

**Tiago Santana Balbuena**

**Proteômica do desenvolvimento da semente de *Araucaria  
angustifolia***

**São Paulo  
2009**

**Tiago Santana Balbuena**

**Proteômica do desenvolvimento da semente de *Araucaria  
angustifolia***

Tese apresentada ao Instituto de  
Biotecnologia da Universidade de São  
Paulo, para a obtenção de título de  
Doutor em Ciências, na área de  
Botânica

Orientadora: Profa. Dra. Eny I. S. Floh

**São Paulo  
2009**

## FICHA CATALOGRÁFICA

B 173p Balbuena, Tiago Santana  
Proteômica do desenvolvimento da semente de  
Araucaria angustifolia / Tiago Santana Balbuena. --  
São Paulo : T. S. B., 2009.  
102 p. : il.

Tese (Doutorado) - Instituto de Biociências da  
Universidade de São Paulo. Departamento de Botânica,  
2009.

1. Coníferas 2. Eletroforese 3. Embriogênese  
4. Proteínas I. Universidade de São Paulo. Instituto de  
Biociências. Departamento de Botânica II. Título

LC QK 494.5.C75

## AUTORIZAÇÃO PARA REPRODUÇÃO

Para fins acadêmicos e científicos está autorizada a reprodução total (X) ou parcial  
( ) por processos fotocopiadores e a disponibilização total (X) ou parcial ( ) em meios  
eletrônicos.

Autor: Tiago Santana Balbuena

---

Orientadora: Profa. Dra. Eny lochevet Segal Floh

---

Candidato: Tiago Santana Balbuena

Título da Tese: Proteômica do desenvolvimento da semente de *Araucaria  
angustifolia*

A comissão julgadora dos trabalhos de defesa da tese de doutorado, em sessão pública realizada a ...../...../....., considerou

Aprovado ( )      Reprovado ( )

### COMISSÃO JULGADORA

---

Prof(a). Dr(a).

---

Prof(a). Dr(a).

---

Prof(a). Dr(a).

---

Prof(a). Dr(a).

---

Prof(a). Dr(a).

Orientadora

**AOS MEUS PAIS, ALBERTO E EUNICE,  
E À MINHA ESPOSA, MARCELLA**

## AGRADECIMENTOS

A Deus pelo amparo constante e pelas oportunidades diárias de aprendizado.

À Professora Dra. Eny I. S. Floh pela confiança, paciência, estímulo e ensinamentos transmitidos durante esses oito anos de convivência acadêmica.

Aos meus queridos pais, Alberto A. A. Balbuena e Eunice S. A. Balbuena, exemplos de caráter e dedicação, pelo carinho empregado durante minha formação pessoal e moral e pelo apoio incondicional em todas as minhas decisões.

À minha esposa, Marcella Cavanilla, pelo companheirismo, amor e compreensão demonstrados desde o início de nossa trajetória.

Aos meus tios, Modesto M. B. Franco e Fanny S. Balbuena (*in memoriam*), pelo incentivo constante em minha formação profissional.

Ao Prof. Dr. Vanildo Silveira e à Dra. Claudete Santa Catarina pelo auxílio na elaboração de experimentos e sugestões durante o desenvolvimento desse trabalho.

Aos meus amigos do Laboratório de Biologia Celular de Plantas, Amanda, Andre, Carmen, Fernanda, Julia, Leo, Leo Jo e Susie pela amizade e excelente convívio (extra) laboratorial.

Ao Prof. Dr. Andrej Shevchenko por ter me recebido em seu laboratório em minhas duas missões científicas, e pela orientação e permissão de livre uso aos recursos do Laboratório de Espectrometria de Massas do Max Planck Institute of Molecular Cell Biology and Genetics (MPI-CBG), Dresden, Alemanha.

Aos colegas do MPI-CBG pelas intensas e produtivas discussões, em especial, ao Magno Junqueira, pelo ensino e amizade construída.

Aos amigos e funcionários do Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo pelo apoio na realização deste trabalho.

À FAPESP e ao DAAD pelas bolsas concedidas.

Meu sincero obrigado...sem vocês certamente não teria chegado até aqui.

# SUMÁRIO

<i>LISTA DE ABREVIATURAS</i>	1
<i>RESUMO</i>	2
<i>ABSTRACT</i>	3
<i>CAPÍTULO 1. Introdução geral</i>	5
1. A espécie em estudo – <i>Araucaria angustifolia</i>	5
2. A proteômica em plantas	11
3. Objetivos	17
4. Referências bibliográficas	17
<i>CAPÍTULO 2. Seleção de metodologia de extração de proteínas de embriões de <i>Araucaria angustifolia</i> para utilização em sistemas de eletroforese bidimensional</i>	25
1. Resumo	25
2. Introdução	25
3. Materiais e métodos	27
4. Resultados e discussão	31
5. Considerações finais	35
6. Referências bibliográficas	36
<i>CAPÍTULO 3. Desenvolvimento de um programa de seleção de espectros para identificação de proteínas em embriões de <i>Araucaria angustifolia</i></i>	39
1. Resumo	39
2. Introdução	40
3. Materiais e métodos	42
4. Resultados e discussão	46
5. Considerações finais	55
6. Referências bibliográficas	55
<i>CAPÍTULO 4. Caracterização do proteoma de <i>Araucaria angustifolia</i> durante o desenvolvimento da semente</i>	58
1. Resumo	58
2. Introdução	58
3. Materiais e Métodos	59
4. Resultados e discussão	66
5. Considerações finais	91
6. Referências bibliográficas	92
<i>CAPÍTULO 5. Considerações gerais, implicações e perspectivas</i>	97
Referências bibliográficas	100

## LISTA DE ABREVIATURAS

2,4-D: ácido-2,4-diclorofenoxiacético  
2-DE: eletroforese bidimensional  
ABA: ácido abscísico  
AIA: ácido indol-3-acético  
APX: ascorbato peroxidase  
BAP: bezilaminopurina  
CLAE: cromatografia líquida de alta eficiência  
DTT: ditioneitol  
EDTA: ácido etilenodiaminotetracético  
GDH: glutamato desidrogenase  
GGAT: glioxalato aminotransferase  
GOGAT: glutamato sintase  
GS: glutamina sintetase  
H<sub>2</sub>O<sub>2</sub>: peróxido de nitrogênio  
IEF: focalização isoeletrica  
IPG: gradiente imobilizado de pH  
KCl: cloreto de potássio  
kDa: kilodaltons  
Kin: cinetina  
LEA: "late embryogenesis abundant"  
LTQ: quadrupolo triplo linear  
MF: massa fresca  
Mr: massa molecular relativa  
MudPIT: tecnologia de identificação multidimensional de proteínas  
PCNA: proteína antígeno nuclear de proliferação celular  
PEG: polietileno glicol  
pI: ponto isoeletrico  
PMF: "peptide mass fingerprinting"  
PMSF: fluoreto de fenilmetilsulfonil  
RBP: proteína ligadora da RuBisCo  
ROS: espécies reativas de oxigênio  
RuBisCo: ribulose-1,5-bisfosfato carboxilase/oxigenase  
SDS-PAGE: eletroforese em gel de poliacrilamida com dodecil sulfato de sódio  
TCA: ácido tricloroacético

## RESUMO

O presente trabalho teve como objetivo caracterizar o desenvolvimento da semente de *Araucaria angustifolia* através da proteômica comparativa, buscando compreender as alterações fisiológicas e metabólicas que ocorrem durante esse processo. Inicialmente, foram avaliados três diferentes metodologias de extração de proteínas. A metodologia composta por solução de extração contendo 7 M de uréia, 2 M de tiouréia, 1% de ditioneitol, 2% de Triton-100, 1 mM de fluoreto de fenilmetilsulfonil e 5  $\mu$ M de pepstatina, seguido de precipitação em 20% de ácido tricloroacético apresentou géis de maior resolução e reprodutibilidade, tendo sido escolhida como metodologia de extração proteica para o estudo das alterações no proteoma da semente de *A. angustifolia*. Uma dificuldade associada ao estudo do proteoma de espécies não sequenciadas é a baixa representatividade nos bancos de dados protéicos, resultando em identificações baseadas em homologia. Estratégias proteômicas baseadas em fracionamento em gel resultam em grandes contaminações por fragmentos de queratina. Sendo assim, foi desenvolvido um programa de remoção de espectros de baixa qualidade para utilização em proteômica baseada em homologia. As análises mostraram que o programa reduz o tempo de busca, melhora a qualidade dos alinhamentos e não resulta em perda de identificações positivas. Finalmente, utilizando as metodologias descritas, foram estudadas as alterações no proteoma durante o desenvolvimento da semente de *A. angustifolia*. Noventa e seis proteínas foram identificadas e agrupadas de acordo com sua função biológica e padrão de detecção. Os resultados obtidos permitiram o estabelecimento de marcadores protéicos no início e final do desenvolvimento embrionário. A análise das proteínas abundantes no início da embriogênese indica um maior controle no metabolismo oxidativo em relação aos estádios finais. Contrariamente, o final da embriogênese é caracterizado por um alto metabolismo de assimilação de carbono e acúmulo de proteínas de reserva. As implicações dos resultados obtidos no controle e melhoramento de sistemas de embriogênese somática na espécie também foram discutidas.

## ABSTRACT

The aim of the present work was to characterize the seed development of *Araucaria angustifolia* through proteomics in order to understand the physiological and biochemical changes during this process. For that, initially, three different protein extraction methods were evaluated. The extraction based on protein solubilization in 7 M urea, 2 M thiourea, 1% dithiothreitol, 2% Triton-100, 1 mM phenylmethylsulphonyl fluoride, 5  $\mu$ M pepstatin, followed by 20% trichloroacetic acid precipitation showed the highest gel resolution and reproductivity and, thus, was chosen to be used in the analysis of the proteome of *A. angustifolia* seeds. One aspect that hampers the proteome study of unsequenced species is the low protein representativity in databases. So, protein identification is usually carried out through homology. Strategies based on 2-DE result in high keratin contamination. In the present work a spectra filtering software was developed and evaluated for use in homology driven proteomics. The software reduced the time of search, improved alignment quality and did not result in lost of positive identifications. Finally, using the described strategies, the changes in the proteome of *A. angustifolia* seeds were studied. Ninety six proteins were identified and classified according to their biological functions and expression profiles during seed development. The identified proteins may be used as protein markers of early and late embryogenesis. Proteins involved in the control of oxidative metabolism were highly expressed during the early stages of seed development; while, carbon metabolism and storage proteins were highly expressed in late stages. Considerations on the improvement and control of somatic embryogenesis through medium manipulation and protein markers screening using data generated are also discussed.

# **CAPÍTULO 1**

## **CAPÍTULO 1. Introdução geral**

### **1. A espécie em estudo – *Araucaria angustifolia***

#### *1.1. Distribuição e descrição da área de ocorrência*

A família Araucariaceae está distribuída principalmente no Hemisfério Sul, ocorrendo na Nova Caledônia, na Nova Guiné, na Austrália, na Nova Zelândia e na América do Sul (SETOGUCHI *et al.*, 1998). Na América do Sul, as principais espécies são a *A. araucana* e a *A. angustifolia*. A *A. angustifolia* é uma espécie nativa com ocorrência natural nos estados do Paraná, Santa Catarina e Rio Grande do Sul e com manchas esparsas em São Paulo, Minas Gerais e Rio de Janeiro, ocupando uma área original de cerca de 200 mil quilômetros quadrados (AULER *et al.*, 2002; GUERRA *et al.*, 2002). Em conjunto a outras arbóreas de grande porte, constitui a Floresta Ombrófila Mista que está circunscrita a uma região de clima pluvial subtropical em altitudes que vão de 500 à 1.200 m e com área original de ocorrência aproximada de 20 milhões de hectares (GUERRA *et al.*, 2002). As médias anuais de precipitação no planalto central dos estados sulinos oscila entre 1.500 e 1.700 mm, sendo maior na borda das Serras do Mar e Geral (MACHADO, 1980). A umidade atmosférica na maior parte da região de ocorrência natural de *A. angustifolia* é bastante alta, apresentando uma média anual em torno de 80% (MACHADO, 1980). A *A. angustifolia* é uma espécie que ocorre em diferentes tipos de solos, mas de modo geral os solos dominantes da região de ocorrência natural da espécie são os siltico-argilosos, podendo ainda ser encontrada em solos oriundos de granitos, dioritos e arenitos (MACHADO, 1980).

#### *1.2. Aspectos morfológicos*

As árvores de *A. angustifolia* possuem ciclo de vida longo, atingindo maturidade após 15 anos de crescimento. As plantas são dióicas, raramente monóicas, apresentando até 50 m de altura, copa cônica, com ramos primários cilíndricos, nos indivíduos jovens e, posteriormente, corimbosa na maturidade, com ramos primários verticilados e ramos secundários alternos, caducos e agrupados no ápice dos ramos primários (GARCIA, 2002). As folhas verticiladas concentram-se na porção terminal dos galhos, medem de 15 à 80 mm

de comprimento, e de 6 à 35 mm de largura com numerosas nervuras finas, que se estendem da base ao ápice, sem nervura central (REITZ e KLEIN, 1966). As folhas podem ser oval-lanceoladas à estreito-lanceoladas, ápice agudo, base decurrente, côncavas, glabras, coriáceas e com estômatos alinhados longitudinalmente (GARCIA, 2002). A estrutura reprodutiva masculina apresenta de 10 à 15 cm de comprimento, e é constituída por escamas coriáceas com 10 à 25 anteras alongadas, presas na base ventral de cada escama. As escamas encontram-se arranjadas em forma espiral e se abrem deixando o pólen livre para ser transportado pelo vento até o estróbilo feminino. A estrutura reprodutiva feminina, ou pinha, localiza-se no ápice do ramo e é constituída por inúmeras brácteas coriáceas com óvulo, inseridas sobre um eixo central (REITZ e KLEIN, 1966).

### 1.3. Embriogênese zigótica e desenvolvimento da semente

De maneira geral, pode-se identificar duas fases distintas na embriogênese de coníferas: a primeira fase inicia-se com a formação do zigoto e termina com os cotilédones completamente formados; e a segunda fase, que compreende o estágio de maturação da semente (Figura 1).

Como em outras espécies de coníferas, a embriogênese zigótica de *A. angustifolia* é caracterizada pela presença de uma etapa de núcleos livres, quando ocorrem sucessivas divisões nucleares sem a biossíntese de parede celular (CAIRNEY e PULLMAN, 2007). Em espécies de Araucariaceae, há a formação de 32 à 64 núcleos livres antes do início da formação das paredes celulares (GUERRA *et al.*, 2008). A polaridade dos pró-embriões ocorre pela organização dos núcleos livres, constituindo três tipos celulares bem definidos: as células do suspensor, da capa e do grupo embrionário (GUERRA *et al.*, 2008).

A poliembrionia também é um fenômeno comum nas coníferas e é definida pela presença de mais de um embrião nos estádios iniciais do desenvolvimento da semente. Existem dois tipos de poliembrionia, a poliembrionia polizigótica e a poliembrionia por clivagem. Em muitas espécies de *Pinus*, ocorrem os dois tipos de poliembrionia, sendo que, por poliembrionia polizigótica, dois ou mais óvulos são fecundados em um mesmo megagametófito, produzindo separadamente pró-embriões, ou seja, cada arquegônio produz

um único pró-embrião. Posteriormente, cada pró-embrião formado sofre poliembrião por clivagem, onde se dividem por processo de clivagem em até oito pró-embriões. Em espécies do gênero *Araucaria* foi observada somente a poliembrião polizigótica (GIFFORD e FOSTER, 1989).

A fase pró-embriônica envolve todos os estádios pós fecundação, até que o pró-embrião cesse o alongamento do suspensor (HAINES e PRAKASH, 1980). Esse fenômeno é de fundamental importância, já que permite o direcionamento do embrião em direção ao megagametófito (DOGRA, 1978). Nos estádios iniciais do desenvolvimento do embrião, o suspensor representa a principal via de transporte de nutrientes para o pró-embrião (ASTARITA, 2000).

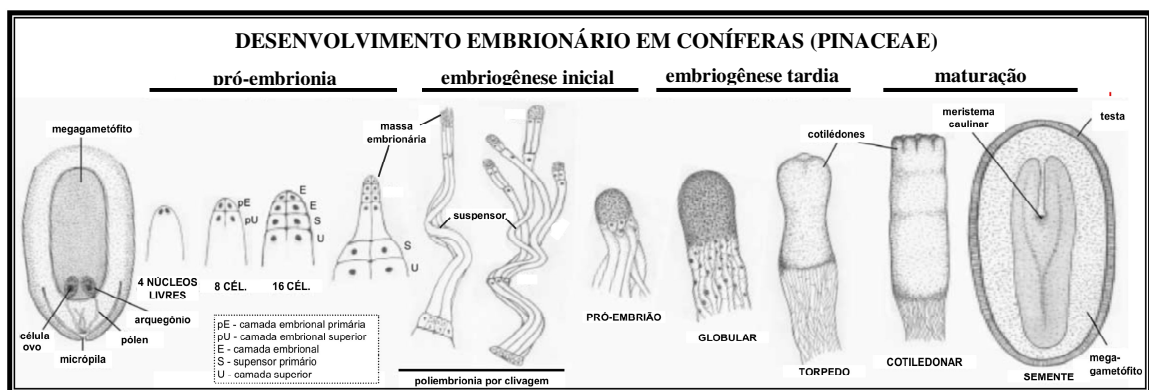


Figura 1: Aspectos do desenvolvimento do embrião de *Pinus taeda*. Adaptado de VON ARNOLD *et al.* (2002).

Após o final do estágio pró-embriônico, o suspensor degenera e ocorre um transiente aumento dos níveis de ácido indol-3-acético (AIA) nos tecidos embrionários (ASTARITA *et al.*, 2003) e deposição de materiais de reserva principalmente nos cotilédones, cujas reservas são supridas por nutrientes presentes na camada de corrosão do megagametófito (DOGRA, 1978). O armazenamento de reservas começa nos tecidos mais externos do megagametófito, possibilitando que o eixo embrionário aumente de tamanho. A formação do tecido de reserva, ou seja, do megagametófito e dos cotilédones, é caracterizada pelo aumento do número e tamanho das células, muitas das quais multinucleadas e com o citoplasma espesso, devido à presença de proteínas e um grande

número de grãos de amido (HAINES, 1983). Finalmente, quando os cotilédones completam a sua formação, tem início a fase de maturação da semente, a qual é caracterizada pela diminuição nos níveis do AIA e ácido giberélico (GA), e aumento da concentração do ácido abscísico (ABA) (ASTARITA *et al.*, 2003; SILVEIRA *et al.*, 2008).

A *A. angustifolia* é uma espécie recalcitrante, cujas sementes maduras perdem a viabilidade, rapidamente, em condições convencionais de armazenamento. Embriões maduros de *A. angustifolia* armazenam grandes quantidades de água (100 à 130% de teor hídrico) e sofrem danos irreversíveis quando submetidos à desidratação, como a degradação de membranas e desintegração nuclear (ESPINDOLA *et al.*, 1994). Análises histológicas de embriões maduros mostram a presença de vacúolos armazenadores de proteínas nos embriões maduros, mitocôndrias diferenciadas e ativas no citoplasma embrionário, além de presença de grande quantidade de água livre, características de recalcitrantes (ESPINDOLA *et al.*, 1994; PANZA *et al.*, 2002). Os embriões de *A. angustifolia* apresentam alterações estruturais e tornam-se inviáveis quando o teor hídrico atinge valores inferiores à 70% (TOMPSETT, 1984; FARRANT *et al.*, 1989).

#### 1.4. Importância econômica e exploração

As utilizações mais comuns da *A. angustifolia* sempre estiveram associadas à obtenção de madeira para tabulados, vigamentos, caixas, móveis, instrumentos musicais, artigos de esporte, ferramentas e fabricação de compensados (GUERRA *et al.*, 2002). A madeira serrada e laminada nessa espécie foi, por um longo período, um dos produtos mais importantes na exportação brasileira. Além disso, a madeira da espécie contém 58% de celulose, 29% de lignina e fibras longas de alto rendimento em celulose, tornando-a muito atraente para as indústrias de papel e celulose (GUERRA *et al.*, 2008).

Um outro produto da árvore muito utilizado por populações locais é a semente, também conhecida como pinhão (CONFORTI e LUPANO, 2007). O pinhão é uma fonte rica de amido, possuindo também proteínas, lipídios, açúcares, fibras, magnésio e cobre, em menores quantidades (CORDENUNSI *et al.*, 2004). Além disso, a coloração avermelhada da casca da semente madura indica a presença de compostos polifenólicos, benéficos ao

sistema vascular e ao combate do câncer. Estes compostos, presentes na casca são transferidos para o gametófito quando do cozimento da semente para consumo (CORDENUNSI *et al.*, 2004).

A exploração de *A. angustifolia* pelo homem intensificou-se a partir de 1934 e teve seu auge entre as décadas de 1950 e 1970, quando o metro cúbico da espécie era o produto madeireiro mais importante do Brasil (SHIMIZU e OLIVEIRA, 1981; SEITZ, 1986). Em virtude do intenso desmatamento e do consumo sem preocupações com o manejo e conservação, a produção de madeira tendeu a um esgotamento das reservas naturais. Esta situação prejudicou o abastecimento de matéria-prima para a indústria e alterou a fisionomia de extensas áreas no Sul do Brasil (GUERRA *et al.*, 2008). O intenso processo de exploração predatória colocou a espécie na lista daquelas criticamente em perigo pela IUCN (The World Conservation Union) e fez com que as reservas fossem severamente reduzidas nos estados do Paraná e Santa Catarina, e praticamente exauridas no estado de São Paulo. Atualmente, as reservas naturais dessa espécie estão limitadas a valores entre 2% a 4% da área original no país (GUERRA *et al.*, 2002).

### 1.5. Aspectos biotecnológicos

Técnicas biotecnológicas baseadas na embriogênese somática podem ser alternativas viáveis para a propagação de *A. angustifolia* (HANDRO *et al.*, 1988; GUERRA *et al.*, 2002). A embriogênese somática é um processo onde através da técnica de cultivo *in vitro*, células isoladas ou um pequeno grupo de células somáticas dão origem a embriões (TAUTORUS *et al.*, 1991), numa sequência morfogênética que se aproxima aos eventos representativos da embriogênese zigótica. Este processo pode ser dividido em quatro fases: 1) a indução em meios de cultura contendo auxinas (mais frequente) e citocininas (menos frequente); 2) multiplicação em meios de cultura contendo auxinas em baixa concentração; 3) maturação em presença de ABA e/ou contendo agentes osmóticos e; 4) germinação em meios de cultura isentos de fitorreguladores (TAUTORUS *et al.*, 1991).

Comparativamente às demais técnicas de micropropagação, a embriogênese somática apresenta as seguintes vantagens: a) permite a obtenção de uma grande

quantidade de propágulos (embriões somáticos); b) permite um alto grau de automatização, reduzindo os custos por unidade produzida; c) os embriões somáticos podem ser produzidos de forma sincronizada, com alto grau de uniformização e pureza genética; d) pode ser utilizada como uma ferramenta integrada a programas de melhoramento genético florestal, em especial, quando associada à técnica de criopreservação; e) pode ser utilizada para a produção de metabólitos secundários (MERKLE e DEAN, 2000).

