MADS-box transcription factor 50 (OsMADS50)
Plântula	<i>Oryza sativa</i>	P49030	Protein mago nashi homolog (Mago nashi-like protein)

Local	Espécie	AC	Proteína
Plântula	<i>Oryza sativa</i>	Q42978	Nonspecific lipid-transfer protein 2 precursor (LTP 2)
Plântula	<i>Oryza sativa</i>	Q42976	Nonspecific lipid-transfer protein 3 precursor (LTP 3)
Plântula	<i>Oryza sativa</i>	P22181	Photosystem I reaction center subunit VI, chloroplast precursor
Plântula	<i>Oryza sativa</i>	P92812	Mitochondrial 60S ribosomal protein L2
Plântula	<i>Oryza sativa</i>	P28756	Superoxide dismutase [Cu-Zn] 1 (EC 1.15.1.1)
Plântula	<i>Oryza sativa</i>	Q40635	Vacuolar ATP synthase 16 kDa proteolipid subunit (EC 3.6.3.14)
Plântula	<i>Phaseolus vulgaris</i>	P39865	Nitrate reductase [NADH] 1 (EC 1.7.1.1) (NR-1)
Plântula	<i>Pinus taeda</i>	Q6J163	Auxin-induced protein 5NG4
Plântula	<i>Pinus taeda</i>	Q41073	Nonspecific lipid-transfer protein precursor (LTP)
Plântula	<i>Pisum sativum</i>	P31239	1-aminocyclopropane-1-carboxylate oxidase (EC 1.14.17.4)
Plântula	<i>Pisum sativum</i>	P00155	Apocytchrome f precursor
Plântula	<i>Pisum sativum</i>	P81406	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
Plântula	<i>Pisum sativum</i>	P16002	Plastocyanin, chloroplast precursor
Plântula	<i>Triticum aestivum</i>	P93596	Cytochrome P450 51 (EC 1.14.13.70)
Plântula	<i>Triticum aestivum</i>	P46525	Cold-shock protein CS120
Plântula	<i>Triticum aestivum</i>	P46526	Cold shock protein CS66
Plântula	<i>Triticum aestivum</i>	Q41578	Protochlorophyllide reductase A, chloroplast precursor (EC 1.3.1.33)
Plântula	<i>Vitis vinifera</i>	P52596	Glutamate dehydrogenase (EC 1.4.1.3) (GDH)
Plântula	<i>Vitis vinifera</i>	P51118	Glutamine synthetase cytosolic isozyme 1 (EC 6.3.1.2)
Plântula	<i>Vitis vinifera</i>	P51119	Glutamine synthetase cytosolic isozyme 2 (EC 6.3.1.2)
Plântula	<i>Zea mays</i>	P13689	Auxin-binding protein 1 precursor (ABP) (ERABP1)
Plântula	<i>Zea mays</i>	P49235	Beta-glucosidase, chloroplast precursor (EC 3.2.1.21)
Plântula	<i>Zea mays</i>	Q43260	Glutamate dehydrogenase (EC 1.4.1.3) (GDH)
Plântula	<i>Zea mays</i>	Q43272	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
Plântula	<i>Zea mays</i>	P46605	Homeobox protein HOX1A
Plântula	<i>Zea mays</i>	P22275	Tubulin alpha-3 chain (Alpha-3 tubulin)
Plântula	<i>Zea mays</i>	Q02245	Tubulin alpha-5 chain (Alpha-5 tubulin)
Plântula	<i>Zea mays</i>	P33627	Tubulin alpha-6 chain (Alpha-6 tubulin)
Sem. Imatura	<i>Cucurbita maxima</i>	Q39548	Ent-kaurene synthase B, chloroplast precursor (EC 4.2.3.19) (KSB)
Sem. Imatura	<i>Fagopyrum esculentum</i>	O23878	13S globulin seed storage protein 1 precursor
Sem. Imatura	<i>Fagopyrum esculentum</i>	O23880	13S globulin seed storage protein 2 precursor (Legumin-like protein 2)
Sem. Imatura	<i>Oryza sativa</i>	Q8H930	Putative UDP-arabinose 4-epimerase 1 (EC 5.1.3.5)
Sem. Imatura	<i>Oryza sativa</i>	Q8H0B6	Putative UDP-arabinose 4-epimerase 2 (EC 5.1.3.5)
Sem. Imatura	<i>Oryza sativa</i>	Q8L4E7	SAP-like protein BP-73 (OsBP-73) (Riaa1)
Sem. Imatura	<i>Oryza sativa</i>	P29611	Catalase isozyme A (EC 1.11.1.6) (CAT-A)
Sem. Imatura	<i>Oryza sativa</i>	Q09151	Glutelin type-A 3 precursor