Em *A. angustifolia*, a embriogênese somática compreende uma sequência de etapas que incluem a indução e a multiplicação de complexos celulares suspensor-embriionários, que são os precursores dos embriões somáticos (GUERRA *et al.*, 2008). A estratégia empregada no estabelecimento da embriogênese somática em *A. angustifolia* consiste em: um ciclo de indução e multiplicação de culturas embriogênicas, originadas a partir do embrião zigótico imaturo pela adição de auxinas e citocininas (Figura 2A); e um ciclo de maturação, promovido por alterações no balanço de reguladores de crescimento e fontes de carbono adicionados ao meio de cultura (Figura 2B). O resultado destas etapas leva à formação de embriões somáticos maduros (STEINER, 2005; GUERRA *et al.*, 2008).

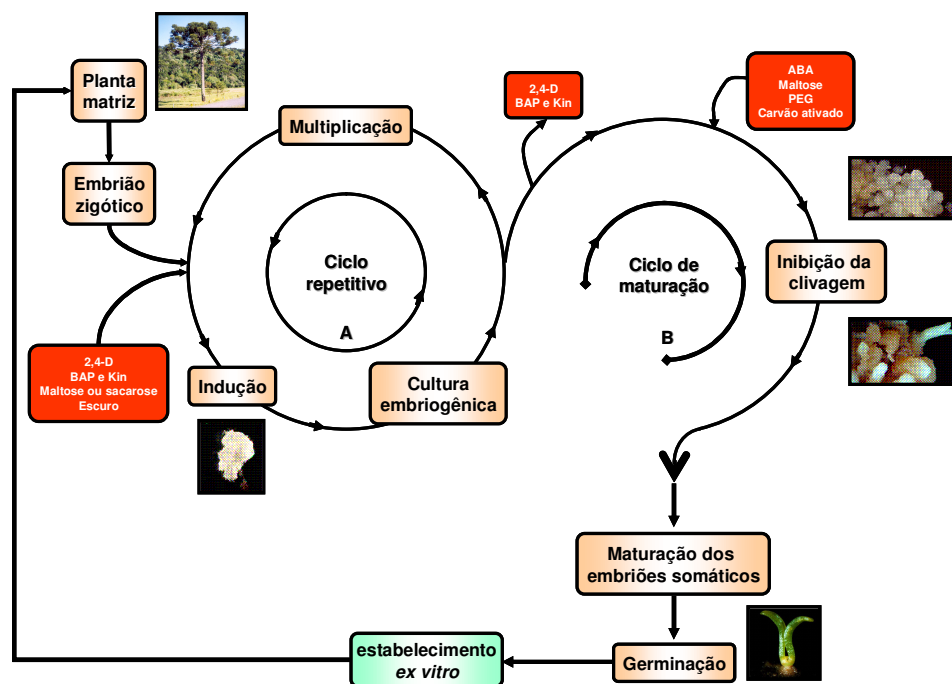


Figura 2: Esquema da modulação de culturas embriogênicas de *A. angustifolia*, por meio de ciclo repetitivo de multiplicação e ciclo de maturação. Adaptado de STEINER (2005).

As condições básicas e os fatores que afetam a indução, manutenção e multiplicação de culturas embriogênicas de *A. angustifolia* já foram estabelecidas em vários estudos (DOS SANTOS *et al.*, 2002; GUERRA *et al.*, 2002; SILVEIRA *et al.*, 2002; STEINER *et al.*, 2005). Adicionalmente, sabe-se que as taxas de multiplicação e rotas morfogenéticas podem ser influenciadas pela adição de ABA, agentes osmóticos, carvão ativado, poliaminas e óxido nítrico nos meios de cultura (ASTARITA, 2000; SILVEIRA *et al.*, 2002; STEINER *et al.*, 2005; STEINER *et al.*, 2007). Entretanto, os eventos tardios da embriogênese somática nesta espécie ainda não estão suficientemente elucidados e frequentemente resultam na produção de embriões somáticos de baixa qualidade. Esta condição é limitante para a utilização desta estratégia em programas de conservação ou como ferramenta no melhoramento genético da espécie. Além disso, em coníferas, métodos de propagação por estaquias são de difícil controle, devido à relação inversa entre a idade das plantas doadoras e a capacidade de enraizamento das estacas (STASOLLA e YEUNG, 2003). Portanto, embora as sementes de *A. angustifolia* sejam recalcitrantes e, conseqüentemente, de difícil armazenamento (ESPINDOLA *et al.*, 1994; PANZA *et al.*, 2002), o reflorestamento ainda é realizado por plantio e dispersão de sementes (GUERRA *et al.*, 2002).

## **2. A proteômica em plantas**

Nas últimas décadas, os códigos genéticos de diversos organismos têm sido sequenciados, representando um importante avanço do conhecimento na área genômica. No contexto das “ômicas”, grande ênfase tem-se dado para a genômica funcional, que visa a identificação da sequência e função dos genes e proteínas através de estudos do transcriptoma e proteoma dos organismos (DIAS *et al.*, 2007).

As avaliações da expressão gênica em nível do transcriptoma podem fornecer informações importantes sobre a carga genética transcrita de um organismo em um determinado momento; entretanto, ela não reflete diretamente a expressão das proteínas (CHEN e HARMON, 2006). Após a transcrição, podem ocorrer modificações pós-transcricionais e pós-traducionais, com alteração da conformação espacial de proteínas, e

que resultam em classes protéicas que bioquímica e estruturalmente podem desempenhar diferentes funções metabólicas (CHEN e HARMON, 2006).

A proteômica é o estudo sistemático de um proteoma, ou conjunto de proteínas expressas pelo genoma de um organismo (WASINGER *et al.*, 1995), que permite avaliações quantitativas e qualitativas de proteínas que atuam no metabolismo celular (PARK, 2004; CHEN e HARMON, 2006). Conseqüentemente, a identificação de proteínas expressas durante o desenvolvimento vegetal permite a caracterização de diferentes eventos fisiológicos que ocorrem nas células, tecidos e órgãos das plantas (KORMUTAK *et al.*, 2006). Em geral, a proteômica possui muitas aplicações como no: a) estudo da expressão diferencial de proteínas; b) estudo de modificações pós-traducionais; c) estudos de interação proteína-proteína; d) estudo estrutural; e e) estudo da função das proteínas (DIAS *et al.*, 2007).

Nos últimos anos vários trabalhos têm focado a caracterização da dinâmica de proteínas ao longo do desenvolvimento vegetal (CÁNOVAS *et al.*, 2004; PARK, 2004; HOCHHOLDINGER *et al.*, 2006). A dinâmica da proteínas em um sistema vivo é influenciada por diversos fatores, internos e externos, que determinam modificações estruturais e a conformação das proteínas. Neste sentido, o estudo e a caracterização de mapas proteômicos apresentam-se como importantes ferramentas complementares aos estudos anatômicos, fisiológicos e de genômica. A análise do proteoma permite examinar simultaneamente alterações e classificar padrões temporais de acúmulo de proteínas que ocorrem durante o desenvolvimento da semente, permitindo a identificação de proteínas marcadoras estágio específicas (DIAS *et al.*, 2007).

Com os recentes avanços nas técnicas de 2-DE, como o aumento da reprodutibilidade pelo uso de gradientes imobilizados de pH, a visualização de imagens e aumento de sensibilidade de detecção; e nas análises por espectrometria de massas, como o aumento de sensibilidade, resolução e velocidade de aquisição de dados, a proteômica tem se tornado uma estratégia alternativa no estabelecimento de marcadores bioquímicos e uma das mais utilizadas para análises metabólicas e de expressão gênica global (SHEORAN *et al.*, 2005; PAWLOWSKI, 2007).

Devido à grande diversidade de proteínas acumuladas nos compartimentos celulares, as diferentes metodologias para o estudo do proteoma não são simples (GORG *et al.*, 2004). As duas metodologias mais utilizadas atualmente envolvem: a) a eletroforese bidimensional (2-DE) combinada à espectrometria de massas; e b) análises em “shotgun”, como a tecnologia de identificação protéica multidimensional (MudPIT) (AEBERSOLD e MANN, 2003).

A 2-DE foi desenvolvida no início da década de 1970 e ainda é uma das técnicas mais utilizadas para o fracionamento protéico (CHEN e HARMON, 2006). A primeira etapa do processo consiste na separação das proteínas de acordo com seus pontos isoelétricos (pIs) e, posteriormente, de acordo com suas massas moleculares relativas (Mr). Esta situação permite o fracionamento de milhares de proteínas, além da obtenção de informações físico-químicas das mesmas. Posteriormente, os géis são corados, digitalizados, e as imagens produzidas analisadas através de programas de análise de imagens (Figura 3A) (CHEN e HARMON, 2006). Além da 2-DE, outras técnicas de fracionamento protéico também têm sido utilizadas com sucesso, como a MudPIT em que os peptídeos, previamente digeridos em uma solução protéica, são fracionados em uma coluna de troca catiônica e, posteriormente, em uma fase reversa (Figura 3B) (AGRAWAL *et al.*, 2008). Alternativamente, a eletroforese unidimensional também pode ser utilizada para fracionamento de misturas protéicas complexas, permitindo análises globais do proteoma em estudo, já que permite a melhor detecção de proteínas pouco abundantes (Figura 3C) (GAO *et al.*, 2008).

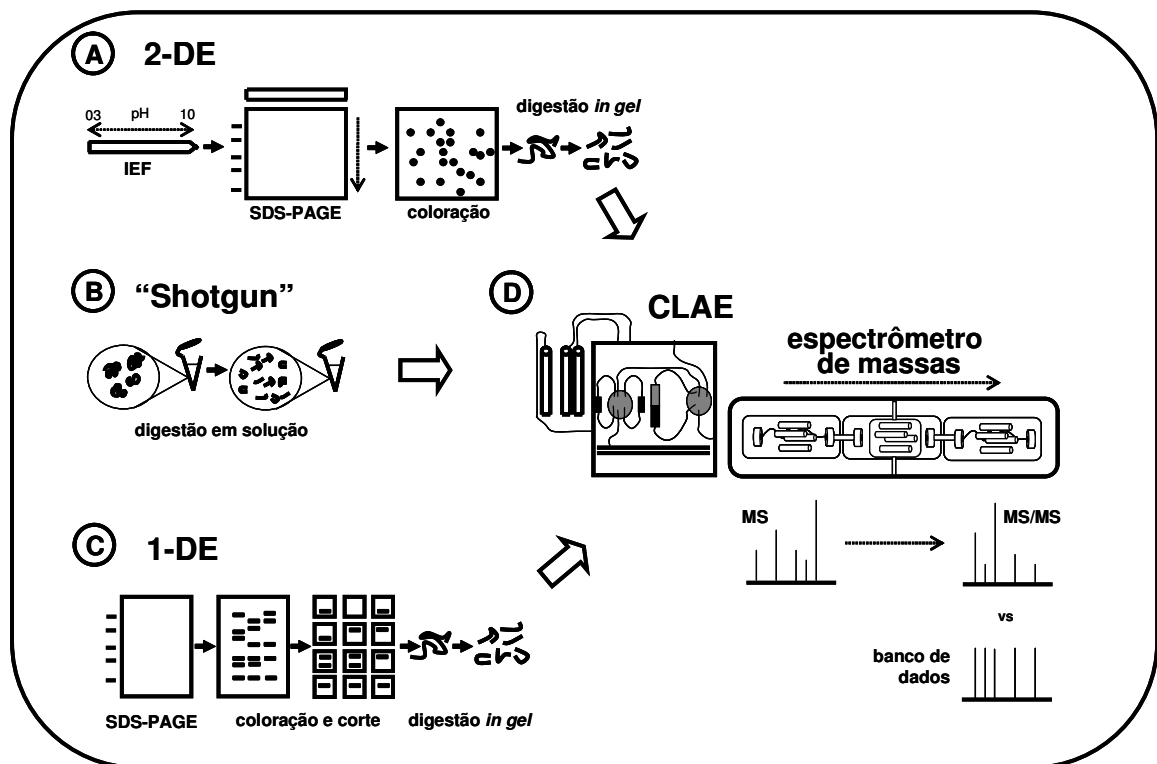


Figura 3: Metodologias de preparo de amostras protéicas para identificação por espectrometria de massas. **A.** Fracionamento protéico por eletroforese bidimensional (2-DE), seguido de excisão e digestão de proteínas *in gel*. **B.** Obtenção de peptídeos trípticos diretamente da solução protéica. **C.** Fracionamento protéico por SDS-PAGE (1-DE), seguido de corte e digestão *in gel*. **D.** Fracionamento por CLAE.

Atualmente, independente da estratégia utilizada, o procedimento padrão para identificação de proteínas inicia-se com a digestão enzimática (OLSEN *et al.*, 2004). Posteriormente, os peptídeos são fragmentados em um espectrômetro de massas, gerando espectros do tipo MS/MS, os quais são utilizados nas buscas em bancos de dados. Os programas utilizados na busca de dados baseiam-se na comparação entre o padrão de fragmentação teórico de uma proteína e o espectro de fragmentação obtido experimentalmente. A estratégia de identificação por MS/MS, que permite a determinação precisa da sequência experimental de aminoácidos, é mais precisa do que a identificação de proteínas baseada somente na comparação entre as massas dos fragmentos trípticos, chamada de "peptide mass fingerprinting" (PMF) (CHEN e HARMON, 2006). Alternativamente, os espectros do tipo MS/MS também permitem buscas por similaridade de sequências entre organismos próximos, através da reconstrução da sequência de

aminoácidos manual ou automaticamente. Essa estratégia é muito interessante quando o organismo estudado possui baixa representatividade nos bancos de dados (CHEN e HARMON, 2006).

Os primeiros estudos envolvendo variações no proteoma durante o desenvolvimento e germinação de sementes foi realizado em *Arabidopsis thaliana* e *Medicago truncatula* (GALLARDO *et al.*, 2001; GALLARDO *et al.*, 2003). Trabalhos semelhantes têm sido realizados em outras espécies modelos e naquelas que apresentam valor econômico, como *Glycine max*, *Oryza sativa*, *Lycopersicon esculentum*, *Triticum aestivum* e *Hordeum vulgare* (Tabela 1). Entretanto, poucos estudos foram realizados com espécies arbóreas. Destacam-se os trabalhos realizados por SALLANDROUZE *et al.* (2002), que estudaram a dinâmica da expressão protéica durante as etapas finais do desenvolvimento embrionário, avaliando a mobilização de proteínas de reserva e a expressão de proteínas “late embryogenesis abundant” (LEA) na conífera *Cupressus sempervirens*; e o trabalho de PAWLOWSKI (2007) que realizou um estudo do acúmulo diferencial de proteínas durante a germinação de sementes dormentes *Fagus sylvatica*.

Recentemente, DOS SANTOS *et al.* (2006) realizaram análises qualitativas e quantitativas dos mapas proteômicos obtidos por 2-DE de diferentes estádios do desenvolvimento da semente de *A. angustifolia*. Embora não tenham sido utilizadas técnicas de espectrometria de massas para a identificação de proteínas, a presença de arabinogalactanos e proteínas do tipo quitinases foi detectada por coloração imunológica e utilização de substratos específicos em análises por eletroforese unidimensional.

SILVEIRA *et al.* (2008) realizaram um estudo comparativo entre o proteoma dos estádios torpedo e maduro em *A. angustifolia*. Foram identificadas alterações no padrão eletroforético entre os diferentes materiais estudados. Entretanto, não foram realizadas análises quantitativas dessas variações. Adicionalmente, os pesquisadores utilizaram a estratégia de identificação de proteínas por PMF para a identificar proteínas observadas exclusivamente no estágio maduro, tendo sido identificadas a enzima fosfoinositídeo fosfatase, uma proteína LEA e uma proteína de reserva do tipo vicilina.

Tabela 1. Espécies cujas sementes foram estudadas utilizando abordagem proteômica.

Família	Espécie	Assunto	Referência
Brassicaceae	<i>Arabidopsis thaliana</i>	Proteômica comparativa durante germinação	(GALLARDO <i>et al.</i> , 2001)
		Proteômica comparativa entre diferentes órgãos e sementes	(GIAVALISCO <i>et al.</i> , 2005)
		Estudo do metabolismo e processamento de proteínas de reserva	(HIGASHI <i>et al.</i> , 2006)
		Mecanismos de controle da dormência em sementes	(CHIBANI <i>et al.</i> , 2006)
		Efeito do ácido salicílico nos mecanismos de defesa durante a germinação	(RAJJOU <i>et al.</i> , 2006)
		Proteômica durante o armazenamento de sementes	(RAJJOU <i>et al.</i> , 2008)
Fabaceae	<i>Medicago truncatula</i>	Proteômica durante a maturação	(GALLARDO <i>et al.</i> , 2003)
		Caracterização do proteoma nuclear durante maturação	(REPETTO <i>et al.</i> , 2008)
	<i>Glycine max</i>	Proteômica durante o desenvolvimento da semente e em relação a sementes oleaginosas	(AGRAWAL <i>et al.</i> , 2008)
		Identificação de proteínas alergênicas em sementes	(MAGNI <i>et al.</i> , 2005)
		Proteômica durante o desenvolvimento da semente	(HAJDUCH <i>et al.</i> , 2005)
Poaceae	<i>Oryza sativa</i>	Proteômica comparativa entre diferentes genótipos e respectivos alergênicos	(XU <i>et al.</i> , 2007)
		Identificação de proteínas alergênicas em sementes	(GOGGIN <i>et al.</i> , 2008)
		Efeito da temperatura no acúmulo diferencial de proteínas durante o desenvolvimento da semente	(LIN <i>et al.</i> , 2005)
Poaceae	<i>Hordeum vulgare</i>	Proteômica durante a germinação	(YANG <i>et al.</i> , 2007)
		Mecanismos de regulação oxidativa durante a germinação	(YANO e KURODA, 2006)
		Proteômica durante o desenvolvimento da semente	(FINNIE <i>et al.</i> , 2002)
	<i>Triticum aestivum</i>	Efeito da transgenia e de fatores ambientais no proteoma da semente	(FINNIE <i>et al.</i> , 2004)
		Proteômica comparativa durante a germinação	(BONSAGER <i>et al.</i> , 2007)
		Caracterização de proteínas de reserva associadas a cromossomos telocêntricos no endosperma	(ISLAM <i>et al.</i> , 2002)
Chenopodiaceae	<i>Beta vulgaris</i>	Proteômica comparativa entre tecidos da semente durante a maturação	(CATUSSE <i>et al.</i> , 2008)
Brassicaceae	<i>Brasica napus</i>	Proteômica durante o desenvolvimento da semente	(HAJDUCH <i>et al.</i> , 2006)
Asteraceae	<i>Helianthus annuus</i>	Proteômica comparativa entre diferentes genótipos	(HAJDUCH <i>et al.</i> , 2007)
Euphorbiaceae	<i>Ricinus communis</i>	Proteômica do retículo endoplasmático durante o desenvolvimento da semente e germinação	(MALTMAN <i>et al.</i> , 2007)
Cupressaceae	<i>Cupressus sempervirens</i>	Proteômica durante o desenvolvimento da semente	(SALLANDROUZE <i>et al.</i> , 2002)
Fagaceae	<i>Fagus sylvatica</i>	Efeito do ácido absísico e de giberelinas na quebra da dormência	(PAWLOWSKI, 2007)
Araucariaceae	<i>Araucaria angustifolia</i>	Proteômica comparativa entre o início e final do desenvolvimento embrionário	(SILVEIRA <i>et al.</i> , 2008)
		Proteômica durante o desenvolvimento da semente	(DOS SANTOS <i>et al.</i> , 2006)

Embora tenham sido obtidos avanços na proteômica da semente de *A. angustifolia*, não existem investigações em larga escala sobre as alterações do proteoma de *A. angustifolia* durante o desenvolvimento da semente e o processo de embriogênese zigótica.

### **3. Objetivos**

O presente trabalho teve como objetivo a caracterização do desenvolvimento da semente de *A. angustifolia* através da proteômica comparativa. Foram propostos os seguintes objetivos específicos:

- Estabelecimento de uma metodologia de extração de proteínas de *A. angustifolia* para uso em sistemas de 2-DE;
- Desenvolvimento de programa computacional para aumento da eficiência de identificações protéicas por similaridade de sequências nesta espécie, cujo genoma ainda não é sequenciado;
- Estudo das alterações no proteoma durante o desenvolvimento da semente.

### **4. Referências bibliográficas**

- Aebersold, R.,M. Mann. Mass spectrometry-based proteomics. Nature, v.422, p.198-207. 2003.
- Agrawal, G. K., M. Hajduch, K. Graham,J. J. Thelen. In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. Plant Physiology, v.148, p.504-518. 2008.
- Astarita, L. V. Aspectos bioquímicos e fisiológicos da embriogênese e de culturas celulares embriogênicas de *Araucaria angustifolia* (Bert.) O. Ktze. Instituto de Biociências, Universidade de São Paulo, São Paulo, 2000. 112 p.
- Astarita, L. V., E. I. S. Floh,W. Handro. Changes in IAA, tryptophan and activity of soluble peroxidase associated with zygotic embryogenesis in *Araucaria angustifolia* (Brazilian pine). Plant Growth Regulation, v.39, p.113-118. 2003.
- Auler, N. M. F., M. S. Reis, M. P. Guerra,R. O. Nodari. The genetics and conservation of *Araucaria angustifolia*: I. genetic structure and diversity of natural populations by means of non-adaptive variation in the state of Santa Catarina, Brazil. Genetics and Molecular Biology, v.25, p.329-338. 2002.
- Bonsager, B. C., C. Finnie, P. Roepstorff,B. Svensson. Spatio-temporal changes in germination and radical elongation of barley seeds tracked by proteome analysis of

- dissected embryo, aleurone layer, and endosperm tissues. Proteomics, v.7, p.4528-40. 2007.
- Cairney, J., G. S. Pullman. The cellular and molecular biology of conifer embryogenesis. New Phytologist, v.176, p.511-536. 2007.
- Canovas, F. M., F. R. Canton, A. Garcia-Gutierrez, F. Gallardo, R. Crespillo. Molecular physiology of glutamine and glutamate biosynthesis in developing seedlings of conifers. Physiologia Plantarum, v.103, p.287-294. 1998.
- Catusse, J., J. M. Strub, C. Job, A. V. Dorsselaer, D. Job. Proteome-wide characterization of sugarbeet seed vigor and its tissue specific expression. Proceedings of the National Academy of Sciences of the United States of America, v.105, p.10262-10267. 2008.
- Chen, S. X., A. C. Harmon. Advances in plant proteomics. Proteomics, v.6, p.5504-5516. 2006.
- Chibani, K., S. Ali-Rachedi, C. Job, D. Job, M. Jullien, P. Grappin. Proteomic analysis of seed dormancy in *Arabidopsis*. Plant Physiology, v.142, p.1493-1510. 2006.
- Conforti, P. A., C. E. Lupano. Starch characterization of *Araucaria angustifolia* and *Araucaria araucana* seeds. Starch-Starke, v.59, p.284-289. 2007.
- Cordenunsi, B. R., E. W. De Menezes, M. I. Genovese, C. Colli, A. G. A. De Souza, F. M. Lajolo. Chemical composition and glycemic index of Brazilian pine (*Araucaria angustifolia*) seeds. Journal of Agricultural and Food Chemistry, v.52, p.3412-3416. 2004.
- Dias, L. L. C., E. I. S. Floh, C. Santa-Catarina, V. Silveira. Proteômica comparativa aplicada à cultura de tecidos de plantas. Revista Brasileira de Horticultura Ornamental, v.13S, p.2002-2008. 2007.
- Dogra, P. D. Morphology, development and nomenclature of conifer embryo. Phytomorphology, v.28, p.307-332. 1978.
- Dos Santos, A. L. W., V. Silveira, N. Steiner, M. Vidor, M. P. Guerra. Somatic embryogenesis in Parana Pine (*Araucaria angustifolia* (Bert.) O. Kuntze). Brazilian Archives of Biology and Technology, v.45, p.97-106. 2002.
- Dos Santos, A. L. W., N. Wietholter, N. E. El Gueddari, B. M. Moerschbacher. Protein expression during seed during in *Araucaria angustifolia*: transient accumulation of class IV chitinases and arabinogalactan proteins. Physiologia Plantarum, v.127, p.138-148. 2006.
- Espindola, L. E., M. Noin, F. Corbineau, D. Côme. Cellular and metabolic damage induced by desiccation in recalcitrant *Araucaria angustifolia* embryos. Seed Science Research, v.4, p.193-201. 1994.
- Farrant, J. M., N. W. Pammenter, P. Berjak. Germination-associated events and the desiccation sensitivity of recalcitrant seeds - a study on 3 unrelated species. Planta, v.178, p.189-198. 1989.

- Finnie, C., S. Melchior, P. Roepstorff, B. Svensson. Proteome analysis of grain filling and seed maturation in barley. Plant Physiology, v.129, p.1308-1319. 2002.
- Finnie, C., T. Steenholdt, O. R. Noguera, S. Knudsen, J. Larsen, H. Brinch-Pedersen, P. B. Holm, O. Olsen, B. Svensson. Environmental and transgene expression effects on the barley seed proteome. Phytochemistry, v.65, p.1619-1627. 2004.
- Gallardo, K., C. Job, S. P. C. Groot, M. Puype, H. Demol, J. Vandekerckhove, D. Job. Proteomic analysis of *Arabidopsis* seed germination and priming. Plant Physiology, v.126, p.835-848. 2001.
- Gallardo, K., C. Le Signor, J. Vandekerckhove, R. D. Thompson, J. Burstin. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiology, v.133, p.664-682. 2003.
- Gao, B.-B., L. Stuart, E. P. Feener. Label-free quantitative analysis of one-dimensional PAGE LC/MS/MS proteome: application on angiotensin II-stimulated smooth muscle cells secretome. Molecular & Cellular Proteomics, v.7, p.2399-2409. 2008.
- Garcia, R. J. F. Araucareaceae. In: M. G. L. Wanderley, G. J. Shepherd, A. M. Giulietti (Ed.). Flora Fanerogâmica do Estado de São Paulo. São Paulo: Hucitec, v.2, 2002. Araucareaceae, p.1-4.
- Giavalisco, P., E. Nordhoff, T. Kreitler, K. D. Kloppel, H. Lehrach, J. Klose, J. Gobom. Proteome analysis of *Arabidopsis thaliana* by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry. Proteomics, v.5, p.1902-1913. 2005.
- Gifford, E. M., A. S. Foster. Morphology and Evolution of Vascular Plants. New York: W. H. Freeman Co. 1989.
- Goggin, D. E., G. Mir, W. B. Smith, M. Stuckey, P. A. C. Smith. Proteomic analysis of lupin seed proteins to identify conglutin beta as an allergen, Lup an 1. Journal of Agricultural and Food Chemistry, v.56, p.6370-6377. 2008.
- Gorg, A., O. Drews, A. Kopf, C. Luck, G. Reil, B. Schroter, W. Weiss. Today's 2-D electrophoresis technology. Molecular & Cellular Proteomics, v.3, p.S283-S283. 2004.
- Guerra, M. P., V. Silveira, M. S. Dos Reis, L. Schneider. Exploração, manejo e conservação da araucária (*Araucaria angustifolia*). In: L. L. Simões, Lino, C. F. (Ed.). Sustentável Mata Atlântica: a exploração se seus recursos florestais. São Paulo: Editora Senac, 2002. Exploração, manejo e conservação da araucária (*Araucaria angustifolia*), p.85-101.
- Guerra, M. P., N. Steiner, A. Mantovani, R. O. Nodari, M. S. Dos Reis, K. L. Dos Santos. Araucária. In: R. L. Barbieri, E. R. T. Stumpf (Ed.). Origem e Evolução de Plantas Cultivadas. Brasília: Embrapa Informação Tecnológica, 2008. Araucária, p.149-184.
- Haines, R. J. Seed development in *Araucaria juss*. Australian Journal of Botany, v.31, p.255-267. 1983.

- Haines, R. J., N. Prakash. Proembryo development and suspensor elongation in *Araucaria juss.* Australian Journal of Botany, v.28, p.511-522. 1980.
- Hajduch, M., J. E. Casteel, K. E. Hurrelmeyer, Z. Song, G. K. Agrawal, J. J. Thelen. Proteomic analysis of seed filling in *Brassica napus*. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiology, v.141, p.1159-1159. 2006.
- Hajduch, M., J. E. Casteel, S. X. Tang, L. B. Hearne, S. Knapp, J. J. Thelen. Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. Journal of Proteome Research, v.6, p.3232-3241. 2007.
- Hajduch, M., A. Ganapathy, J. W. Stein, J. J. Thelen. A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiology, v.137, p.1397-1419. 2005.
- Handro, W., E. I. S. Floh, C. M. Ferreira, M. P. Guerra. Tissue, cell-culture and micropropagation of *Mandevilla velutina*, a natural source of a bradykinin antagonist. Plant Cell Reports, v.7, p.564-566. 1988.
- Higashi, Y., M. Y. Hirai, T. Fujiwara, S. Naito, M. Noji, K. Saito. Proteomic and transcriptomic analysis of *Arabidopsis* seeds: molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. Plant Journal, v.48, p.557-571. 2006.
- Hochholdinger, F., M. Sauer, D. Dembinsky, N. Hoecker, N. Muthreich, M. Saleem, Y. Liu. Proteomic dissection of plant development. Proteomics, v.6, p.4076-4083. 2006.
- Islam, N., S. H. Woo, H. Tsujimoto, H. Kawasaki, H. Hirano. Proteome approaches to characterize seed storage proteins related to ditelocentric chromosomes in common wheat (*Triticum aestivum* L.). Proteomics, v.2, p.1146-1155. 2002.
- Kormutak, A., T. Salaj, B. Vookova. Storage protein dynamics in zygotic and somatic embryos of white fir. Biologia (Bratislava), v.61, p.479-486. 2006.
- Lin, S. K., M. C. Chang, Y. G. Tsai, H. S. Lur. Proteomic analysis of the expression of proteins related to rice quality during caryopsis development and the effect of high temperature on expression. Proteomics, v.5, p.2140-2156. 2005.
- Machado, S. A. Distribuição natural da *Araucaria angustifolia* (Bert.) O. Ktze. In: N. I. S. Carneiro (Ed.). Conferência IUFRO: Problemas Forestais do Gênero Araucaria. Curitiba: Fundação de Pesquisas Florestais do Paraná, 1980. Distribuição natural da *Araucaria angustifolia* (Bert.) O. Ktze., p.4-9.
- Magni, C., C. Ballabio, P. Restani, E. Sironi, A. Scarafoni, C. Poiesi, M. Duranti. Two-dimensional electrophoresis and western-blotting analyses with anti Ara h 3 basic subunit IgG evidence the cross-reacting polypeptides of *Arachis hypogaea*, *Glycine max*, and

- Lupinus albus* seed proteomes. Journal of Agricultural and Food Chemistry, v.53, p.2275-2281. 2005.
- Maltman, D. J., S. M. Gadd, W. J. Simon, A. R. Slabas. Differential proteomic analysis of the endoplasmic reticulum from developing and germinating seeds of castor (*Ricinus communis*) identifies seed protein precursors as significant components of the endoplasmic reticulum. Proteomics, v.7, p.1513-1528. 2007.
- Merkle, S. A., J. F. D. Dean. Forest tree biotechnology. Current Opinion in Biotechnology, v.11, p.298-302. 2000.
- Olsen, J. V., S. E. Ong, M. Mann. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Molecular & Cellular Proteomics, v.3, p.608-614. 2004.
- Panza, V., V. Lainez, H. Maroder, I. Prego, S. Maldonado. Storage reserves and cellular water in nature seeds of *Araucaria angustifolia*. Botanical Journal of the Linnean Society, v.140, p.273-281. 2002.
- Park, O. K. Proteomics studies in plants. Journal of Biochemistry and Molecular Biology, v.37, p.133-138. 2004.
- Pawlowski, T. A. Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: Influence of abscisic and gibberellic acids. Proteomics, v.7, p.2246-2257. 2007.
- Rajjou, L., M. Belghazi, R. Huguet, C. Robin, A. Moreau, C. Job, D. Job. Proteomic investigation of the effect of salicylic acid on *Arabidopsis* seed germination and establishment of early defense mechanisms. Plant Physiology, v.141, p.910-923. 2006.
- Rajjou, L., Y. Lovigny, S. P. C. Groot, M. Belghazi, C. Job, D. Job. Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. Plant Physiology, v.148, p.620-641. 2008.
- Reitz, R., R. M. Klein. Araucariaceae. Itajaí. 1966. 62 p. (Flora Ilustrada Catarinense-Herbário Barbosa Rodriguez).
- Repetto, O., H. Rogniaux, C. Firnhaber, H. Zuber, H. Kuster, C. Larre, R. Thompson, K. Gallardo. Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. Plant Journal, v.56, p.398-410. 2008.
- Sallandrouze, A., M. Faurobert, M. El Maataoui. Characterization of the developmental stages of cypress zygotic embryos by two-dimensional electrophoresis and by cytochemistry. Physiologia Plantarum, v.114, p.608-618. 2002.
- Seitz, R. Crow development of *Araucaria angustifolia* in its natural environment during sixty years. In: T. Fujimori, D. Whitehead (Ed.). Crow and Canopy structure in Relation to Productivity. Ibaraki: Forestry and Forest Products Research Institute, 1986. Crow development of *Araucaria angustifolia* in its natural environment during sixty years, p.129-145.

- Setoguchi, H., T. A. Osawa, J. C. Pintaud, T. Jaffre, J. M. Veillon. Phylogenetic relationships within Araucariaceae based on rbcL gene sequences. American Journal of Botany, v.85, p.1507-1516. 1998.
- Sheoran, I. S., D. J. H. Olson, A. R. S. Ross, V. K. Sawhney. Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics, v.5, p.3752-3764. 2005.
- Shimizu, J. Y., Y. M. M. Oliveira. Distribuição, variação e usos dos recursos genéticos da Araucaria no sul do Brasil. Curitiba: Embrapa-URPFCS. 1981. 9 p.
- Silveira, V., C. Santa-Catarina, T. S. Balbuena, F. M. S. Moraes, C. A. O. Ricart, M. V. Sousa, M. P. Guerra, W. Handro, E. I. S. Floh. Endogenous abscisic acid and protein contents during seed development of *Araucaria angustifolia*. Biologia Plantarum, v.52, p.101-104. 2008.
- Silveira, V., N. Steiner, A. L. W. Dos Santos, R. O. Nodari, M. P. Guerra. Biotechnology tools in *Araucaria angustifolia* conservation and improvement: inductive factors affecting somatic embryogenesis. Crop Breeding and Applied Biotechnology, v.2, p.463-470. 2002.
- Stasolla, C., E. C. Yeung. Recent advances in conifer somatic embryogenesis: improving somatic embryo quality. Plant Cell Tissue and Organ Culture, v.74, p.15-35. 2003.
- Steiner, N. Parâmetros fisiológicos e bioquímicos durante a embriogênese zigótica e somática em Araucaria angustifolia (Bert.) O. Kuntze. (Dissertação de Mestrado). Genética e Melhoramento Vegetal, Universidade Federal de Santa Catarina, Santa Catarina, 2005. 129 p.
- Steiner, N., C. Santa-Catarina, V. Silveira, E. I. S. Floh, M. P. Guerra. Polyamine effects on growth and endogenous hormones levels in *Araucaria angustifolia* embryogenic cultures. Plant Cell, Tissue and Organ Culture, v.89, p.55-62. 2007.
- Steiner, N., F. D. Vieira, S. Maldonado, M. P. Guerra. Effect of carbon source on morphology and histodifferentiation of *Araucaria angustifolia* embryogenic cultures. Brazilian Archives of Biology and Technology, v.48, p.895-903. 2005.
- Taurus, T. E., L. C. Fowke, D. I. Dunstan. Somatic embryogenesis in conifers. Canadian Journal of Botany, v.69, p.1873-1899. 1991.
- Tompsett, P. B. Desiccation studies in relation to the storage of *Araucaria* seed. Annals of Applied Biology, v.105, p.581-586. 1984.
- Von Arnold, S., I. Sabala, P. Bozkov, J. Dyachok, L. Filanova. Developmental pathways of somatic embryogenesis. Plant Cell Tissue and Organ Culture, v.69, p.233-249. 2002.
- Wasinger, V. C., S. J. Cordwell, A. Cerpapoljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, I. Humpherysmith. Progress with gene-product mapping of the mollicutes - *Mycoplasma genitalium*. Electrophoresis, v.16, p.1090-1094. 1995.
- Xu, C., T. J. Caperna, W. M. Garrett, P. Cregan, H. Bae, D. L. Luthria, S. Natarajan. Proteomic analysis of the distribution of the major seed allergen in wild, landrace, ancestral, and

- modern soybean genotypes. Journal of the Science of Food Agriculture, v.87, p.2511-2518. 2007.
- Yang, P., X. Li, X. Wang, H. Chen, F. Chen, S. Shen. Proteomic analysis of rice (*Oryza sativa*) seeds during germination. Proteomics, v.7, p.3358-3368. 2007.
- Yano, H., M. Kuroda. Disulfide proteome yields a detailed understanding of redox regulations: a model study of thioredoxin-linked reactions in seed germination. Proteomics, v.6, p.294-300. 2006.

## **CAPÍTULO 2**

## **CAPÍTULO 2. Seleção de metodologia de extração de proteínas de embriões de *Araucaria angustifolia* para utilização em sistemas de eletroforese bidimensional**

### **1. Resumo**

O preparo de amostras protéicas é uma etapa crucial na eletroforese bidimensional (2-DE). Tecidos vegetais são ricos em fenóis, carboidratos e lipídios, fato que pode interferir na obtenção de géis reprodutíveis e de alta resolução. No presente trabalho, foram avaliados três diferentes metodologias de extração de proteínas de embriões cotiledonares de *A. angustifolia*: a) solubilização e precipitação em ácido tricloroacético (TCA); b) extração bifásica em fenol seguido de precipitação em acetato de amônio; c) extração em solução de uréia e tiouréia seguido de precipitação em TCA. Os géis 2-DE obtidos, através da terceira metodologia de extração apresentaram maior resolução, reprodutibilidade e número total e específico de “spots” detectados.

### **2. Introdução**

A estratégia mais utilizada para análise do proteoma em plantas consiste na sequência: 1) extração de proteínas em tampão com agentes caotrópicos, 2) separação protéica através de 2-DE em gel de poliacrilamida, 3) digestão de “spots” *in gel*, 4) identificação da proteína por espectrometria de massas, e 5) buscas em bancos de dados disponíveis (GIAVALISCO *et al.*, 2003).

A 2-DE é uma das ferramentas mais poderosas da proteômica na separação e quantificação de proteínas (NATARAJAN *et al.*, 2005). O uso de gradientes imobilizados de pH (IPG) na focalização isoeletrica, e a otimização no processo de aplicação das amostras têm permitido a utilização de quantidades cada vez menores de proteínas e maior reprodutibilidade nos géis estudados (WESTERNMEIER, 2001). Na 2-DE, o preparo da amostra é uma etapa crucial. A solubilização e extração incompleta de proteínas podem

afetar o processo e, conseqüentemente, comprometer o estudo do proteoma de um tecido, órgão ou organismo (ISLAM *et al.*, 2004; CARPENTIER *et al.*, 2005). Em tecidos vegetais, a etapa de extração requer maior atenção, já que grande parte do volume da célula é ocupada pela parede celular e vacúolos, os quais acumulam grande quantidade de moléculas que podem interferir na proteômica, como os ácidos orgânicos, carboidratos, lipídios e fenóis (HOCHHOLDINGER *et al.*, 2006).

Durante o preparo das amostras para 2-DE, as proteínas precisam ser solubilizadas, desnaturadas e reduzidas sem alterar suas cargas e massas moleculares (VÂLCU e SCHLINK, 2006). Atualmente, diferentes metodologias de extração têm sido descritas para contornar os problemas referentes à quantidade de interferentes e à baixa concentração de proteínas em células vegetais. Apesar dos avanços, a maioria dos trabalhos faz uso da metodologia tradicional de extração, onde é utilizado TCA e acetona, seguido de ressolubilização das proteínas em tampões contendo agentes caotrópicos e detergentes (SARAVANAN e ROSE, 2004). Dessa maneira pode-se obter extratos protéicos em alta concentração e livre de contaminantes; embora compostos poliméricos, como os polissacarídeos e os polifenóis, não são retirados através dessa metodologia (SARAVANAN e ROSE, 2004).

Além dos compostos fenólicos, terpenos e ácidos graxos, os carboidratos podem interferir na 2-DE (CARPENTIER *et al.*, 2005). Carboidratos, como o amido, podem bloquear os poros dos géis de acrilamida, causando precipitação, aumento no tempo de focalização, rastros e perda de proteínas durante a focalização isoeletrica (CARPENTIER *et al.*, 2005). Sementes de *A. angustifolia* possuem uma grande quantidade de compostos polifenólicos que na sua maioria está limitada aos tecidos da testa da semente. A sua transferência para o megagametófito ocorre apenas após o cozimento da semente para consumo, mais especificamente, para a alimentação humana (CORDENUNSI *et al.*, 2004). Em *A. angustifolia*, o metabólito de reserva mais abundante é o amido (CORDENUNSI *et al.*, 2004; CONFORTI e LUPANO, 2007; 2008). Sendo assim, o objetivo deste trabalho foi estabelecer uma metodologia adequada para a extração de amostras protéicas a ser utilizada em sistemas 2-DE nos embriões de *A. angustifolia*.

### **3. Materiais e métodos**

#### *3.1. Material vegetal*

Foram utilizadas sementes de *A. angustifolia*, coletadas no município de Bom Retiro – SC (27°47'S e 49°29'W) em abril 2005. Os embriões, no estágio cotiledonar, foram isolados e armazenados à -80°C para avaliações posteriores.

#### *3.2. Extração de proteínas*

Para cada extrato protéico, 10 embriões cotiledonares foram pulverizados em nitrogênio líquido. Foram utilizados 500 mg do pó e avaliadas as seguintes metodologias de extração (Figura 1).

##### *3.2.1. Metodologia 1: Precipitação em solução de ácido tricloroacético*

Essa metodologia foi baseada naquela descrita por DAMERVAL *et al.* (1986). O material pulverizado foi ressuspendido em solução de acetona contendo 20% de TCA, 0,2% de ditioneitol (DTT), 1 mM de fluoreto de fenilmetilsulfonil (PMSF) e 5 µM de pepstatina. As proteínas foram precipitadas, por 12 h, à temperatura de -20°C. Posteriormente, as amostras foram centrifugadas por 30 min à 12000 *g* (4°C), o sobrenadante descartado e as proteínas lavadas por três vezes em acetona gelada contendo 0,2% de DTT. Finalmente, as proteínas foram ressuspendidas em solução contendo 7 M de uréia, 2 M de tiouréia, 1% de DTT, 2% de Triton-100, 2% de IPG Buffer pH 3-10 (GE Healthcare, Chalfont St. Giles, England), 1 mM de PMSF e 5 µM de pepstatina.

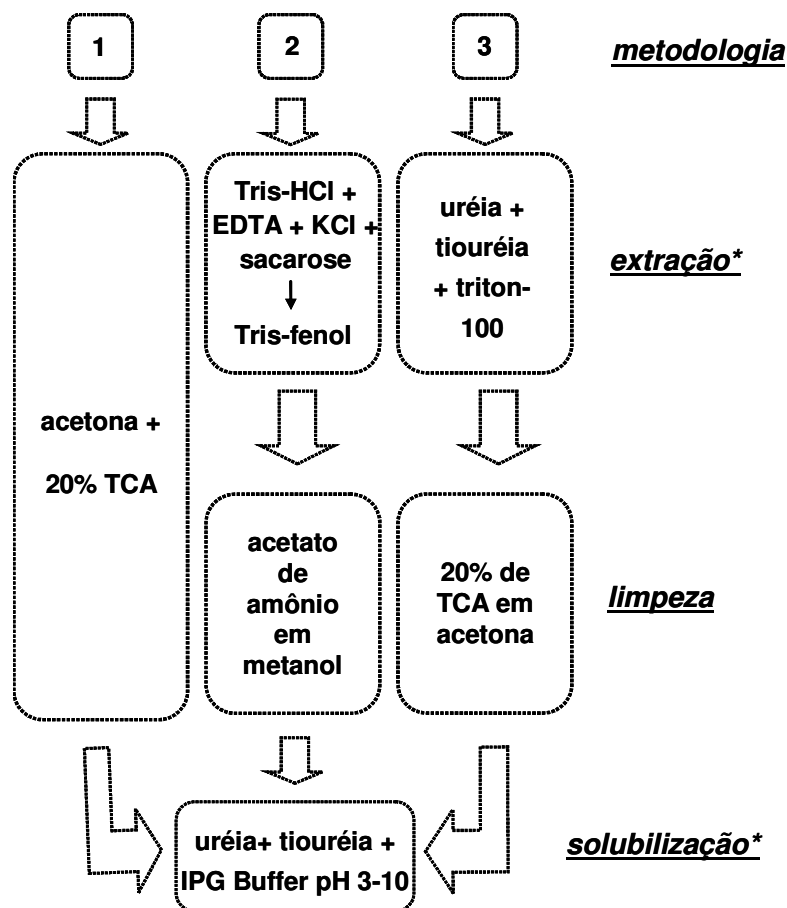


Figura 1: Metodologias de extração de proteínas de embriões de *A. angustifolia*. **Metodologia 1:** solubilização e precipitação direta em ácido tricloroacético (TCA). **Metodologia 2:** extração em tampão fenol e precipitação em acetato de amônio. **Metodologia 3:** extração em solução de uréia-tiouréia seguido de precipitação em TCA. **EDTA:** ácido etilenodiaminotetracético. \* adicionado 1% de ditioneitol, 1 mM de fenilmetilsulfonil e 5  $\mu$ M de pepstatina nas soluções.

### 3.2.2. Metodologia 2: Extração em tampão fenol e precipitação em acetato de amônio

Essa metodologia foi baseada naquela descrita por CARPENTIER *et al.* (2005). O material pulverizado foi ressuscitado em tampão de extração contendo 50 mM de Tris HCl pH 8,8, 5 mM de ácido etilenodiaminotetracético (EDTA), 100 mM de KCl, 1 mM de PMSF, 5  $\mu$ M de pepstatina, 30% de sacarose e 1% de DTT. Após agitação por 30 s, foi adicionado o tampão Tris-fenol (pH 8,0, GE Healthcare) e o material agitado novamente por 15 min. Posteriormente, a amostra foi centrifugada por 3 min (6000 *g* à 4°C), a fração fenólica coletada e as proteínas precipitadas por 12 h pela adição de cinco volumes de solução metanólica contendo 100mM de acetato de amônio à -20°C. Finalmente, as proteínas foram

ressuspendidas em tampão contendo 7M de uréia, 2M de tiouréia, 1% de DTT, 2% de Triton-100, 2% de IPG Buffer pH 3-10, 1mM de PMSF e 5µM de pepstatina, como descrito na seção 3.2.1.