Sem. Imatura	<i>Oryza sativa</i>	Q8S857	Probable histone H2A variant 2
Sem. Imatura	<i>Oryza sativa</i>	P69247	Histone H3
Sem. Imatura	<i>Oryza sativa</i>	P20423	Plastocyanin, chloroplast precursor
Sem. Imatura	<i>Oryza sativa</i>	Q944F4	Fructokinase-1 (EC 2.7.1.4) (Fructokinase I) (OsFKI)
Sem. Imatura	<i>Oryza sativa</i>	Q944F5	Fructokinase-2 (EC 2.7.1.4) (Fructokinase II) (OsFKII)
Sem. Imatura	<i>Phaseolus lunatus</i>	Q9ZSK5	Zeatin O-glucosyltransferase (EC 2.4.1.203)
Sem. Imatura	<i>Prunus dulcis</i>	Q43804	Oleosin 1
Semente	<i>Triticum aestivum</i>	P01085	Alpha-amylase inhibitor
Semente	<i>Triticum aestivum</i>	P27736	Granule-bound starch synthase 1, chloroplast precursor (EC 2.4.1.242)
Semente	<i>Triticum aestivum</i>	O64392	Wheatwin1 precursor (Pathogenesis-related protein 4a) (Protein 0.14)
Semente	<i>Triticum aestivum</i>	O64393	Wheatwin2 precursor (Pathogenesis-related protein 4b)
Semente	<i>Zea mays</i>	P19023	ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14)
Semente	<i>Zea mays</i>	P24067	Luminal-binding protein 2 precursor (BiP2) (Heat shock protein)
Semente	<i>Zea mays</i>	O24581	Luminal-binding protein 3 precursor (BiP3)
Semente	<i>Zea mays</i>	Q9TON8	Cytokinin dehydrogenase 1 precursor (EC 1.5.99.12)
Semente	<i>Zea mays</i>	Q93XP7	Cis-zeatin O-glucosyltransferase 1 (EC 2.4.1.215) (cisZOG1)
Semente	<i>Triticum aestivum</i>	Q43654	Starch synthase 1, chloroplast precursor (EC 2.4.1.21) (SS I)
Semente	<i>Zea mays</i>	P31726	Cystatin-1 precursor (Cystatin I)
Sementes	<i>Abrus precatorius</i>	P11140	Abrin-a precursor [Contains: Abrin-a A chain (EC 3.2.2.22)]
Sementes	<i>Abrus precatorius</i>	Q06077	Abrin-b precursor [Contains: Abrin-b A chain (EC 3.2.2.22)]
Sementes	<i>Allium cepa</i>	Q41258	Antimicrobial protein Ace-AMP1 precursor
Sementes	<i>Arabidopsis thaliana</i>	P30224	Cysteine-rich antifungal protein 1 precursor (AFP1)
Sementes	<i>Arabidopsis thaliana</i>	Q9SQH9	Ammonium transporter 1, member 3 (AtAMT1;3)
Sementes	<i>Arabidopsis thaliana</i>	Q9ZPJ8	Ammonium transporter 1, member 2 (AtAMT1;2)
Sementes	<i>Avena sativa</i>	P80356	Avenin-3 precursor (Prolamin)
Sementes	<i>Avena sativa</i>	Q09114	Avenin-E (Alpha-2 avenin) (Avenin N9) (Prolamin)
Sementes	<i>Bauhinia bauhinioides</i>	P83051	Kunitz-type proteinase inhibitor BbCI
Sementes	<i>Bauhinia bauhinioides</i>	P83052	Kunitz-type serine protease inhibitor BbKI
Sementes	<i>Brassica campestris</i>	P07088	Acyl carrier protein SF2, chloroplast precursor (ACP)
Sementes	<i>Brassica juncea</i>	P80207	Allergen Bra j 1-E (Bra j I)
Sementes	<i>Brassica napus</i>	P80208	Napin-3 (Napin BrIII) (Napin III)
Sementes	<i>Brassica napus</i>	P24565	Napin-1A (Napin BrIa)

Local	Espécie	AC	Proteína
Sementes	<i>Brassica napus</i>	P10352	Acyl carrier protein, chloroplast precursor (ACP)
Sementes	<i>Brassica napus</i>	P17650	Acyl carrier protein, chloroplast precursor (ACP)
Sementes	<i>Brassica napus</i>	P32887	Acyl carrier protein, chloroplast precursor (ACP)
Sementes	<i>Brassica napus</i>	P08971	Acyl carrier protein, chloroplast precursor (ACP)
Sementes	<i>Canavalia gladiata</i>	P10562	Canavalin precursor
Sementes	<i>Capparis masaikai</i>	P80351	Sweet protein mabinlin-1 (Mabinlin I) (MAB I)