### 3.2.3. Metodologia 3: Extração em solução de uréia-tiouréia

Essa metodologia foi baseada naquelas descritas por CARPENTIER *et al.* (2005) e NATARAJAN *et al.* (2005). O material pulverizado foi ressuspendido em solução de extração contendo 7 M de uréia, 2 M de tiouréia, 1% de DTT, 2% de Triton-100, 1 mM de PMSF e 5 µM de pepstatina. Cada uma das amostras foi agitada por 15 min, centrifugada por 5 min (12000 *g* à 4°C) e, no sobrenadante coletado, as proteínas foram precipitadas, por 1 h, em solução de acetona contendo 20% TCA e 0,2% de DTT. Após centrifugação (12000 *g* à 4°C) por 30 min, as proteínas lavadas por três vezes em acetona gelada contendo 0,2% de DTT. Finalmente, as proteínas foram ressuspendidas em tampão contendo 7 M de uréia, 2 M de tiouréia, 1% de DTT, 2% de Triton-100, 2% de IPG Buffer pH 3-10 (GE Healthcare), 1 mM de PMSF e 5 µM de pepstatina, como descrito na seção 3.2.1.

### 3.3. Quantificação de proteínas

O conteúdo de proteínas nos extratos foi determinado pelo 2-D Quant Kit (GE Healthcare), contendo albumina sérica bovina como padrão.

### 3.4. Eletroforese bidimensional

As amostras foram diluídas em tampão de reidratação, contendo 7 M de uréia, 2 M de tiouréia, 2% de CHAPS, 2% de IPG Buffer pH 4-7, 1% de DTT e 0,002% de azul de bromofenol, até a concentração de 570 µg de proteínas por 1 mL de extrato. Para cada gel, foram utilizados 200 µg de proteínas, carregados simultaneamente à reidratação, em tira de IPG (GE Healthcare). Foram utilizadas tiras de 18 cm com faixa de separação linear de pH 4-7. A reidratação foi realizada por 12 h, e a focalização isoelétrica foi realizada em uma unidade de IPG Phor II (GE Healthcare), à temperatura de 20°C, com limite de corrente de 50 µA por tira, segundo as etapas: 0,5 kVh à 500 V, 8 h à 1000 V, 3 h à 8000 V, 20 kVh à

8000 V. Antes de iniciar a segunda dimensão, as proteínas foram reduzidas, por 15 min, em solução de equilíbrio (50 mM de Tris-HCl pH 8,8, 6 M de uréia, 30% de glicerol, 2% de SDS, 0,002% de azul de bromofenol) adicionado de 125 mM de DTT e, posteriormente, alquiladas (solução de equilíbrio adicionada de 125 mM de iodoacetamida). A segunda dimensão foi realizada em gel de 12% de poliacrilamida, em condições desnaturantes (SDS-PAGE). A separação eletroforética foi realizada com corrente constante de 25 mA por gel, em um sistema vertical Protean II (BioRad, Hercules, EUA).

### 3.5. Coloração dos géis

As proteínas foram visualizadas pela coloração com nitrato de prata (SHEVCHENKO *et al.*, 1996). Inicialmente, os géis foram fixados, por 12 h, em solução de fixação contendo 50% de metanol, 12% de ácido acético e 0,05% de formaldeído. Após duas lavagens por 30 min em solução de 50% de etanol, os géis foram incubados, por 2 min, em solução de 0,02% de tiosulfato de sódio e, posteriormente, lavados em água. Em seguida, os géis foram incubados por 20 min em solução de 0,2% de nitrato de prata e 0,08% de formaldeído. Após lavagem em água, a revelação foi realizada durante 3 min pela incubação dos géis em solução de 6% de carbonato de sódio, 0,16% de tiosulfato de sódio e 0,05% de formaldeído e interrompida utilizando a mesma solução de fixação do início do processo de coloração.

### 3.6. Análise dos géis

Os géis bidimensionais (2-D) foram digitalizados e analisados no programa Image Master Platinum 6.0 (GE Healthcare). A autenticidade e o contorno de cada “spot” foram determinados automaticamente e, posteriormente, validados por inspeção visual e editados quando necessários. Duas repetições foram realizadas para cada metodologia de extração, sendo avaliados o número de “spots” detectados e a resolução dos géis obtidos. Os alinhamentos entre os géis das repetições foi realizado automaticamente, utilizando três “spots” de referência, seguido de inspeção e correção manual quando necessário.

#### **4. Resultados e discussão**

No presente trabalho, foram avaliados três diferentes metodologias de extração de proteínas para uso em sistemas 2-DE em embriões de *A. angustifolia*. Proteínas, lipídios e amido são as principais fontes de reserva de energia em sementes, sendo sintetizadas em abundância durante o desenvolvimento dos cotilédones (YANG *et al.*, 2007). ASTARITA *et al.* (2003) estudaram o perfil de acúmulo de proteínas durante o desenvolvimento do embrião zigótico de *A. angustifolia*. Estes pesquisadores verificaram um aumento na quantidade de proteínas durante a embriogênese, embora não tenham sido observadas diferenças significativas no conteúdo de proteínas entre os estádios cotiledonar e maduro. A escolha do estágio cotiledonar para o estabelecimento da metodologia de extração deve-se à grande quantidade de proteínas nestes embriões e ao alto teor de amido, em sementes desta espécie, após o desenvolvimento cotiledonar (CORDENUNSI *et al.*, 2004).

Em todas as análises foram utilizadas tiras de IPG de 4-7. Estudos eletroforéticos preliminares demonstraram que a disposição preferencial de proteínas, quando submetidas ao fracionamento em gradiente amplo de pH (3-10), é em torno do pH 6 (Figura 2). Sendo assim, para o experimental, foram utilizadas tiras de IPG de 4-7, as quais permitiram melhor visualização, detecção e resolução dos géis 2-D.

Por ser uma estratégia simples, rápida e eficiente, a metodologia de extração e precipitação de proteínas em soluções de TCA (metodologia 1) é o procedimento mais utilizado em tecidos vegetais (GIAVALISCO *et al.*, 2003). Em embriões de *A. angustifolia*, seu uso resultou na detecção de  $824 \pm 144$  “spots” de proteína, nos géis 2-D (Tabela 1).

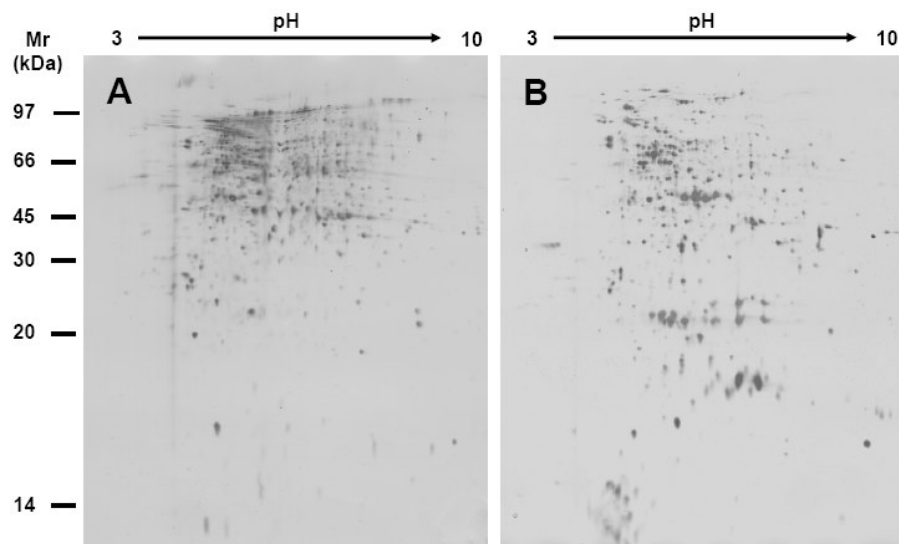


Figura 2: Géis bidimensionais de proteínas de embriões globulares (**A**) e maduros (**B**) de *A. angustifolia*. O fracionamento protéico foi realizado em tiras de poliacrilamida com gradiente imobilizado de pH de 3-10 (18 cm). A SDS-PAGE foi realizada em géis de 12% de poliacrilamida com corrente constante de 25 mA por gel.

A maior desvantagem de metodologias que empregam precipitação por TCA é a difícil ressolubilização das proteínas (NANDAKUMAR *et al.*, 2003). Alternativamente, estratégias distintas de fracionamento e limpeza das soluções protéicas têm sido empregadas. Dentre essas, a utilização da extração por fenol (metodologia 2) é a mais utilizada em tecidos com altos teores de interferentes (SARAVANAN e ROSE, 2004; CARPENTIER *et al.*, 2005; VINCENT *et al.*, 2006). Recentemente, SARAVANAN e ROSE (2004) e CARPENTIER *et al.* (2005) adaptaram esse método, primeiramente proposto por HURKMAN e TANAKA (1986), pela adição de sacarose. Esta adaptação permite a inversão e obtenção de uma fase aquosa rica em carboidratos, ácidos nucleicos e demais interferentes aquosos. Além disso, o uso de KCl, EDTA, tampões alcalinos e inibidores de proteases foi proposto numa tentativa de otimizar o processo de extração bifásico, e de inibir a ação de enzimas proteolíticas (CARPENTIER *et al.*, 2005). Esta metodologia de extração por fenol (metodologia 2) também foi avaliada para extração de proteínas em embriões de *A. angustifolia*. Foram detectados  $737 \pm 66$  “spots” de proteínas nos géis 2-D (Tabela 1) e observados rastros horizontais na região ácida do gel (Figura 3).

Tabela 1: Conteúdo protéico e número total de “spots” detectados através das três diferentes metodologias de extração em embriões de *A. angustifolia*. MF: massa fresca. TCA: ácido tricloroacético.

Metodologia	Conteúdo protéico (mg <sup>-1</sup> .g MF)	“Spots” detectados
1 (extração em TCA)	5,17 ± 0,07	824 ± 144
2 (extração em fenol)	5,75 ± 0,19	737 ± 66
3 (extração em uréia e tiouréia)	5,22 ± 1,32	1009 ± 48

Uma terceira metodologia de extração foi avaliada, utilizando solução de uréia e tiouréia para solubilização de proteínas. Embora os protocolos originais descrevam somente o uso da uréia, a tiouréia é atualmente adicionada por possuir capacidade de aumentar a solubilização de proteínas conjugadas a açúcares e proteínas hidrofóbicas (GIAVALISCO *et al.*, 2003; CARPENTIER *et al.*, 2005). Esta metodologia, seguida de precipitação em TCA resultou na detecção do maior número de “spots” dentre as metodologias avaliadas: 1009 ± 48 (Tabela 1). Adicionalmente, os géis obtidos não apresentaram rastros horizontais nos géis 2-DE (Figura 3).

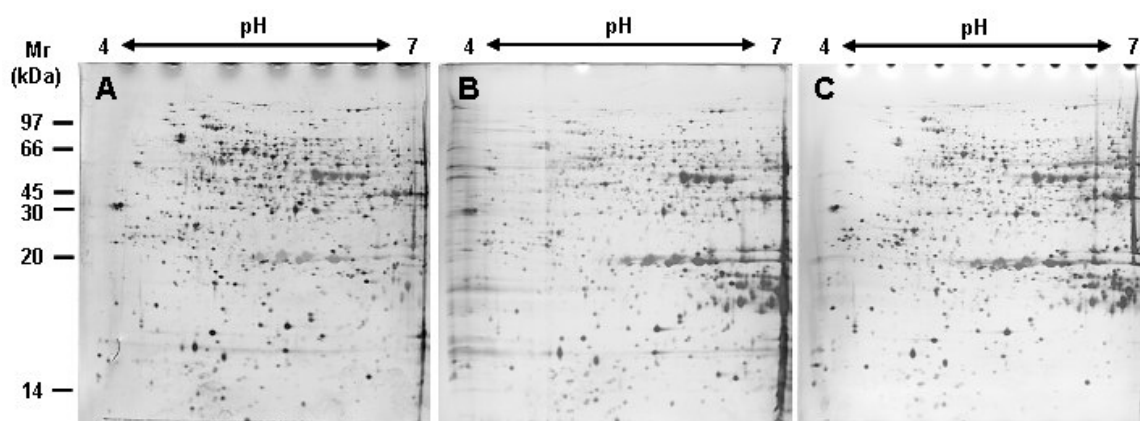


Figura 3: Géis bidimensionais de proteínas de embriões cotiledonares de *A. angustifolia* extraídas em ácido tricloroacético (TCA) (metodologia 1) (A), fenol seguido de precipitação em acetato de amônio (metodologia 2) (B) e uréia e tiouréia seguido de precipitação em TCA (metodologia 3) (C).

Embora tenham sido detectadas diferenças no número total de “spots” nos géis 2-DE de cada metodologia de extração, não foram observadas diferenças entre os conteúdos

protéicos (Tabela 1). Assim, a maior quantidade de “spots” detectados a partir da extração com uréia e tiouréia (metodologia 3) deve estar relacionada à obtenção de géis de maior resolução.

O fator reprodutibilidade também foi avaliado quantitativamente, pela obtenção de gráficos de dispersão de intensidade dos “spots” entre as repetições de cada metodologia de extração. A metodologia de solubilização de proteínas em TCA (metodologia 1) foi aquela com menor reprodutibilidade entre as repetições, apresentando um coeficiente de correlação linear de 0,782 (Figura 4A). Contrariamente, nas repetições obtidas a partir das metodologias de extração em fenol (metodologia 2) e em uréia e tiouréia (metodologia 3) observaram-se altos coeficientes de correlação linear: 0,940 e 0,944, respectivamente (Figura 4B e C).

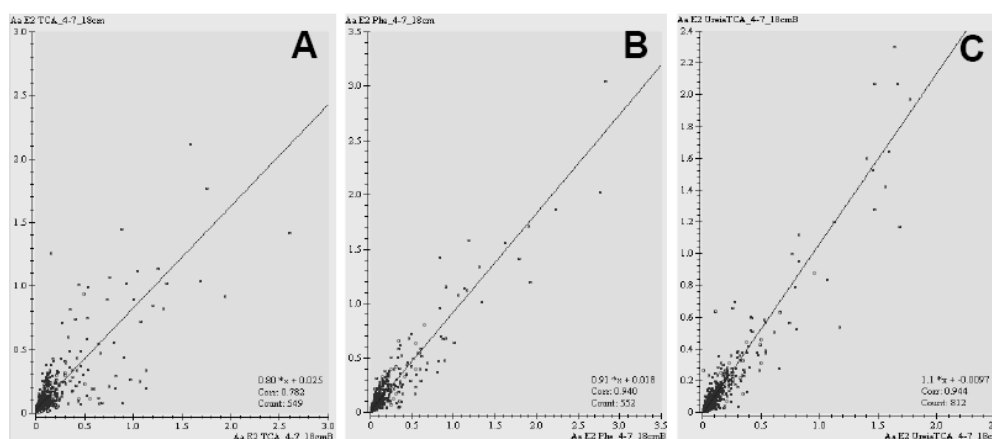


Figura 4: Gráficos de dispersão de intensidade de “spots” entre as repetições de cada metodologia de extração de proteínas de embriões cotiledonares de *A. angustifolia*. **A:** solubilização em ácido tricloroacético (TCA) (metodologia 1), **B:** extração em fenol seguido de precipitação em acetato de amônio (metodologia 2), **C:** extração em uréia e tiouréia seguido de precipitação em TCA (metodologia 3).

Os padrões eletroforéticos, obtidos a partir dos mapas 2-D, foram avaliados nas três metodologias de extração (Figura 5). Para tanto, os géis analisados foram alinhados, visando atribuir identidades comuns entre os “spots” das repetições, e os “spots” presentes nas duas repetições foram adicionados aos mapas. Não foram visualizadas variações no perfil eletroforético de distribuição dos “spots” de acordo com a massa molecular relativa

(Mr) entre as metodologias avaliadas, sendo que, a classe de “spots” contendo Mr entre 20 e 40 kDa foi a mais abundante (Figura 5). Posteriormente, os mapas 2-DE foram alinhados, e o número de “spots” especificamente detectados em cada metodologia foi computado. As metodologias de extração por TCA e fenol apresentaram a menor quantidade de “spots” específicos, ou seja, 245 e 130, respectivamente. Os mapas 2-D obtidos a partir da extração em uréia e tiouréia apresentaram a maior quantidade de “spots” específicos (380).

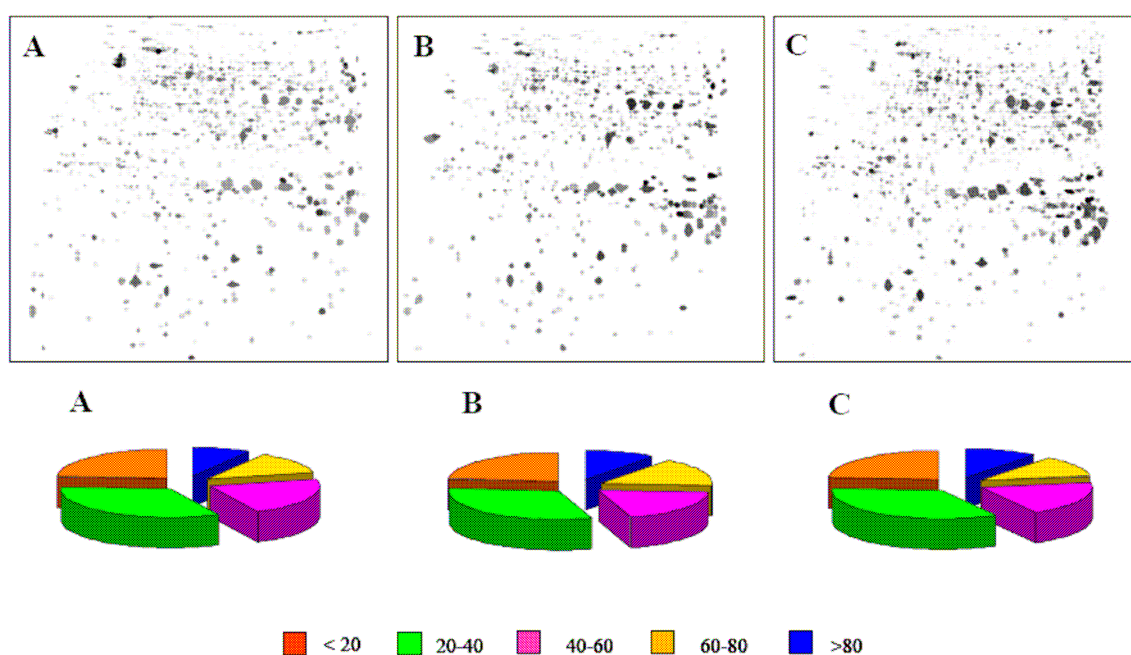


Figura 5: Mapas 2-D de proteínas isoladas de embriões cotiledonares de *A. angustifolia* e distribuição de “spots” de acordo com a massa molecular relativa. **A**: extração em ácido tricloroacético (TCA) (metodologia 1), **B**: extração em fenol seguido de precipitação em acetato de amônio (metodologia 2), **C**: extração em tiouréia e uréia seguido de precipitação em TCA (metodologia 3).

## 5. Considerações finais

Estudos comparativos do proteoma de um organismo por 2-DE requerem a maximização no número de “spots” detectados e a obtenção de géis reprodutíveis. No presente trabalho foram testados três metodologias de extração de proteínas de embriões cotiledonares de *A. angustifolia*. Em todos os métodos, foi adicionada uma etapa de precipitação objetivando a limpeza dos extratos protéicos. Embora a precipitação esteja

associada à perda de proteínas, no presente trabalho verificamos que essas podem ser reproduzíveis e, assim, não comprometer a sua utilização em estudos comparativos.

O fenômeno de comigração de “spots”, ou padrão similar de distribuição eletroforética, é comum em análises bidimensionais de amostras protéicas complexas e pode resultar em géis 2-D de baixa resolução. Tal fato, associado às impurezas e rastros presentes nos géis, podem interferir na visualização e detecção dos mesmos, gerando mapas 2-D com baixo número de “spots”. Embora não tenham sido verificados diferentes padrões de distribuição de Mr entre as metodologias de extração avaliadas, houve variação no número total de “spots” detectados. Além disso, foram verificadas diferentes quantidades de “spots” específicos, indicando que tais variações podem refletir a heterogeneidade físico-química das proteínas, e as distintas propriedades de solubilização das metodologias testadas.

Para embriões de *A. angustifolia*, a metodologia de extração em uréia e tiouréia (metodologia 3) foi aquela que resultou em um maior número de “spots” totais e específicos, além de maior reprodutibilidade. Assim, essa metodologia foi escolhida para a extração de proteínas e estudo da variação do proteoma durante a embriogênese zigótica de *A. angustifolia*.

## 6. Referências bibliográficas

- Astarita, L. V., E. I. S. Floh, W. Handro. Free amino acid, protein and water content changes associated with seed development in *Araucaria angustifolia*. Biologia Plantarum, v.47, p.53-59. 2003.
- Carpentier, S. C., E. Witters, K. Laukens, P. Deckers, R. Swennen, B. Panis. Preparation of protein extracts from recalcitrant plant tissues: and evaluation of different methods for two-dimensional gel electrophoresis. Proteomics, v.5, p.2497-2507. 2005.
- Conforti, P. A., C. E. Lupano. Starch characterization of *Araucaria angustifolia* and *Araucaria araucana* seeds. Starch-Starke, v.59, p.284-289. 2007.
- Conforti, P. A., C. E. Lupano. Comparative study of the starch digestibility of *Araucaria angustifolia* and *Araucaria araucana* seed flour. Starch-Starke, v.60, p.192-198. 2008.

- Cordenunsi, B. R., E. W. De Menezes, M. I. Genovese, C. Colli, A. G. A. De Souza, F. M. Lajolo. Chemical composition and glycemic index of Brazilian pine (*Araucaria angustifolia*) seeds. Journal of Agricultural and Food Chemistry, v.52, p.3412-3416. 2004.
- Damerval, C., D. Devienne, M. Zivy, H. Thiellement. Technical improvements in two-dimensional electrophoresis increase the level of genetic-variation detected in wheat-seedling proteins. Electrophoresis, v.7, p.52-54. 1986.
- Giavalisco, P., E. Nordhoff, H. Lehrach, J. Gobom, J. Klose. Extraction of proteins from plant tissues for two-dimensional electrophoresis analysis. Electrophoresis, v.24, p.207-216. 2003.
- Hochholdinger, F., M. Sauer, D. Dembinsky, N. Hoecker, N. Muthreich, M. Saleem, Y. Liu. Proteomic dissection of plant development. Proteomics, v.6, p.4076-4083. 2006.
- Hurkman, W. J., C. K. Tanaka. Solubilization of plant membrane-proteins for analysis by two-dimensional gel-electrophoresis. Plant Physiology, v.81, p.802-806. 1986.
- Islam, N., M. Lonsdale, N. M. Upadhyaya, T. J. Higgins, H. Hirano, R. Akhurst. Protein extraction from mature rice leaves for two-dimensional gel electrophoresis and its application in proteome analysis. Proteomics, p.1903-1908. 2004.
- Nandakumar, M. P., J. Shen, B. Raman, M. R. Marten. Solubilization of trichloroacetic acid (TCA) precipitated microbial proteins via NaOH for two-dimensional electrophoresis. Journal of Proteome Research, v.2, p.89-93. 2003.
- Natarajan, S., C. Xu, T. J. Carpena, W. M. Garrett. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. Analytical Biochemistry, v.342, p.214-220. 2005.
- Saravanan, R. S., J. K. C. Rose. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. Proteomics, v.4, p.2522-2532. 2004.
- Shevchenko, A., M. Wilm, O. Vorm, M. Mann. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Analytical Chemistry, v.68, p.850-858. 1996.
- Vâlcu, C. M., K. Schlink. Reduction of proteins during sample preparation and two-dimensional gel electrophoresis of woody plant samples. Proteomics, v.6, p.1599-1605. 2006.
- Vincent, D., M. D. Wheatley, G. R. Cramer. Optimization of protein extraction and solubilization for mature grape berry clusters. Electrophoresis, v.27, p.1853-1865. 2006.
- Westernmeier, R. Electrophoresis in practice. Weinheim: Wiley. 2001. 257 p.
- Yang, P., X. Li, X. Wang, H. Chen, F. Chen, S. Shen. Proteomic analysis of rice (*Oryza sativa*) seeds during germination. Proteomics, v.7, p.3358-3368. 2007.

## **CAPÍTULO 3**

### **CAPÍTULO 3. Desenvolvimento de um programa de seleção de espectros para identificação de proteínas em embriões de *Araucaria angustifolia*<sup>1,2</sup>**

#### **1. Resumo**

A análise e fragmentação de peptídeos tripticos por espectrometria de massas geram espectros que são utilizados na identificação de proteínas, após buscas em bancos de dados. Entretanto, apenas uma pequena parte dos espectros obtidos resulta em identificação positiva. Em organismos cujo genoma ainda não é conhecido, como em *A. angustifolia*, espectros interferentes podem afetar o tempo de processamento de dados, qualidade de alinhamentos obtidos e número de identificações, principalmente quando utilizadas estratégias de identificação baseadas em homologia e similaridade de sequências. Nesse trabalho, foi desenvolvido um programa para remoção de espectros de fragmentos de queratina e tripsina, interferentes comuns em identificações de proteínas por espectrometria de massas após eletroforese bidimensional (2-DE). Avaliações de desempenho, realizadas a partir de espectros obtidos de proteínas digeridas de géis bidimensionais (2-D) de embriões de *A. angustifolia*, mostraram que o programa diminui o número de espectros interferentes, resultando em maior número de identificações positivas, redução no tempo de processamento de dados e no número de identificações limítrofes, as quais requerem cuidadosa validação manual. Devido ao alto desempenho, tanto em buscas restritivas por MASCOT, como por similaridade de sequências (MS BLAST), o programa foi incorporado na estratégia de análise do proteoma da embriogênese zigótica de *A. angustifolia*.

---

<sup>1</sup> Este capítulo foi desenvolvido no Max Planck Institute of Molecular Cell Biology and Genetics, sob orientação do Prof. Dr. Andrej Shevchenko.

<sup>2</sup> Este capítulo encontra-se publicado no periódico *Journal of Proteome Research* 7, 3382-3395 (2008).

## 2. Introdução

Atualmente, a estratégia mais utilizada para identificação de proteínas é a fragmentação de peptídeos tripticos em um espectrômetro de massas, seguido de alinhamento dos espectros obtidos com sequências depositadas em bancos de dados (ONG e MANN, 2005). Embora seja possível obter milhares de espectros de qualidade em uma única análise, apenas 5 à 50% resultam em alinhamentos, sendo que uma porcentagem ainda menor resulta em identificações positivas, ou seja, referentes a proteínas alvo (JOHNSON *et al.*, 2005).

Os espectros não alinhados podem ser agrupados em três tipos. O primeiro grupo é formado por aqueles oriundos de proteínas alvo, mas que não foram alinhados nos bancos de dados devido à presença de sequências polimórficas ou modificações pós traducionais (LISKA *et al.*, 2004; SHEVCHENKO *et al.*, 2005). Entretanto, a identificação dessas proteínas ainda pode ser realizada pelo uso de estratégias de busca tolerante a erros (Figura 1). O segundo grupo de espectros origina-se de fontes não protéicas, como detergentes e plásticos, cujos íons precursores podem ser facilmente reconhecidos por espectrômetros de alta resolução e excluídos dos eventos de fragmentação, se necessário, por possuírem apenas uma carga (SCHLOSSER e VOLKMER-ENGERT, 2003).

Fragmentos de queratina e tripsina formam o terceiro grupo de espectros não alinhados. Eles são os principais contaminantes em experimentos envolvendo espectrometria de massas e, em abundância, podem gerar um grande número de sequências polimórficas e adutos de sódio, além de sobrecarregar os arquivos, aumentando os tempos de busca nos bancos de dados.

Em organismos cujo genoma ainda não é totalmente conhecido, a identificação de proteínas pode ser realizada através de buscas por similaridade de sequências (KATZ *et al.*, 2007; WARINDEL *et al.*, 2007) (Figura 1). Na proteômica dirigida por homologia, sequências redundantes, degeneradas e não precisas, obtidas por sequenciamento *de novo* automatizado de espectros, são alinhadas com sequências de proteínas de organismos filogeneticamente distantes daquele em estudo (HABERMANN *et al.*, 2004; LISKA *et al.*, 2004). Entretanto, quanto menor a restringência no tamanho do banco de dados utilizado,

maior a obtenção de alinhamentos não relativos à proteína alvo. Adicionalmente, sequências de queratina são ricas em fragmentos repetitivos e de baixa complexidade e, portanto, podem gerar identificações que, em primeira instância não referem-se a queratinas. A validação dessas identificações requer inspeção manual de cada alinhamento obtido pelo BLAST com a sequência completa da proteína identificada, o que inviabiliza seu uso em programas de identificação em larga escala.

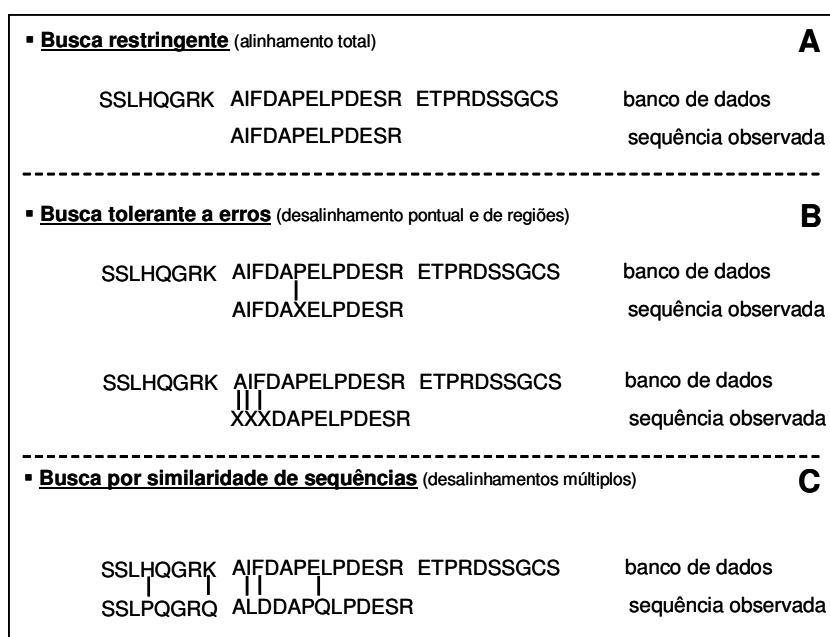


Figura 1: Diferentes estratégias de alinhamento utilizadas na identificação de proteínas por sequência de fragmentos peptídicos. **A**: Busca restrigente. **B**: Busca tolerante a erros. **C**: Busca por similaridade de sequências. A letra X indica imprecisão na determinação do aminoácido.

Devido à heterogeneidade dos íons precursores, os espectros dos fragmentos contaminantes são pouco reprodutíveis dificultando a seleção dos mesmos em listas de exclusão. Sendo assim, o objetivo desse trabalho foi desenvolver e avaliar um programa para rápido reconhecimento e exclusão de espectros de contaminantes, independente do uso de banco de dados, para identificação de proteínas de *A. angustifolia*, um organismo cujo genoma é desconhecido.

### **3. Materiais e métodos**

#### *3.1. Obtenção de extratos peptídicos de embriões de *A. angustifolia* e de queratina*

Embriões de *A. angustifolia*, obtidos a partir de sementes coletadas no município de Bom Retiro - SC (27°47'S e 49°29'W) em abril 2005, foram isolados e pulverizados em nitrogênio líquido. As proteínas foram extraídas em solução contendo 7 M de uréia, 2 M de tiouréia, 1% de ditioneína (DTT), 2% de Triton-100, 1 mM de fluoreto de fenilmetilsulfonil (PMSF) e 5 µM de pepstatina, como descrito no método de extração em solução de uréia-tiouréia (Capítulo 2). Cada amostra foi agitada por 15 min, centrifugada por 5 min (12000 g à 4°C) e o sobrenadante coletado. Posteriormente, as proteínas foram precipitadas por 1 h em solução de acetona contendo 20% TCA e 0,2% de DTT. Após centrifugação (12000 g à 4°C) por 30 min, as proteínas lavadas por três vezes em acetona gelada contendo 0,2% de DTT. Finalmente, as proteínas foram ressuspensas em tampão contendo 7 M de uréia, 2 M de tiouréia, 1% de DTT, 2% de Triton-100, 2% de IPG Buffer pH 3-10 (GE Healthcare, Little Chalfort, England), 1 mM de PMSF e 5 µM de pepstatina. A quantidade de proteínas em cada extrato foi determinada pelo 2-D Quant Kit (GE Healthcare), contendo albumina sérica bovina como padrão.

Previamente à 2-DE, as amostras de embriões foram diluídas em tampão de reidratação, contendo 7 M de uréia, 2 M de tiouréia, 2% de CHAPS, 2% de IPG Buffer pH 4-7, 1% de DTT e 0,002% de azul de bromofenol, até a concentração de 570 µg de proteínas por 1 mL de extrato. Para cada gel, foram utilizados 200 µg de proteínas, carregados simultaneamente à reidratação em tira de gradiente de pH imobilizado (IPG, GE Healthcare). Foram utilizadas tiras de IPG de 18 cm com faixa de separação linear de pH 4-7. A reidratação foi realizada por 12 h e a focalização isoeletrica foi conduzida em uma unidade de IPG Phor II (GE Healthcare) à 20°C com limite de corrente de 50 µA por tira, segundo as etapas: 0,5 kVh à 500 V, 8 h à 1000 V, 3 h à 8000 V, 20 kVh à 8000 V. Antes de iniciar a segunda dimensão, as proteínas foram reduzidas, por 15 min, em solução de equilíbrio (50 mM de Tris-HCl pH 8,8, 6 M de uréia, 30% de glicerol, 2% de SDS, 0,002% de azul de bromofenol) adicionado de 125 mM de DTT e, posteriormente, alquiladas (solução de equilíbrio adicionada de 125 mM de iodoacetamida). A segunda dimensão foi realizada em

gel de 12% de poliacrilamida, em condições desnaturantes (SDS-PAGE). A separação eletroforética foi conduzida com corrente constante de 25 mA por gel, em um sistema vertical Protean II (BioRad, Hercules, EUA) e a visualização dos “spots” foi realizada pela coloração com nitrato de prata (SHEVCHENKO *et al.*, 1996).

A digestão de proteínas foi realizada *in gel* (SHEVCHENKO *et al.*, 2006). Para tanto, “spots” selecionados foram isolados dos géis 2-D, cortados em pequenos fragmentos e desidratados por meio de adição de acetonitrila por 10 min. Posteriormente, as proteínas foram reduzidas e alquiladas em soluções de 100 mM de bicarbonato de amônio contendo 10 mM de DTT, por 30 min à 56° C, e 55 mM de iodoacetamida, por 10 min em temperatura ambiente no escuro, respectivamente. A digestão enzimática foi realizada através de adição de 16 ng.L<sup>-1</sup> tripsina (grau de sequenciamento, Promega, Madison, US) à 37° C por 12 h. Finalmente, os peptídeos foram extraídos em solução contendo 5% de ácido fórmico e acetonitrila (1:2, v/v) por 15 min à 37° C. O sobrenadante foi coletado, seco e os peptídeos ressuspensos em 0,05% de ácido trifluoroacético (TFA).

Os extratos peptídicos de queratina foram obtidos por meio de digestão trípica de fragmentos de 1 cm<sup>2</sup> de géis do tipo SDS-PAGE 12% sem prévio carregamento protéico.

### 3.2. Fracionamento e análises dos peptídeos em espectrômetro de massas

As misturas peptídicas, resultantes do processo de digestão, foram analisadas em um nanoLC-MS/MS por um cromatógrafo líquido Ultimate 3000 (Dionex, Sunnyvale, US) acoplado a um espectrômetro de massas do tipo LTQ-Orbitrap (ThermoFisher Scientific, Waltham, US). Para tanto, os peptídeos trípticos foram ressuspensos em 0,05% de TFA e eluídos em uma pré-coluna de carregamento C18 PepMAP100, (Dionex) (5 x 300 mm, 5 µm) sob um fluxo de 20 µL/min.

Após o carregamento, concentração e lavagem, os peptídeos foram separados em cromatografia líquida de alta eficiência (CLAE) usando uma nano-coluna C18 PepMAP100, 3 µm (Dionex) (15 cm x 75 µm). Foram utilizados como solventes uma solução de 5% acetonitrila e 1% de ácido fórmico (solvente A) e uma solução de 80% acetonitrila e 1% de ácido fórmico (solvente B). A separação foi realizada sob um fluxo de 200 nL/min e

a mudança na proporção do solvente B em relação ao solvente A definiu o gradiente de corrida, conforme descrito a seguir: de 5 à 20% nos primeiros 20 min, de 20 à 50% de 20 à 36 min, de 50 à 100% de 36 à 41 min, 100% de 41 à 51 min e 5% entre 51 e 56 min. Após a separação em CLAE, os peptídeos foram injetados em tandem no espectrômetro de massas utilizando um sistema de nanospray (ThermoFisher Scientific) equipado com uma agulha de sílica de 10 µm de diâmetro interno (New Objective, Woburn, US). A voltagem do spray utilizada foi de 1,8 kV e a temperatura do capilar de transferência iônica foi ajustada a 180°C. O espectrômetro de massas operou no modo dependente de dados (DDA). O controle de ganhos automáticos (AGC) foi ajustado à  $5 \times 10^5$  e  $5 \times 10^4$  para os eventos de fragmentação MS/MS no quadrupolo triplo linear (LTQ). Cada ciclo de aquisição de dados consistiu em uma varredura de moléculas entre 200-2000  $m/z$  no analisador Orbitrap seguido de fragmentação MS/MS, sob energia de colisão de 35% na célula de colisão linear, dos quatro íons precursores mais intensos. Íons de carga igual a +1 foram excluídos dos eventos de MS/MS e as  $m/z$  dos íons precursores fragmentados foram inseridos em uma lista de exclusão por 90 segundos.

### 3.3. Identificação de proteínas

Os espectros obtidos em cada ciclo de análise foram convertidos em um único arquivo tipo .mgf, utilizando o programa BioWorks 3.2 (ThermoFisher Scientific). As buscas foram realizadas no banco de dados MSDB através do programa MASCOT v.2.2 (Matrix Science, London, UK) instalado em dois servidores locais. As configurações de busca foram: 10 ppm para tolerância de massas do íon precursor e 0,5 Da para os fragmentos. Foi permitido uma falha na fragmentação. Carbamidometilação da cisteína, oxidação de metionina e acetilação de N-proteínas foram adicionadas como modificações variáveis. Foram consideradas identificações as proteínas com pelo menos três peptídeos de escore maior que 20. Para identificações a partir de somente um ou dois espectros, ao menos um peptídeo deve possuir escore acima de 50.

Paralelamente à busca restrigente por MASCOT, sequenciamento *de novo* automatizado foi realizado. Para tanto, o programa PepNovo foi utilizado e apenas

sequências candidatas com escore acima de 6 foram consideradas (FRANK e PEVZNER, 2005). Nas buscas por MS BLAST, todas as sequências peptídicas selecionadas foram combinadas em uma única entrada e as buscas foram realizadas no NCBI através do sítio <http://genetics.bwh.harvard.edu/msblast/>. A confirmação das proteínas identificadas por MASCOT, ou a identificação de novas proteínas foi realizada de acordo com a tabela de identificação do MS BLAST (HABERMANN *et al.*, 2004).

#### 3.4. “Eagle Eye”: o programa de seleção de espectros

O programa aqui descrito e denominado Eagle Eye foi desenvolvido em linguagem C++ em colaboração com os alunos de doutorado Victor Spirin (Brigham and Women’s Hospital - Harvard Medical School) e Magno Junqueira (Max Planck Institute of Molecular Cell Biology and Genetics) e pode ser acessado pelo sítio <http://genetics.bwh.harvard.edu/cgi-bin/msfilter/eagleeye.cgi>.

Basicamente, o programa desenvolvido busca dissimilaridades entre espectros, contidos nos arquivos do tipo .mgf, e aqueles presentes em uma biblioteca de interferentes. Visando sua utilização em diferentes experimentos, o Eagle Eye possui bibliotecas de espectros obtidos a partir de dois diferentes espectrômetros de massas (LTQ ou LTQ-Orbitrap). Além disso, o usuário pode utilizar sua própria biblioteca para processamento dos espectros experimentais. O alinhamento, e conseqüente filtragem, pode ser otimizado pelo uso de parâmetros de tolerância a erros nas massas dos precursores e fragmentos e por ajuste do valor  $p$ , referente à probabilidade de que espectros não oriundos de queratina e tripsina sejam filtrados aleatoriamente. Além disso, o Eagle Eye pode criar uma biblioteca não redundante por combinação de espectros provenientes de diferentes análises.

#### 3.5. Processamento da biblioteca de interferentes

Espectros de fragmentos de queratina e tripsina provenientes de cinco digestões independentes de géis de poliacrilamida 12% foram agrupados e, visando a inclusão de maiores quantidades de espectros interferentes, foram adicionados 2.183 espectros oriundos de auto digestões de tripsina em solução, identificados por MASCOT e MS BLAST,

e 1.750 espectros que não haviam sido removidos em filtragens anteriores. Para a obtenção de uma biblioteca não redundante, cada espectro foi alinhado à biblioteca pré-existente e apenas adicionado se não observado nenhum alinhamento (item 3.4).

## 4. Resultados e discussão

### 4.1. Desenvolvimento do programa

O processo de filtragem desenvolvido inicia-se por meio de comparação entre a carga e  $m/z$  dos íons precursores. Caso não sejam correspondentes, os espectros são considerados diferentes e o espectro candidato não é removido da análise. Contrariamente às buscas tradicionais em que se visa a detecção de alinhamentos significativos, o Eagle Eye busca dissimilaridades entre dois espectros. Tal estratégia deve-se à necessidade de detecção de informações suficientes para a não exclusão de espectros de interesse contendo pequeno número de sinais interferentes (Figura 2).

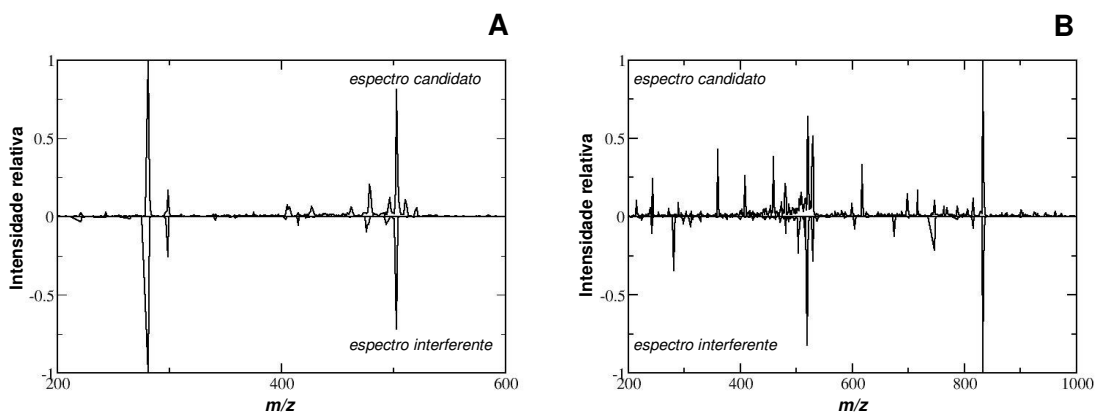


Figura 2: Diagramas comparativos entre espectros candidatos e interferentes em duas situações distintas. **A**: Espectro candidato possui grande quantidade de alinhamentos, indicando tratar-se de um espectro interferente. **B**: Espectro com grande quantidade de picos não alinhados e, embora possua alinhamentos interferentes, não deve ser removido.

Sendo assim, foi necessário o estabelecimento de uma variável de controle de exclusão, além das tolerâncias a erros de massas dos precursores e fragmentos, ajustável pelo usuário e que fosse baseada em medidas de dissimilaridade entre os espectros

comparados. Em ensaios anteriores, foi observado que o alinhamento entre espectros candidatos e interferentes ocorre principalmente por correspondência de sinais de pequena abundância e em regiões espectrais pouco abaixo da  $m/z$  do íon precursor alinhado. Portanto, para a determinação das medidas de dissimilaridade, os espectros foram empiricamente particionados em três regiões de  $m/z$  distintas ( $A$ ,  $B$  e  $C$ ) de acordo com a densidade de picos interferentes.

Em cada região, foram determinados valores referentes à soma das intensidades relativas dos picos presentes no espectro candidato ( $i$ ) e que não foram observados em um espectro depositado na biblioteca de interferentes ( $j$ ). Esse valor foi denominado  $S_{i,j}^n$  e a somatória de todas as intensidades relativas do espectro  $i$  foi denominado  $S_i^t$ . Analogamente,  $S_{j,i}^n$  e  $S_j^t$  referem-se aos mesmos valores, porém para o espectro  $j$ .

$$D_{ij} = \frac{c_A S_{A,i,j}^n + c_B S_{B,i,j}^n + c_C S_{C,i,j}^n + c_A S_{A,j,i}^n + c_B S_{B,j,i}^n + c_C S_{C,j,i}^n}{c_A S_{A,i}^t + c_B S_{B,i}^t + c_C S_{C,i}^t + c_A S_{A,j}^t + c_B S_{B,j}^t + c_C S_{C,j}^t} \quad (1)$$

Considerando que os sinais de maior massa encontram-se em regiões do espectro com menor incidência de interferentes (região  $C$ ) e que aqueles de menor massa encontram-se em regiões com alta densidade de interferentes (região  $B$ ), foram realizadas pequenas simulações computacionais entre espectros conhecidos não interferentes e espectros de queratina e tripsina, previamente identificados por buscas no MASCOT, visando otimizar a remoção de espectros interferentes sem comprometimento daqueles de interesse. Sendo assim, para cada valor da equação (1) foram multiplicados os seguintes coeficientes:  $c_A=2$ ;  $c_B=1$ ;  $c_C=4$ .