Sementes	<i>Capparis masaikai</i>	P30233	Sweet protein mabinlin-2 precursor (Mabinlin II) (MAB II)
Sementes	<i>Capparis masaikai</i>	P80352	Sweet protein mabinlin-3 (Mabinlin III) (MAB III)
Sementes	<i>Capparis masaikai</i>	P80353	Sweet protein mabinlin-4 (Mabinlin IV)
Sementes	<i>Coffea arabica</i>	Q42656	Alpha-galactosidase precursor (EC 3.2.1.22)
Sementes	<i>Cyamopsis tetragonoloba</i>	P14749	Alpha-galactosidase precursor (EC 3.2.1.22)
Sementes	<i>Fagus sylvatica</i>	Q39752	Calmodulin (CaM)
Sementes	<i>Glycine max</i>	P19594	2S albumin precursor (GM2S-1)
Sementes	<i>Glycine max</i>	P13917	Basic 7S globulin precursor (Bg) (SBg7S)
Sementes	<i>Glycine max</i>	Q8RVH5	Basic 7S globulin 2 precursor (Bg) (SBg7S)
Sementes	<i>Glycine max</i>	Q39817	Calnexin homolog precursor
Sementes	<i>Helianthus annuus</i>	P23110	Albumin-8 precursor (Methionine-rich 2S protein) (SFA8)
Sementes	<i>Hordeum vulgare</i>	P28814	Barwin
Sementes	<i>Impatiens balsamina</i>	O24006	Antimicrobial peptides precursor (IB-AMP)
Sementes	<i>Lupinus angustifolius</i>	P30364	L-asparaginase precursor (EC 3.5.1.1)
Sementes	<i>Lupinus arboreus</i>	P30362	L-asparaginase precursor (EC 3.5.1.1)
Sementes	<i>Macadamia integrifolia</i>	P80915	Antimicrobial peptide 1 precursor (AMP1) (MiAMP1)
Sementes	<i>Malva parviflora</i>	P83141	Antifungal protein 1 large subunit (CW-1)
Sementes	<i>Malva parviflora</i>	P83140	Antifungal protein 1 small subunit (CW-1)
Sementes	<i>Oryza sativa</i>	P27941	Alpha-amylase isozyme C2 precursor (EC 3.2.1.1)
Sementes	<i>Oryza sativa</i>	Q8H0B2	Putative UDP-arabinose 4-epimerase 3 (EC 5.1.3.5)
Sementes	<i>Oryza sativa</i>	Q42456	Aspartic proteinase oryzasin-1 precursor (EC 3.4.23.-)
Sementes	<i>Oryza sativa</i>	P27519	Chlorophyll a-b binding protein, chloroplast precursor
Sementes	<i>Oryza sativa</i>	P37891	Serine carboxypeptidase 3 precursor (EC 3.4.16.5)
Sementes	<i>Phaseolus vulgaris</i>	Q42460	Arcelin-5A precursor
Sementes	<i>Phaseolus vulgaris</i>	Q41116	Arcelin-5B precursor
Sementes	<i>Phaseolus vulgaris</i>	Q43629	Arcelin-4 precursor
Sementes	<i>Pisum sativum</i>	P62926	Albumin 1 A precursor (PA1 A)
Sementes	<i>Pisum sativum</i>	P62927	Albumin 1 B precursor (PA1 B)
Sementes	<i>Pisum sativum</i>	P62928	Albumin 1 C precursor (PA1 C)
Sementes	<i>Pisum sativum</i>	P62929	Albumin 1 D precursor (PA1 D)
Sementes	<i>Pisum sativum</i>	P62930	Albumin 1 E precursor (PA1 E)
Sementes	<i>Pisum sativum</i>	P62931	Albumin 1 F precursor (PA1 F)
Sementes	<i>Pisum sativum</i>	P08688	Albumin-2 (PA2)
Sementes	<i>Psophocarpus tetragonolobus</i>	P15465	Albumin-1 (WBA-1)
Sementes	<i>Ricinus communis</i>	P06750	Agglutinin precursor (RCA)
Sementes	<i>Sesamum indicum</i>	Q9XHP0	11S globulin seed storage protein 2 precursor
Sementes	<i>Sesamum indicum</i>	Q9XHP1	2S seed storage protein 1 precursor (beta-globulin)
Sementes	<i>Sinapis alba (Brassica hirta)</i>	P15322	Allergen Sin a 1 precursor (Sin a I)
Sementes	<i>Theobroma cacao</i>	P32765	21 kDa seed protein precursor
Sementes	<i>Triticum aestivum</i>	Q02066	Abscisic acid-inducible protein kinase (EC 2.7.11.1) (Fragment)
Sementes	<i>Triticum aestivum</i>	Q43199	Adenine phosphoribosyltransferase 1 (EC 2.4.2.7) (APRT 1)

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)