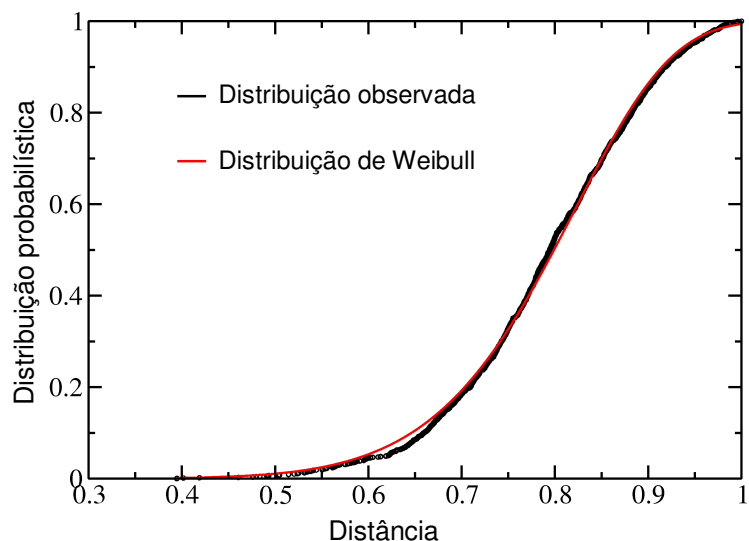


Figura 3: Distribuição cumulativa dos melhores alinhamentos entre espectros conhecidos e espectros interferentes em relação à distribuição de Weibull. O eixo das abscissas indica a menor distância, ou medida de dissimilaridade, observada entre o espectro conhecido e espectros da biblioteca, enquanto o eixo das ordenadas indica a probabilidade de obter a mesma distância aleatoriamente.

No presente trabalho, o limite de remoção de espectros é esperado somente para aqueles que apresentem a menor medida de dissimilaridade em relação ao espectros interferentes. As propriedades estatísticas pela busca da menor distância, ou dissimilaridade, entre elementos de uma amostra enquadram-se na teoria dos valores extremos, a qual é muito utilizada para avaliar a significância estatística de alinhamentos oriundos de bancos de dados (PALAGI *et al.*, 2006; FORNER *et al.*, 2007). Embora existam três diferentes formas assintóticas de distribuição probabilística (GUMBEL, 1958; KENDALL *et al.*, 1977; ALDOUS, 1989), as distribuições de Gumbel e Frechet não se enquadram na análise, já que não existem valores que ultrapassem os limites de dissimilaridade 0 e 1. Sendo assim, o perfil de distribuição dos valores extremos de dissimilaridade entre os espectros deve seguir a distribuição probabilística de Weibull.

Para a determinação dos parâmetros da distribuição de Weibull, foram utilizados 2.000 espectros conhecidos com score acima de 45, em buscas via MASCOT. Cada espectro conhecido foi alinhado a outros de uma biblioteca de 6.000 espectros de queratina e tripsina e a distribuição cumulativa dos scores das medidas de distância foi computada. A

curva obtida seguiu a distribuição de Weibull (Figura 3) e a partir de sua equação (2) os parâmetros de distribuição foram obtidos ( $\alpha=1,20$ ;  $\beta=8,87$ ).

$$F(D_i) = 1 - \exp(-\alpha D_i^\beta) \quad (2)$$

Finalmente, a partir da distribuição obtida, foi possível estabelecer a variável de controle de filtragem ao Eagle Eye, o limite ou valor  $p$ , o qual é utilizado pelo programa para a determinação da medida de dissimilaridade limite ( $D_p$ ) para remoção de um espectro. Este valor indica a fração de espectros não interferentes que serão removidos pelo programa. Ou seja, se compararmos 10.000 espectros possuindo a mesma carga e razão  $m/z$  de um espectro da biblioteca será esperado a remoção de 10 espectros não interferentes quando ajustado um valor  $p$  igual a 0,001. A estratégia de remoção de espectros está ilustrada na Figura 4.

#### 4.2. Desempenho na remoção de espectros interferentes de amostras de *A. angustifolia*

Para avaliação do desempenho do programa Eagle Eye, foi criada uma biblioteca de espectros interferentes oriunda de 4 digestões independentes de géis de poliacrilamida acrescido de espectros provenientes de auto digestões de tripsina em solução. Sendo assim, a biblioteca utilizada nesta simulação continha 10.348 espectros redundantes de fragmentos de queratina e tripsina. O desempenho de filtragem foi avaliado pelo resultado da seleção de espectros de cinco grupos distintos: 1.659 espectros interferentes, provenientes de outra digestão de gel de poliacrilamida, e quatro grupos de 100 espectros não interferentes, ou espectros modelo, provenientes de “spots” digeridos de géis de embriões de *A. angustifolia*, porém classificados de acordo com os escores peptídicos atribuídos pelo MASCOT. Ressalta-se que, após a utilização do programa, a informação referente à qualidade dos espectros modelo foi facilmente recuperada, já que a mesma foi adicionada ao nome dos arquivos.

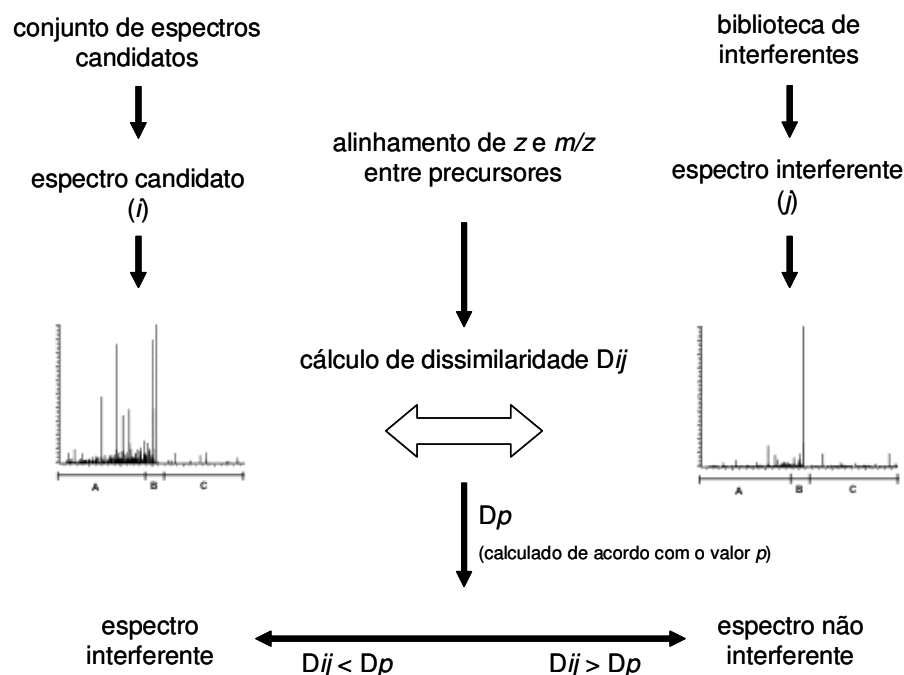


Figura 4: Estratégia de filtragem de espectros candidatos. Para cada espectro candidato, o programa Eagle Eye seleciona um espectro interferente de mesma massa e carga e computa a dissimilaridade entre os mesmos ( $D_{ij}$ ). Posteriormente, o programa calcula a dissimilaridade limite ( $D_p$ ), utilizando a equação de distribuição probabilística e baseando-se no valor  $p$  definido pelo usuário. Finalmente, a medida de dissimilaridade obtida ( $D_{ij}$ ) é comparada com aquela calculada pelo valor  $p$  selecionado ( $D_p$ ) e, se menor, o espectro é removido.

O programa apresentou alta eficiência e precisão de filtragem em simulações variando-se o valor  $p$  e a tolerância a erros na massa dos precursores (Figura 5). Para um valor de  $p$  igual a 0,01, apenas um espectro de baixa qualidade foi removido (escore entre 25 e 40), enquanto que mais de 90% de espectros interferentes foram filtrados. O aumento no valor  $p$ , até seu valor máximo igual 1, resulta em aumento na fração de espectros de queratina e tripsina removidos; porém, também implica na remoção de espectros não interferentes de boa qualidade. Ressalta-se que, em princípio, o ajuste do valor  $p$  igual 1 deveria resultar na remoção de todos os espectros interferentes; entretanto, isso não ocorre já que a estratégia de filtragem adotada é baseada em dissimilaridades e resulta em remoção de somente aqueles que possuem as mesmas massas e razão  $m/z$  em relação aos íons precursores de espectros presentes na biblioteca. Consequentemente, menos de 100% dos espectros interferentes foram removidos quando ajustado o valor de  $p$  igual a 1. De

maneira semelhante, aumentos na tolerância a erros na massa dos precursores também resultaram, embora em menor proporção, em aumento no número de espectros removidos (Figura 5B).

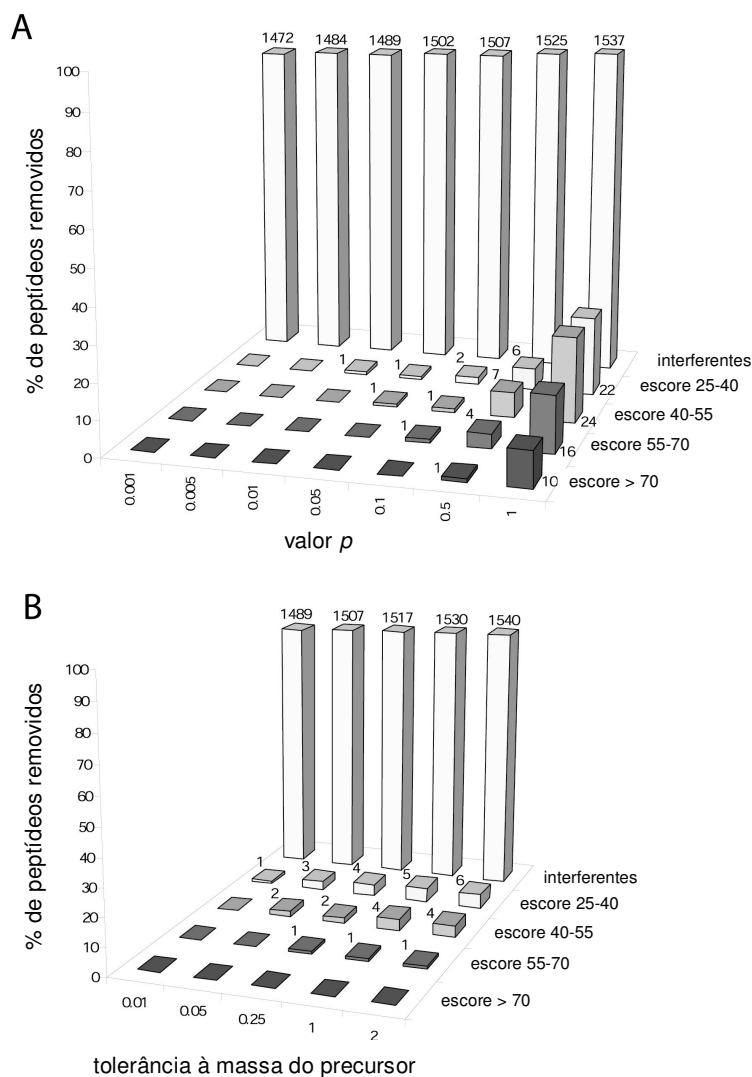


Figura 5: Avaliação do desempenho de filtragem de espectros pelo programa Eagle Eye utilizando amostras embrionárias de *A. angustifolia* e variando-se o valor  $p$  (**A**) e a tolerância à massa do precursor (**B**). O número total de espectros interferentes foi de 1.659. Quatrocentos espectros candidatos foram agrupados em quatro classes de acordo com sua qualidade. Os valores absolutos estão indicados próximos às barras. No painel **A**, a tolerância à massa do precursor e dos fragmentos foi ajustada em 0,01 e 0,6 Da, respectivamente. No painel **B**, o valor  $p$  foi ajustado em 0,01 e a tolerância à massa dos fragmentos em 0,6 Da.

#### 4.3. Desempenho na identificação de proteínas de embriões de *A. angustifolia* por buscas restritivas (MASCOT)

O desempenho do programa Eagle Eye na remoção de espectros de baixa qualidade, potencialmente geradores de falso-positivos e interferentes foi avaliado para a identificação de proteínas em buscas restritivas no banco MSDB, pelo MASCOT. Para tanto, dez “spots” de diferentes massas moleculares relativas foram excisados, digeridos e o número de espectros obtidos antes e após filtragem foi avaliado, utilizando o valor  $p$  igual a 0,01 e tolerância de erros na massa do precursor e fragmentos igual a 0,01 e 0,6 Da, respectivamente. Além disso, para determinação do desempenho do programa na redução do número de falso-positivos, o número de identificações também foi computado em buscas em banco randômico DECOY.

Como esperado, foi observada uma redução no número de espectros após filtragem em todas as avaliações (Tabela 1). A porcentagem de remoção de espectros apresentou uma média de 62% e variou entre 73% e 37%. Tal fato deve-se, possivelmente, aos diferentes graus de contaminação por manipulação durante a digestão ou presentes na própria malha de poliacrilamida entre as amostras de um gel 2-D. Embora tenham sido observados diferentes desempenhos de retenção de espectros entre as amostras protéicas de embriões em *A. angustifolia*, não foram detectadas reduções no número de identificações obtidas em buscas no banco MSDB, resultando em uma relação constante e igual a 1 entre o número de identificações antes e após a seleção. Ressalta-se que houve redução no número de identificações falso-positivas após utilização do programa Eagle Eye. Oito dos dez “spots” avaliados apresentaram uma redução de 4 à 15 vezes no número de identificações falso-positivas, sendo que dois “spots” não apresentaram nenhum alinhamento significativo, ou seja, três peptídeos de escore acima de 20 ou pelo menos um acima de 50, quando avaliados em um banco de sequências aleatórias, baseada no MSDB.

Tabela 1: Desempenho de programa Eagle Eye na seleção de espectros interferentes e obtenção de identificações positivas em amostras de embriões de *A. angustifolia*.

“Spots”	Número de espectros		% remoção	Identificação MSDB <sup>a</sup>	Identificação DECOY <sup>b</sup>	
	Antes seleção	Após seleção			Relação antes / após seleção	Antes seleção
1	1869	694	63	1	18	5
2	1821	631	65	1	7	3
3	1436	560	61	1	16	3
4	858	299	65	1	9	2
5	1746	1102	37	1	17	11
6	1633	545	67	1	14	2
7	1862	676	64	1	9	0
8	1777	572	68	1	15	1
9	1794	483	73	1	10	0
10	1555	616	60	1	8	1

a: relação do número de identificações obtidas antes e após seleção de espectros no banco MSDB pelo MASCOT

b: relação do número de identificações obtidas antes e após seleção de espectros no banco de sequências randômicas DECOY pelo MASCOT

#### 4.4. Desempenho na identificação de proteínas de embriões de *A. angustifolia* por similaridade de sequências (MS BLAST)

Estratégias de identificação de proteínas por similaridade de sequências possuem duas grandes dificuldades relacionadas à existência de espectros interferentes (WARINDEL *et al.*, 2007). A primeira refere-se às necessidades computacionais extremas para processamento de grandes arquivos, contendo espectros que, na maioria das vezes, não resultam em identificação. A outra refere-se a alinhamentos significativos, cuja relação com fragmentos de queratina ou tripsina requer inspeção manual de cada alinhamento obtido pelo MS BLAST com a sequência completa da proteína identificada. Tais alinhamentos são denominados falso-positivos.

Para avaliar a performance do Eagle Eye em buscas por similaridade de sequências, proteínas de um “spot” de 42 kDa de gel 2-D em embriões maduros de *A. angustifolia* foram digeridas, os espectros obtidos sequenciados automaticamente e analisados por MS BLAST. Os alinhamentos positivos e limítrofes (HABERMANN *et al.*, 2004) foram contabilizados quanto ao número de identificações positivas, número de falso-positivos e número de identificações interferentes (fragmentos de queratina ou tripsina).

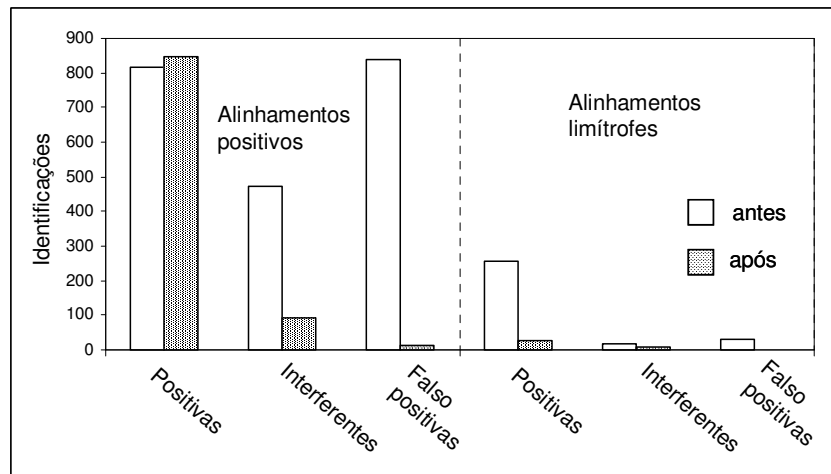


Figura 6: Gráficos representativos da distribuição das identificações obtidas por MS BLAST. Foram consideradas identificações positivas aquelas que resultaram em alinhamentos com actinas de diferentes espécies de plantas. As identificações interferentes relacionam-se a descrições de fragmentos de queratina e tripsina e as identificações falso-positivas referem-se a anotações cuja relação com fragmentos de queratina e tripsina são identificados somente após visualização manual.

A análise via MASCOT das proteínas contidas no “spot” do embrião resultou em alinhamentos com diferentes actinas de plantas, sendo que a identificação mais significativa foi o acesso Q9SPI7 referente à espécie *Picea rubens*. O sequenciamento automatizado seguido de busca por MS BLAST confirmou a identificação e resultou em inúmeros alinhamentos com outras proteínas homólogas de espécies de plantas, animais, fungos e bactérias. Além disso, 470 identificações foram relacionadas a fragmentos de queratina e tripsina; enquanto 838 alinhamentos correspondiam a sequências de queratina e tripsina, porém com outra anotação (identificações falso-positivas).

Conforme esperado, após análise no Eagle Eye utilizando valor  $p$  de 0,01, o número de alinhamentos interferentes e de falso-positivos foi reduzido de 470 para 92 e de 838 para 15, respectivamente (Figura 6). Como consequência, o tempo de envio e processamento dos dados foi reduzido drasticamente de 38 para 08 min. Vale ressaltar que, nesse caso, o número de identificações positivas aumentou após filtragem. Tal fato pode estar relacionado com o processamento incorreto do MS BLAST de sequências positivas em alinhamentos de queratina e tripsina ou limítrofes.

## **5. Considerações finais**

O programa Eagle Eye mostrou-se eficiente na remoção de espectros interferentes sem comprometer a identificação de proteínas alvo. Devido à alta especificidade utilizada em buscas restritivas, o programa não é essencial para identificação de componentes protéicos principais. Entretanto, em análises descritivas do proteoma e/ou dependentes de identificação de proteínas por similaridade de sequências ou busca tolerante a erros, o Eagle Eye é fundamental, devido ao maior universo de organismos e os diferentes padrões de alinhamentos utilizados nessas abordagens.

O algoritmo desenvolvido baseia-se somente na dissimilaridade de espectros, independente do padrão de fragmentação e da sequência do peptídeo. Portanto, ele pode ser utilizado para filtragem de espectros oriundos de diferentes espectrômetros de massas, excetuando-se naqueles cujas identificações baseiam-se em íons mono carregados devido ao modelo de alinhamento utilizado.

Ajustes nos parâmetros de filtragem podem ser facilmente realizados por alterações na tolerância a erros na massa dos íons precursores e pelo valor  $p$ . Simulações realizadas a partir de proteínas digeridas *in gel* em embriões de *A. angustifolia* resultaram em redução no tempo de processamento computacional, na quantidade de identificações de queratinas e tripsinas e de falso-positivos, sem afetar o número de identificações positivas e de peptídeos de alta qualidade em buscas por MASCOT e MS BLAST.

## **6. Referências bibliográficas**

- Aldous, D. J. Probability approximations via the poisson clumping heuristic. New York: Springer. 1989. 45 p.
- Forner, F., L. J. Foster, S. Toppo. Mass spectrometry data analysis in the proteomics era. Current Bioinformatics, v.2, p.63-93. 2007.
- Frank, A., P. Pevzner. PepNovo: De novo peptide sequencing via probabilistic network modeling. Analytical Chemistry, v.77, p.964-973. 2005.
- Gumbel, E. J. Statistics of Extremes. New York: Columbia University Press. 1958. 156 p.

- Habermann, B., J. Oegema, S. Sunyaev, A. Shevchenko. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. Molecular & Cellular Proteomics, v.3, p.238-249. 2004.
- Johnson, R. S., M. T. Davis, J. A. Taylor, S. D. Patterson. Informatics for protein identification by mass spectrometry. Methods, v.35, p.223-2236. 2005.
- Katz, A., P. Waridel, A. Shevchenko, U. Pick. Salt-induced changes in the plasma membrane proteome of the halotolerant alga *Dunaliella salina* as revealed by blue-native gel electrophoresis and nanoLC-MS/MS. Molecular & Cellular Proteomics, v.6, p.1459-1472. 2007.
- Kendall, M. G., A. Stuart, J. K. Ord. The Advanced Theory of Statistics. New York: Oxford University Press. 1977. 347 p.
- Liska, A. J., A. V. Popov, S. Sunyaev, P. Coughlin, B. Habermann, A. Shevchenko, P. Bork, E. Karsenti, A. Shevchenko. Homology-based functional proteomics by mass spectrometry: Application to the *Xenopus* microtubule-associated proteome. Proteomics, v.4, p.2707-2721. 2004.
- Ong, S. E., M. Mann. Mass spectrometry-based proteomics turns quantitative. Nature Chemical Biology, v.1, p.252-262. 2005.
- Palagi, P. M., P. Hernandez, D. Walther, R. D. Appel. Proteome informatics I: bioinformatics tools for processing experimental data. Proteomics, v.6, p.5435-5444. 2006.
- Schlosser, A., R. Volkmer-Engert. Volatile polydimethylcyclsiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry. Journal of Mass Spectrometry, v.38, p.523-525. 2003.
- Shevchenko, A., M. M. L. De Sousa, P. Waridel, S. T. Bittencourt, M. V. De Sousa, A. Shevchenko. Sequence similarity-based proteomics in insects: Characterization of the larvae venom of the Brazilian moth *Cerodirphia speciosa*. Journal of Proteome Research, v.4, p.862-869. 2005.
- Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen, M. Mann. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature Protocols, v.1, p.2856-2860. 2006.
- Shevchenko, A., M. Wilm, O. Vorm, M. Mann. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Analytical Chemistry, v.68, p.850-858. 1996.
- Waridel, P., A. Frank, H. Thomas, V. Surendranath, S. Sunyaev, P. Pevzner, A. Shevchenko. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. Proteomics, v.7, p.2318-2329. 2007.

## **CAPÍTULO 4**

## **CAPÍTULO 4. Caracterização do proteoma de *Araucaria angustifolia* durante o desenvolvimento da semente<sup>3</sup>**

### **1. Resumo**

O objetivo desse trabalho foi caracterizar e identificar alterações no proteoma durante o desenvolvimento de sementes *A. angustifolia*. Para tanto, as proteínas presentes nas sementes foram estudadas por eletroforese bidimensional (2-DE) e espectrometria de massas, segundo as metodologias estabelecidas nos capítulos anteriores. A análise do teor de proteínas mostrou um aumento a partir do início do desenvolvimento cotiledonar. Não foram observadas diferenças marcantes no perfil de acúmulo de proteínas por 2-DE durante o início e final da embriogênese, sendo que as maiores diferenças foram observadas no estágio cotiledonar inicial. A estratégia de identificação protéica permitiu a identificação de 96 proteínas, que foram agrupadas de acordo com sua função biológica e o padrão de acúmulo durante o desenvolvimento embrionário. A análise das proteínas abundantes no início da embriogênese indica um maior controle no metabolismo oxidativo em relação aos estádios finais. Por outro lado, o final da embriogênese é caracterizado por um alto metabolismo de assimilação de carbono e acúmulo de proteínas de reserva.

### **2. Introdução**

O desenvolvimento das sementes é um dos processos mais estudados dentro da fisiologia vegetal. De acordo com sua tolerância à desidratação, as sementes podem ser agrupadas em dois tipos (ROBERTS, 1973). O primeiro grupo abrange as sementes ortodoxas que adquirem tolerância à desidratação, podendo ser armazenadas e mantendo a viabilidade por longos períodos. O segundo grupo, abrange as sementes recalcitrantes que não possuem adaptações anatômicas e fisiológicas para o processo de desidratação e, portanto, mantêm-se metabolicamente ativas durante a maturação, não podendo ser

---

<sup>3</sup> Este capítulo foi aceito para publicação no periódico Journal of Proteomics, doi: 10.1016/j.jprot.2009.01.011

armazenadas por longos períodos de tempo (ROBERTS, 1973; ANGUELOVA-MERHAR *et al.*, 2003).

Em sistemas biológicos complexos, diversos fatores podem afetar a dinâmica de expressão dos genes e a biossíntese de proteínas. O estudo comparativo do proteoma de um organismo em diferentes estádios de desenvolvimento é uma ferramenta importante para a determinação de funções biológicas desempenhadas por proteínas diferencialmente expressas. Entretanto, embora a proteômica possua o potencial de gerar uma grande quantidade de informações sobre o desenvolvimento embrionário, o único trabalho abordando aspectos do desenvolvimento da semente em uma espécie recalcitrante foi aquele realizado em *Fagus sylvatica*, durante a quebra da dormência (PAWLOWSKI, 2007).

De maneira semelhante, poucos são os estudos abordando alterações no proteoma durante a embriogênese zigótica em espécies de gimnospermas. O primeiro trabalho descrito investigou a dinâmica de acúmulo de proteínas de reserva em sementes de *Picea glauca* (FLINN *et al.*, 1991), enquanto que a caracterização do acúmulo de proteínas durante a embriogênese foi descrito para *Cupressus sempervirens* (SALLANDROUZE *et al.*, 2002). SILVEIRA *et al.* (2008) identificaram quatro proteínas envolvidas no desenvolvimento embrionário em *A. angustifolia* e DOS SANTOS *et al.* (2006) estudaram o acúmulo das quitinases e arabinogalactanos durante esse processo.

O objetivo desse trabalho foi caracterizar e identificar alterações no proteoma durante o desenvolvimento de sementes de *A. angustifolia* através da eletroforese bidimensional (2-DE) e espectrometria de massas, considerando as adequações metodológicas já estabelecidas nos capítulos anteriores.

### **3. Materiais e Métodos**

#### *3.1. Material vegetal*

Sementes de *A. angustifolia* foram mensalmente coletadas no município de Bom Retiro-SC (27°47'S e 49°29'W), entre dezembro de 2006 e maio de 2007. As sementes foram armazenadas à -80 °C até a realização das análises. Embriões nos estádios pró-

embrionário, globular, torpedo, cotiledonar inicial, cotiledonar tardio e maduro foram identificados de acordo com a classificação proposta por ASTARITA *et al.* (2003) (Figura 1). Devido ao pequeno tamanho e baixo conteúdo protéico, os pró-embriões, embriões globulares e torpedos foram processados em conjunto com os gametófitos femininos. Embriões nos estádios cotiledonar inicial, cotiledonar tardio e maduro foram isolados dos gametófitos, e as proteínas extraídas separadamente.

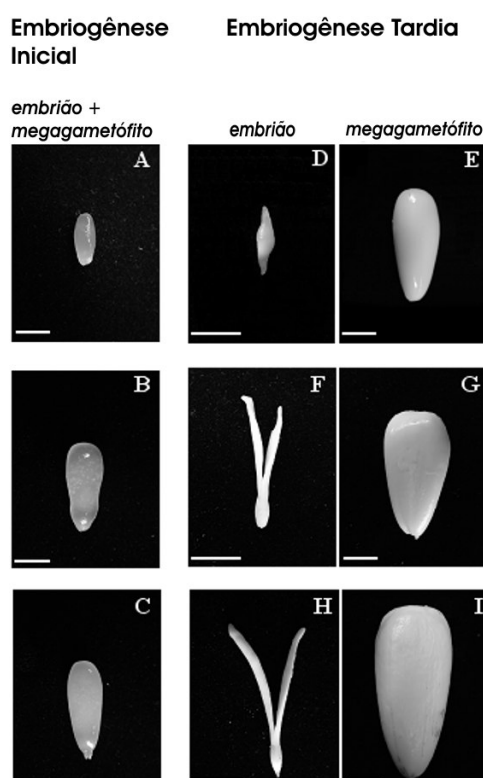


Figura 1: Embriões e megagametófitos de *A. angustifolia*. Megagametófitos contendo pró-embriões (A), embriões no estágio globular (B) e estágio torpedo (C). Embriões e megagametófitos nos estádios cotiledonar inicial (D, E), cotiledonar tardio (F, G) e maduro (H, I). Barras: 1,25 cm.

### 3.2. Extração de proteínas

Extratos protéicos foram preparados em três repetições biológicas. Devido às diferenças nas massas frescas entre os diferentes materiais estudados (Tabela 1), cada repetição foi composta por 15 sementes.

Os tecidos analisados foram pulverizados em nitrogênio líquido e as proteínas extraídas em solução contendo 7 M de uréia, 2 M de tiouréia, 1% de ditiotretitol (DTT), 2% de

Triton-100, 1 mM de fluoreto de fenilmetilsulfonil (PMSF) e 5  $\mu$ M de pepstatina, como descrito no método de extração em solução de uréia-tiouréia (Capítulo 2). Cada amostra foi agitada por 15 min, centrifugada por 5 min (12000 *g* à 4°C) e o sobrenadante coletado. Posteriormente, as proteínas foram precipitadas durante 1 h em solução de acetona contendo 20% TCA e 0,2% de DTT. Após centrifugação (12000 *g* à 4°C) por 30 min, as proteínas foram lavadas por três vezes em acetona gelada contendo 0,2% de DTT. Finalmente, as proteínas foram ressuspensas em tampão contendo 7 M de uréia, 2 M de tiouréia, 1% de DTT, 2% de Triton-100, 2% de IPG Buffer pH 3-10 (GE Healthcare, Little Chalfort, England), 1 mM de PMSF e 5  $\mu$ M de pepstatina. A quantidade de proteínas em cada extrato foi determinada pelo 2-D Quant Kit (GE Healthcare), contendo albumina sérica bovina como padrão.

Tabela 1: Massa fresca (g) dos diferentes materiais durante o desenvolvimento da semente de *A. angustifolia*. Estádios embrionários: pró-embrião (1), globular (2), torpedo (3), cotiledonar inicial (4), cotiledonar tardio (5), maduro (6).

Material	Estádios embrionários					
	1	2	3	4	5	6
<i>Semente</i>	1,51 $\pm$ 0,35	3,66 $\pm$ 0,24	4,23 $\pm$ 0,47	9,24 $\pm$ 0,91	9,58 $\pm$ 0,85	11,80 $\pm$ 0,31
<i>Embrião + Megagametófito</i>	0,12 $\pm$ 0,05	0,53 $\pm$ 0,04	0,56 $\pm$ 0,06	-	-	-
<i>Embrião</i>	-	-	-	0,05 $\pm$ 0,01	0,16 $\pm$ 0,02	0,20 $\pm$ 0,03
<i>Megagametófito</i>	-	-	-	2,60 $\pm$ 0,20	4,96 $\pm$ 0,56	7,47 $\pm$ 0,27

### 3.3. Eletroforese bidimensional

As amostras foram diluídas em tampão de reidratação, contendo 7 M de uréia, 2 M de tiouréia, 2% de CHAPS, 2% de IPG Buffer pH 4-7, 1% de DTT e 0,002% de azul de bromofenol, até a concentração de 570  $\mu$ g de proteínas por 1 mL de extrato. Para cada gel, foram utilizados 200  $\mu$ g de proteínas, carregados simultaneamente à reidratação em tira de gradiente de pH imobilizado (IPG, GE Healthcare). Foram utilizadas tiras de IPG de 18 cm,

com faixa de separação linear de pH 4-7. A reidratação foi realizada por 12 h e a focalização isoeletrica foi realizada em uma unidade de IPG Phor II (GE Healthcare) à 20°C, com limite de corrente de 50 µA por tira, segundo as etapas: 0,5 kVh à 500 V, 8 h à 1000 V, 3 h à 8000 V, 20 kVh à 8000 V. Antes de iniciar a segunda dimensão, as proteínas foram reduzidas, por 15 min, em solução de equilíbrio (50 mM de Tris-HCl pH 8,8, 6 M de uréia, 30% de glicerol, 2% de SDS, 0,002% de azul de bromofenol) adicionado de 125 mM de DTT e, posteriormente, alquiladas (solução de equilíbrio adicionada de 125 mM de iodoacetamida). A segunda dimensão foi realizada em gel de 12% de poliacrilamida, em condições desnaturantes (SDS-PAGE). A separação eletroforética foi conduzida com corrente constante de 25 mA por gel, em um sistema vertical Protean II (BioRad, Hercules, EUA).

#### 3.4. Coloração dos géis

As proteínas foram visualizadas pela coloração com nitrato de prata ((SHEVCHENKO *et al.*, 1996). Inicialmente, os géis bidimensionais (2-D) foram fixados por 12 h em solução de fixação contendo 50% de metanol, 12% de ácido acético e 0,05% de formaldeído. Após duas lavagens de 30 min em solução de 50% de etanol, os géis foram incubados por 2 min em solução de 0,02% de tiosulfato de sódio e, posteriormente, lavados em água. Em seguida, os géis foram incubados por 20 min em solução de 0,2% de nitrato de prata e 0,08% de formaldeído. Após lavagem em água, a revelação foi conduzida por 3 min pela incubação dos géis em solução de 6% de carbonato de sódio, 0,16% de tiosulfato de sódio e 0,05% de formaldeído e interrompida utilizando a mesma solução de fixação utilizada no início do processo de coloração.

#### 3.5. Análise de imagens

Os géis 2-D foram analisados pelo programa Image Master Platinum 6.0 (GE Healthcare). Os parâmetros para detecção automatizada de “spots” foram otimizados através de análise em diferentes regiões dos géis 2-D. Posteriormente, o contorno de cada “spot” foi avaliado manualmente para isolamento, remoção ou adição dos mesmos, caso necessário. Para a caracterização dos géis, mapas 2-D foram obtidos, contendo apenas os

“spots” detectados em todas as repetições. Para tanto, os géis analisados foram alinhados visando atribuir identidades comuns entre os “spots” das diferentes repetições. Os mapas criados foram utilizados na caracterização do perfil eletroforético das proteínas detectadas.

### 3.6. Agrupamentos de expressão

Para identificar padrões similares de acúmulo de proteínas durante a embriogênese zigótica, foi utilizado o programa EPCLUST, disponível no sítio [www.bioinfo.ebc.ee/EP/EP/EPCLUST/](http://www.bioinfo.ebc.ee/EP/EP/EPCLUST/). Para tanto, as porcentagens dos volumes de cada “spot” dos diferentes acessos detectados nos mapas 2-D, nos embriões e megagametófitos, foram somados e convertidos a valores logarítmicos ( $\log_2$ ). Distâncias euclidianas e o algoritmo UPGMA (“unweighted pair group method with arithmetic mean”) foram utilizados nos agrupamentos.

### 3.7. Digestão de proteínas

A digestão de proteínas *in gel* foi realizada conforme descrito por SHEVCHENKO *et al.* (2006). Para tanto, cada “spot” foi isolado dos géis 2-D, cortados em pequenos pedaços e desidratados pela adição de acetonitrila por 10 min. Posteriormente, as proteínas foram reduzidas e alquiladas em soluções de 100 mM de bicarbonato de amônio contendo 10 mM de DTT, por 30 min à 56° C, e 55 mM de iodoacetamida, por 10 min, em temperatura ambiente no escuro, respectivamente. A digestão enzimática foi realizada através de adição de 16 ng.L<sup>-1</sup> tripsina (grau de sequenciamento, Promega, Madison, US), à 37° C, por 12 h. Os peptídeos foram extraídos em solução contendo 5% de ácido fórmico e acetonitrila (1:2, v/v) por 15 min à 37° C. O sobrenadante foi coletado, seco e os peptídeos ressuspensos em 0,05% de ácido trifluoroacético (TFA).

### 3.8. Fracionamento e análise dos peptídeos em espectrômetro de massas

As misturas peptídicas, resultantes do processo de digestão, foram analisadas por nanoLC-MS/MS em um cromatógrafo líquido Ultimate 3000 (Dionex, Sunnyvale, US) acoplado a um espectrômetro de massas do tipo LTQ-Orbitrap (ThermoFisher Scientific,

Waltham, US), segundo o procedimento descrito no Capítulo 3. Para tanto, os peptídeos tripticos foram ressolubilizados em 0,05% de TFA e eluídos em uma pré-coluna de carregamento C18 PepMAP100 (Dionex, 5 x 300 mm, 5  $\mu$ m), sob um fluxo de 20  $\mu$ L/min.

Após o carregamento, concentração e lavagem, os peptídeos foram separados em cromatografia líquida de alta eficiência (CLAE) utilizando uma nano-coluna C18 PepMAP100, 3  $\mu$ m (Dionex, 15 cm x 75  $\mu$ m). Foram utilizados como solventes uma solução de 5% acetonitrila e 1% de ácido fórmico (solvente A), e uma solução de 80% acetonitrila e 1% de ácido fórmico (solvente B). A separação foi realizada sob um fluxo de 200 nL/min e a mudança na proporção, do solvente B em relação ao solvente A, definiu o gradiente de corrida, conforme descrito a seguir: de 5 à 20% nos primeiros 20 min, de 20 à 50% de 20 à 36 min, de 50 à 100% de 36 à 41 min, 100% de 41 à 51 min e 5% entre 51 e 56 min. Após a separação em CLAE, os peptídeos foram injetados em tandem no espectrômetro de massas utilizando um sistema de “nanospray” (ThermoFisher Scientific) equipado com uma agulha de sílica de 10  $\mu$ m de diâmetro interno (New Objective, Woburn, US). A voltagem do spray utilizada foi de 1,8 kV e a temperatura do capilar de transferência iônica foi ajustada à 180 °C. O espectrômetro de massas operou no modo dependente de dados (DDA). O controle de ganhos automáticos (AGC) foi ajustado à  $5 \times 10^5$  e  $5 \times 10^4$  para os eventos de fragmentação MS/MS no quadrupolo triplo linear (LTQ). Cada ciclo de aquisição de dados consistiu em uma varredura de moléculas entre 200-2000  $m/z$  no analisador Orbitrap, seguido de fragmentação MS/MS, sob energia de colisão de 35% na célula de colisão linear, dos quatro íons precursores mais intensos. Íons de carga igual a +1 foram excluídos dos eventos de MS/MS, e as  $m/z$  dos íons precursores fragmentados foram inseridos em uma lista de exclusão por 90 segundos.

### *3.9. Identificação das proteínas*

Os espectros obtidos em cada ciclo de análise foram convertidos em um único arquivo tipo .mgf, utilizando o programa BioWorks 3.2 (ThermoFisher Scientific). Previamente à identificação de proteínas em bancos de dados, os espectros foram filtrados para remoção de espectros de queratina e tripsina utilizando o programa Eagle Eye, como

descrito no Capítulo 3. Foi utilizada a própria biblioteca disponível no sítio <http://genetics.bwh.harvard.edu/cgi-bin/msfilter/eagleeye.cgi>, contendo 12.009 espectros interferentes. Os parâmetros de seleção utilizados foram: tolerância a erros na massa dos fragmentos igual a 0,6 Da, tolerância a erros na massa dos precursores igual a 0,01 Da e valor  $p$  igual a 0,01 (Figura 2).

As buscas foram realizadas no banco de dados MSDB através do programa MASCOT v.2.2 (Matrix Science, London, UK) instalado em dois servidores locais. As configurações de busca foram: 10 ppm para tolerância de massas do íon precursor e 0,6 Da para os fragmentos. Foi permitido uma falha na fragmentação e carbamidometilação da cisteína, oxidação de metionina e acetilação de N-proteínas foram adicionadas como modificações variáveis. Foram consideradas identificações as proteínas com pelo menos três peptídeos de escore maior que 20. Para identificações a partir de somente um ou dois espectros, ao menos um peptídeo deve possuir escore acima de 50 (Figura 2).

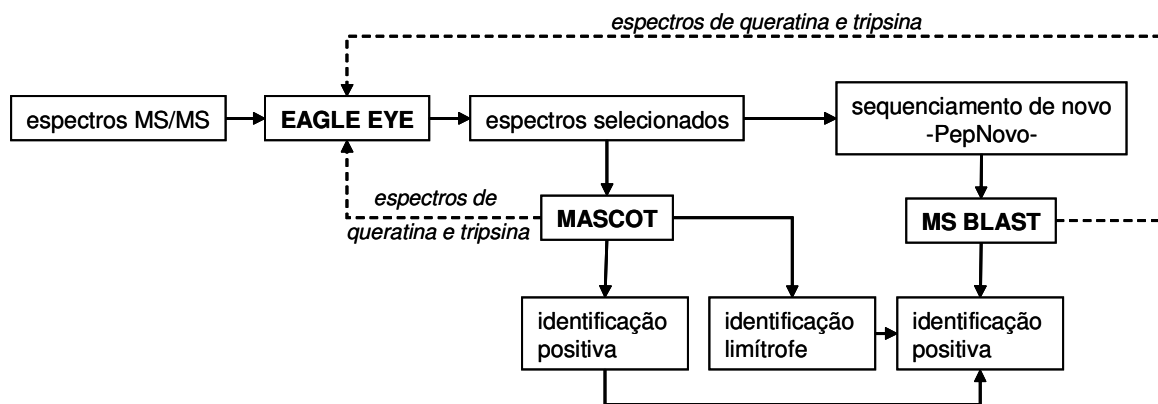


Figura 2: Estratégia de identificação de proteínas de embriões e gametófitos femininos de *A. angustifolia*.

Paralelamente à busca restrigente por MASCOT, o sequenciamento *de novo* automatizado foi realizado (Figura 2). Para tanto, o programa PepNovo foi utilizado e apenas sequências candidatas com escore acima de 6 foram consideradas (FRANK e PEVZNER, 2005). Nas buscas por MS BLAST, todas as sequências peptídicas selecionadas foram combinadas em uma única entrada. As buscas foram realizadas no NCBI através do sítio <http://genetics.bwh.harvard.edu/msblast/>. A confirmação ou identificação de proteínas foi

realizada de acordo com a tabela de identificação do MS BLAST (HABERMANN *et al.*, 2004).

#### 4. Resultados e discussão

##### 4.1. Variação no conteúdo de proteínas e caracterização dos mapas protéicos

Durante o desenvolvimento da semente de *A. angustifolia*, foi observado o mesmo padrão de variação do conteúdo protéico nos tecidos embrionários e gametofíticos. Entretanto, o teor de proteínas nos megagametófitos foi menor em relação àquele observado para os tecidos embrionários. A avaliação do acúmulo de proteínas entre os diferentes estádios de desenvolvimento indicou um aumento de até 17 vezes na concentração de proteínas (Figura 3). Esse aumento é evidente nos embriões em estágio cotiledonar tardio, quando observou-se a maior concentração protéica durante o processo ( $19 \mu\text{g}\cdot\text{mg}^{-1}$  de massa fresca).

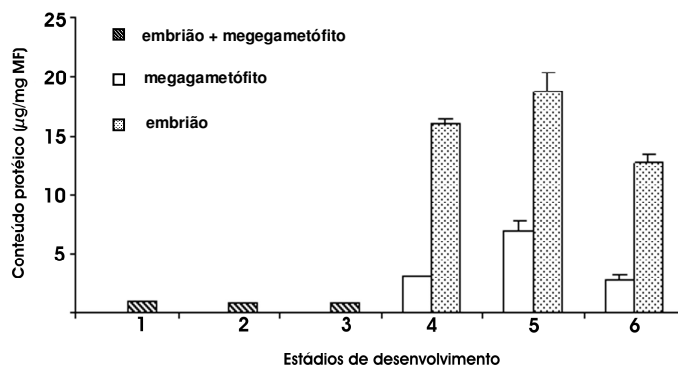


Figura 3: Conteúdo de proteínas de *A. angustifolia* durante a embriogênese zigótica. Estádios de desenvolvimento: **1**: pró-embrião, **2**: globular, **3**: torpeda, **4**: cotiledonar inicial, **5**: cotiledonar tardio, **6**: maduro.

ASTARITA *et al.* (2003) constataram um grande aumento na quantidade de proteínas solúveis acumuladas em tecidos embrionários entre as fases torpeda e cotiledonar inicial. Entretanto, essas proteínas não foram avaliadas qualitativamente. PANZA *et al.* (2002) observaram a presença de vacúolos protéicos em cotilédones e embriões maduros de *A. angustifolia*, indicando que os tecidos embrionários também são responsáveis pelo

acúmulo de proteínas de reserva. Tal fato sugere um aumento e manutenção de alta concentração de proteínas durante maturação. A redução na concentração protéica no embrião maduro de *A. angustifolia* indica que o transiente aumento observado durante o desenvolvimento cotiledonar, seguido de redução no estágio maduro, pode estar relacionado tanto com o acúmulo de proteínas de reserva, como também com um aumento transitório da expressão gênica, tradução protéica e atividade metabólica, já que a mobilização de proteínas de reserva ocorre somente após a germinação (BEWLEY, 1997).

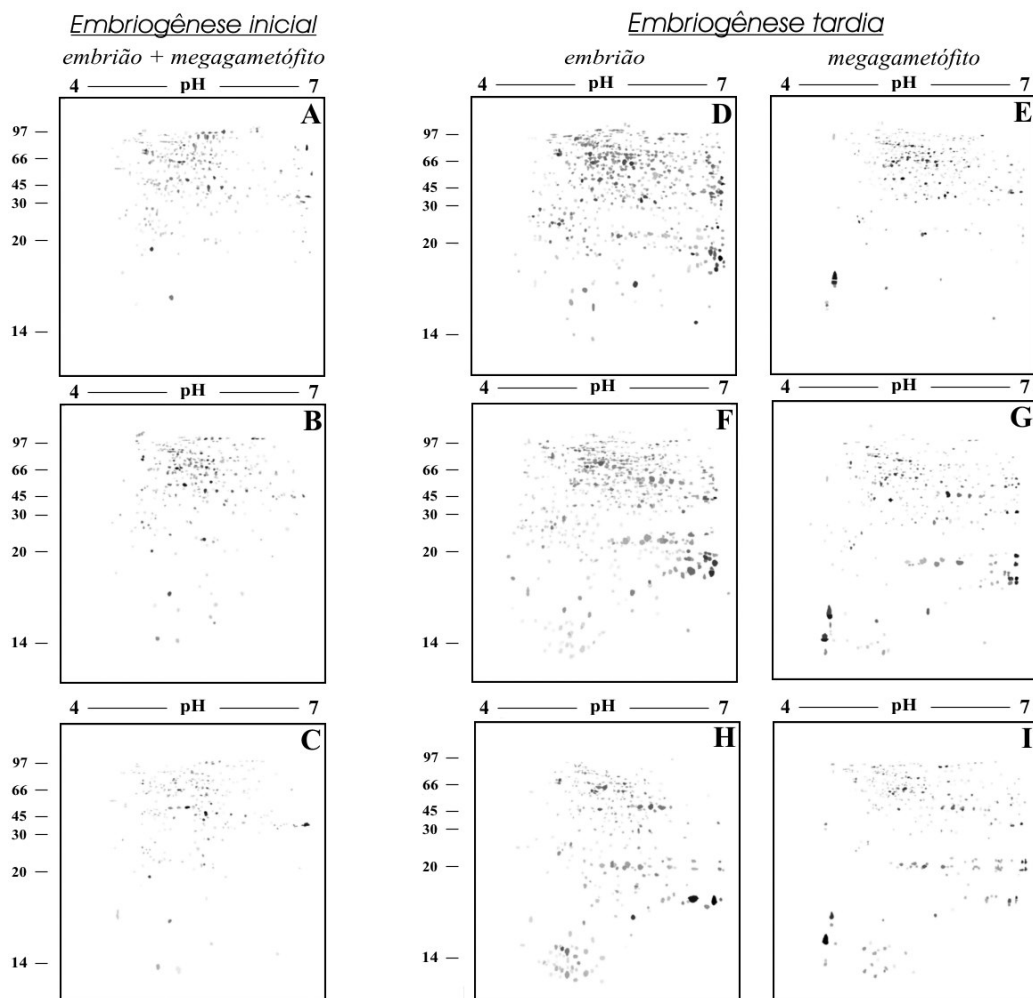


Figura 4: Mapas 2-D das proteínas detectadas durante o desenvolvimento da semente de *A. angustifolia*. Estádios de desenvolvimento: **A**: pró-embrião, **B**: globular, **C**: torpedo, **D** e **E**: cotiledonar inicial, **F** e **G**: cotiledonar tardio, **H** e **I**: maduro. O gradiente de pH utilizado e as massas moleculares relativas (em kDa) estão representados na porção superior e lateral dos mapas.

A análise do mapa 2-D das sementes no estágio cotiledonar inicial mostra a presença de um grande número de proteínas, já que  $955 \pm 21$  “spots” foram detectados no tecido embrionário e  $545 \pm 74$  no megagametófito (Figura 4 e Tabela 2). Embora o final do processo de embriogênese seja caracterizado por reduções nas taxas de divisão celular e início do acúmulo de substâncias de reserva, a alta expressão protéica na fase de diferenciação e desenvolvimento cotiledonar sugere a importância dos eventos que ocorrem nesse período.

A transição entre o estágio cotiledonar tardio e maduro foi caracterizada por uma redução no número de “spots” detectados, semelhante àquela já evidenciada na quantificação de proteínas totais. Esta situação indica uma possível queda na ativação gênica e consequente diminuição na biossíntese de proteínas. Adicionalmente, não foi observada alteração no número de “spots” detectados nos mapas 2-D de tecidos gametofíticos, considerados tecidos de reserva, importantes para o final do processo embriogenético (Tabela 2).

Tabela 2: Número total de “spots” detectados nos géis 2-D durante o desenvolvimento das sementes de *A. angustifolia* (média  $\pm$  DP, n=3). Estádios de desenvolvimento: pró-embrião (1), globular (2), torpedo (3), cotiledonar inicial (4), cotiledonar tardio (5), maduro (6).

<b>Material vegetal</b>	<b>Estádio de desenvolvimento</b>	<b>Número total de “spots”</b>
<i>Embrião + Megagametófito</i>		
	<b>1</b>	426 $\pm$ 42
	<b>2</b>	425 $\pm$ 82
	<b>3</b>	353 $\pm$ 124
<i>Embrião</i>		
	<b>4</b>	955 $\pm$ 21
	<b>5</b>	773 $\pm$ 232
	<b>6</b>	496 $\pm$ 56
<i>Megagametófito</i>		
	<b>4</b>	545 $\pm$ 74
	<b>5</b>	473 $\pm$ 111
	<b>6</b>	427 $\pm$ 93

A 2-DE permite a detecção de propriedades físico-químicas das proteínas fracionadas, como por exemplo a massa molecular relativa (Mr) (CHEN e HARMON, 2006). Durante o início da embriogênese, não foram identificadas alterações nos perfis de distribuição dos “spots” de acordo com a Mr (Figura 5). Foram identificados poucos “spots” com pequena Mr (menores de 20 kDa), sendo que a maior porcentagem ocorre nos

embriões maduros (13%). Ressalta-se que nos estádios iniciais, o material era constituído de embriões e megagametófitos, não tendo sido verificadas alterações na percentagem de “spots” de baixa Mr nessa fase. Entretanto, nos mapas 2-D de tecidos embrionários e megagametofíticos isolados (Figura 5B e C) foram observadas alterações expressivas na classe de “spots” com Mr entre 20 e 40 kDa. Além disso, o aumento na percentagem de “spots” com Mr entre 20 e 40 kDa é mais evidente nos megagametófitos dos estádios cotiledonar inicial, tardio e maduro (9%, 29% e 18%, respectivamente, Figura 5C). Em embriões nos estádios cotiledonar inicial, tardio e maduro (Figura 5B), também foi possível verificar um aumento na percentagem de “spots” de pequena Mr (2%, 6% e 13%), indicando que houve um aumento na quantidade de polipeptídeos de pequena Mr em ambos os tecidos estudados durante a embriogênese.

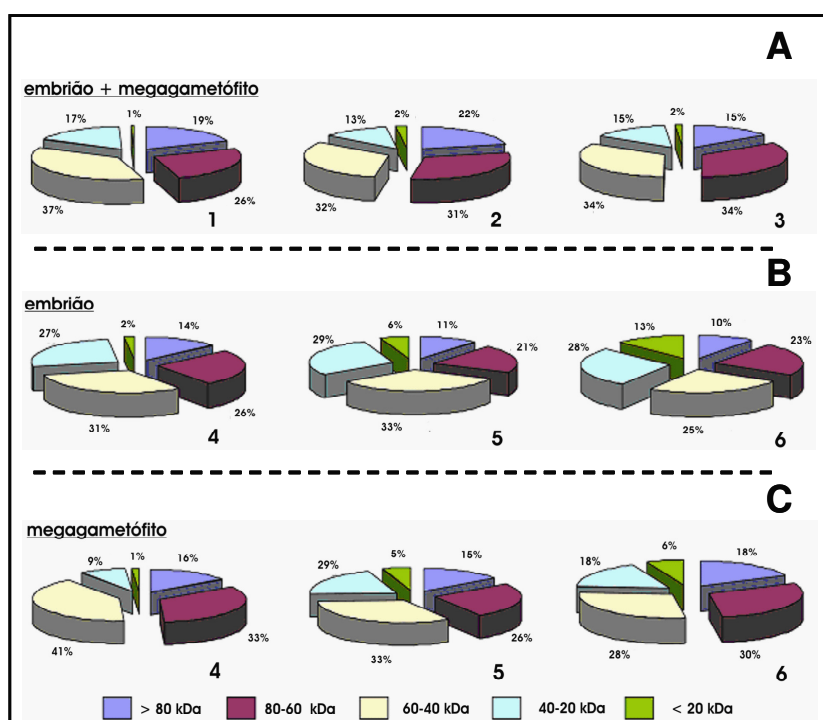


Figura 5: Distribuição dos “spots” detectados nos mapas 2-D, de acordo com a massa molecular relativa (kDa), durante o desenvolvimento das sementes de *A. angustifolia*. Estádios de desenvolvimento: pró-embrião (1), globular (2), torpedo (3), cotiledonar inicial (4), cotiledonar tardio (5), maduro (6).

Visando uma melhor caracterização das variações no padrão de acúmulo de proteínas durante a embriogênese, os mapas 2-D dos diferentes estádios foram alinhados e

o número de spots comuns quantificados. A análise do número de “spots” comuns indicou que as menores alterações no perfil protéico ocorrem durante o início da embriogênese. Os mapas 2-D dos estádios pró-embriônicos e globulares apresentam 24% de “spots” comuns, enquanto que aqueles referentes aos estádios globulares e torpedo apresentaram 23% de “spots” comuns (Tabela 3).

Tabela 3: Tabela comparativa dos “spots” comuns entre os mapas 2-D durante o desenvolvimento das sementes de *A. angustifolia*. Estádios de desenvolvimento: pró-embrião (1), globular (2), torpedo (3), cotiledonar inicial (4), cotiledonar tardio (5), maduro (6).

<b>Material vegetal</b>	<b>Estádios de desenvolvimento</b>	<b>Número de “spots” comuns</b>	<b>Porcentagem de “spots” comuns</b>
<b>Embrião + Megagametófito</b>			
	<b>1-2</b>	203	24
	<b>2-3</b>	179	23
<b>Embrião</b>			
	<b>3-4</b>	168	13
	<b>4-5</b>	414	24
	<b>5-6</b>	235	19
<b>Megagametófito</b>			
	<b>3-4</b>	154	17
	<b>4-5</b>	167	16
	<b>5-6</b>	175	19
<b>Embrião e Megagametófito</b>			
	<b>4</b>	257	17
	<b>5</b>	201	16
	<b>6</b>	258	28

A transição entre o estágio torpedo e cotiledonar tardio é caracterizada pelas maiores alterações no perfil de acúmulo de proteínas. Embora não tenham sido verificadas diferenças na distribuição dos “spots” de acordo com a Mr (Figura 5), os números de “spots” comuns entre os mapas 2-D dos estádios torpedo e cotiledonar inicial foram os menores durante a embriogênese zigótica. Assim, foram detectados 168 e 154 “spots” em relação aos mapas de embriões e megagametófitos dos estádios torpedo e cotiledonar inicial, respectivamente. Tal fato, em associação com os resultados observados quanto às variações no conteúdo protéico durante a embriogênese (Figura 3), indicam uma alteração no perfil metabólico durante a diferenciação e crescimento cotiledonar.

Em coníferas, a interação entre embriões e megagametófitos é de extrema importância no desenvolvimento das sementes (SALLANDROUZE *et al.*, 2002). A análise

comparativa entre os “spots” detectados em embriões e megagametófitos apresenta variações importantes de acordo com o estágio de desenvolvimento avaliado (Tabela 3). Nos mapas dos estádios cotiledonar inicial e tardio ocorrem as menores porcentagens de “spots” comuns (17 e 16%, respectivamente). Entretanto, pela análise comparativa entre os mapas de embriões e megagametófitos do estágio maduro foi verificada uma grande porcentagem de “spots” comuns (28%). Esta situação indica que os possíveis processos de diferenciação e crescimento celular no embrião, e a sua interação com o tecido materno estejam sendo finalizados.

## *4.2. Alterações metabólicas durante e embriogênese zigótica*

### *4.2.1. Identificação de proteínas*

Visando a identificação de alterações no metabolismo de *A. angustifolia* durante a embriogênese zigótica, foram selecionados os 155 “spots” mais abundantes em % de volume dos mapas 2-D, presentes nos estádios globular e maduro. Seus correspondentes nos géis 2-D foram excisados, digeridos e analisados em espectrômetro de massas (Figura 6).

Para a identificação das proteínas foram utilizadas buscas seriadas por MASCOT e por sequenciamento *de novo* automatizado seguido de alinhamentos no MS BLAST. A estratégia adotada mostrou-se bem sucedida, já que *A. angustifolia* é um organismo cujo genoma é desconhecido e para a qual existem somente 6 proteínas depositadas nos bancos NCBI e SwissProt. Foram obtidas identificações de 56 “spots” nos estádios globular e 81 “spots” no maduro. Destes, 15 do globular e 40 do maduro foram exclusivos dos respectivos estádios. Esta situação torna estes materiais potenciais candidatos a marcadores do início e final do desenvolvimento embrionário (Tabela 4).

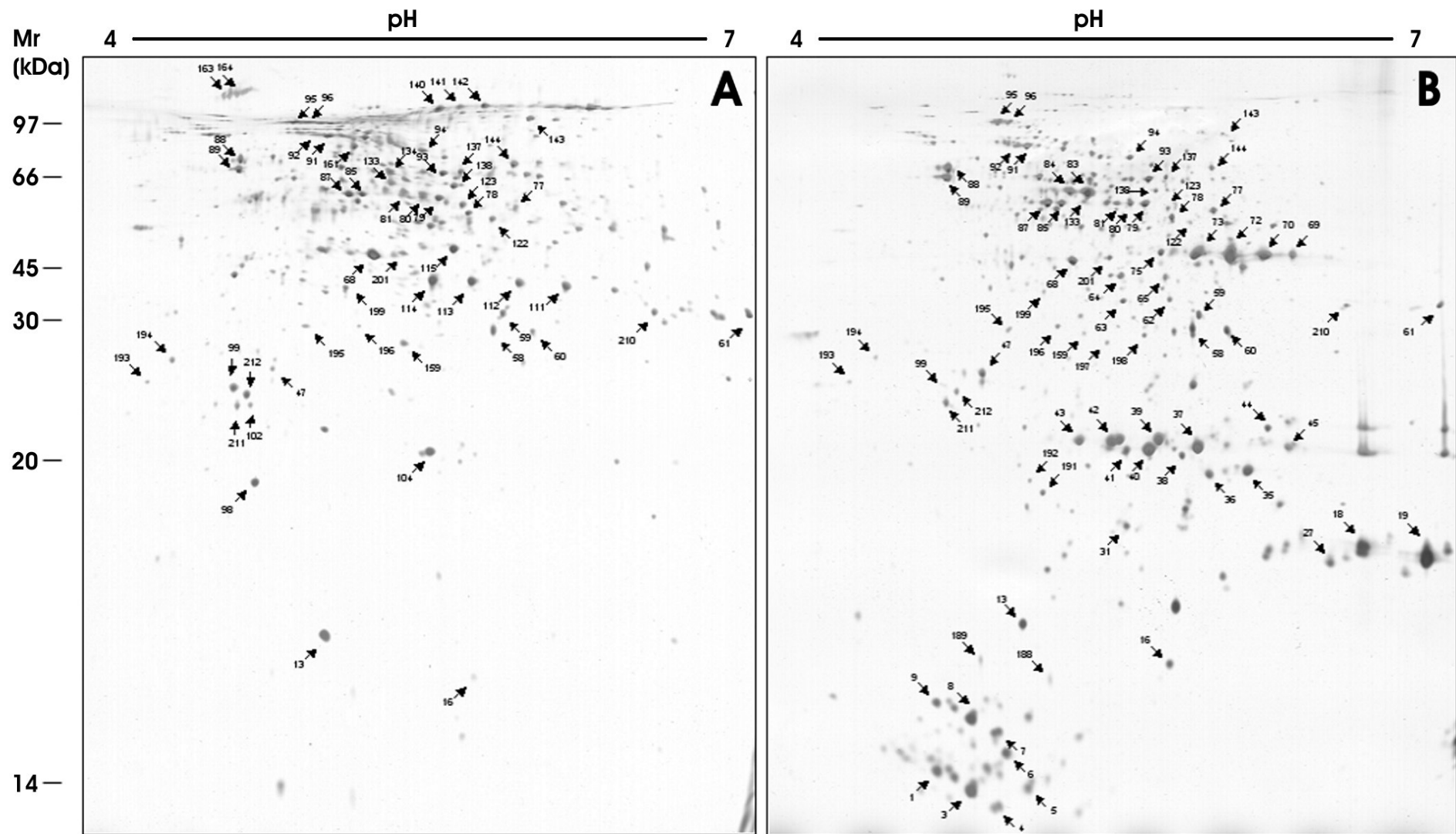


Figura 6: Géis 2-D dos estádios globular (A) e maduro (B) do desenvolvimento de sementes de *A. angustifolia*. As proteínas identificadas estão indicadas com setas e referem-se àquelas descritas na Tabela 4.

Tabela 4: Lista de proteínas identificadas durante a embriogênese de *A. angustifolia*.

Spot <sup>a</sup>	Acúmulo <sup>b</sup>	MASCOT					MS BLAST					Mr (kDa)		pI		
		Proteína	N acesso <sup>c</sup>	Organismo	Escore <sup>d</sup>	Peptídeos únicos <sup>e</sup>	Cobertura <sup>f</sup> (%)	Proteína	N acesso <sup>c</sup>	Organismo	Seq alinhadas <sup>g</sup>	Cobertura <sup>h</sup> (%)	Teo.	Obs.	Teo.	Obs.
<b>Processamento protéico</b>																
16		Heat shock protein 17.0	Q40851	<i>P. glauca</i>	94	2	11	Heat shock protein 17.0	Q40851	<i>P. glauca</i>	3	23	17	17	5.8	5.7
36		Chloroplast chaperonin 21	Q6B4V3	<i>V. vinifera</i>	70	1	11	Putative chaperonin 21	Q6Y679	<i>P. vittata</i>	1	17	13	27	4.6	5.84
37		Peptidyl-prolyl cis-trans isomerase	P21568	<i>S. lycopersicum</i>	74	1	8	Peptidyl-prolyl cis-trans isomerase CYP20-2	Q9ASS6	<i>A. thaliana</i>	3	15	18	29	8.8	5.8
63		Putative elongation factor 2	Q6H4L2	<i>O. sativa</i>	164	3	3	OSJNBa0020P07.3 protein	Q7XTK1	<i>O. sativa</i>	4	6	94	41	5.9	5.5
75		Protein disulfide isomerase	Q5EUD6	<i>Z. mays</i>	146	3	4	Putative disulfide-isomerase	Q75M08	<i>O. sativa</i>	3	9	40	47	6.3	5.6
88		Putative rubisco subunit binding-protein alpha subunit	Q8L5U4	<i>A. thaliana</i>	236	4	7	RuBisCO large subunit-binding protein subunit alpha	P21238	<i>A. thaliana</i>	10	19	62	63	5.1	4.7
89		Putative rubisco subunit binding-protein alpha subunit	Q8L5U4	<i>A. thaliana</i>	391	5	9	RuBisCO large subunit-binding protein subunit alpha	P21238	<i>A. thaliana</i>	21	43	62	61	5.1	4.7
91		Hsc70	Q40151	<i>L. esculentum</i>	454	8	20	Heat shock protein 70	Q40693	<i>O. sativa</i>	6	11	72	70	5.2	5.0
92		Hsc70	Q40151	<i>L. esculentum</i>	318	6	14	Heat shock protein 70	Q40693	<i>O. sativa</i>	4	8	72	70	5.2	4.9
93		Heat shock protein 60	Q8H6U4	<i>P. dulcis</i>	485	7	13	Heat shock protein 60	Q8H6U4	<i>P. dulcis</i>	16	33	58	61	5.3	5.6
95		Molecular chaperone Hsp90-1	Q6UJX6	<i>N. benthamiana</i>	975	20	22	Heat shock protein 83	P51819	<i>I. nil</i>	19	30	80	80	4.9	4.9
96		Heat shock protein	Q71EE1	<i>H. brasiliensis</i>	913	18	19	Heat shock protein 81-1	Q0J4P2	<i>O. sativa</i>	19	30	80	80	5.0	4.9

115		Protein disulfide isomerase	Q5EUD6	<i>Z. mays</i>	65	2	6	Probable protein disulfide-isomerase	P38661	<i>M. sativa</i>	3	8	40	52	6.3	5.6
133		Leucine aminopeptidase 1	P30184	<i>A. thaliana</i>	245	5	10	Leucine aminopeptidase 1	P30184	<i>A. thaliana</i>	9	20	55	56	5.7	5.3
134		Heat shock protein 60	Q8H6U4	<i>P. dulcis</i>	116	3	4	Heat shock protein 60	Q8H6U4	<i>P. dulcis</i>	5	10	58	70	5.3	5.4
137		Putative TCP-1	Q6AV23	<i>O. sativa</i>	432	8	17	Putative TCP-1	Q6AV23	<i>O. sativa</i>	16	34	58	61	5.6	5.7
159		Putative elongation factor 2	Q6H4L2	<i>O. sativa</i>	164	3	3	OSJNBa0020P07.3 protein	Q7XTK1	<i>O. sativa</i>	4	6	94	38	5.9	5.3
161		NI	-	-	-	-	-	70 kDa peptidyl-prolyl isomerase	Q38931	<i>A. thaliana</i>	3	5	62	82	5.2	5.2
188		NI	-	-	-	-	-	40S ribosomal protein S12	Q9XHS0	<i>H. vulgare</i>	5	41	15	16	5.4	5.1
189		Heat shock protein 17.0	Q40851	<i>P. glauca</i>	122	3	15	Heat shock protein 17.0	Q40851	<i>P. glauca</i>	4	32	17	17	5.8	4.8
192		20 kDa chaperonin	O65282	<i>A. thaliana</i>	110	3	6	Chloroplast chaperonin 21	Q6B4V4	<i>V. vinifera</i>	8	36	27	26	8.9	5.1
195		Putative 60S acidic ribosomal protein P0	Q8LNX8	<i>Z. elegans</i>	241	5	36	60S acidic ribosomal protein P0	P50345	<i>L. luteus</i>	12	39	22	39	9.6	4.9
196		Putative 60S acidic ribosomal protein P0	Q8LNX8	<i>Z. elegans</i>	255	6	38	60S acidic ribosomal protein P0	Q24573	<i>Z. mays</i>	12	44	22	39	9.6	5.2
210		Translational elongation factor Tu	Q851Y8	<i>O. sativa</i>	503	9	22	Translational elongation factor Tu	Q8W2C4	<i>O. sativa</i>	12	27	48	41	6.0	6.5

Metabolismo basal

47		Putative uncharacterized protein	O04428	<i>C. paradisi</i>	291	5	17	Putative uncharacterized protein	O04428	<i>C. paradisi</i>	10	34	33	34	5.5	4.8
60		Cysteine synthase	P32260	<i>S. oleracea</i>	131	4	8	Cysteine synthase	Q9XE66	<i>O. sativa</i>	7	7	41	38	6.8	5.9
64		Glutamine synthetase	Q9AXD8	<i>A. marina</i>	498	10	17	Glutamine synthetase	Q9AXD8	<i>A. marina</i>	12	36	39	44	6.2	5.4
72		GDP-mannose-3,5-epimerase	Q93VR3	<i>A. thaliana</i>	143	3	7	GDP-mannose 3,5-epimerase	Q93VR3	<i>A. thaliana</i>	1	4	43	47	5.9	5.9
73		GDP-mannose-3,5-epimerase	Q93VR3	<i>A. thaliana</i>	395	7	20	GDP-mannose-3,5-epimerase	Q93VR3	<i>A. thaliana</i>	4	14	43	47	5.9	5.8
77		Adenosylhomocysteinase	Q4H1G1	<i>B. vulgaris</i>	211	5	13	Adenosylhomocysteinase	P93253	<i>M. crystallinum</i>	3	6	54	55	5.8	5.9
83		Granule-bound glycogen (Starch) synthase	Q9ZSQ5	<i>A. membranaceus</i>	213	4	6	Granule-bound glycogen (Starch) synthase	Q9ZSQ5	<i>A. membranaceus</i>	8	16	67	58	7.6	5.3
84		Granule-bound starch synthase 1	Q00775	<i>S. tuberosum</i>	60	1	3	Granule-bound starch synthase 1	Q43092	<i>P. sativum</i>	2	4	67	58	6.9	5.2
94		Cytosolic phosphoglucomutase	Q6S3D6	<i>P. tomentosa</i>	459	6	14	Phosphoglucomutase	P93804	<i>Z. mays</i>	15	27	63	66	5.5	5.5
112		Type IIIa membrane protein cp-wap13	O24548	<i>V. unguiculata</i>	124	3	8	Type IIIa membrane protein cp-wap13	O24548	<i>V. unguiculata</i>	3	10	39	46	6.2	5.9
113		Type IIIa membrane protein cp-wap13	O24548	<b>V. unguiculata</b>	358	8	30	Type IIIa membrane protein cp-wap13	O24548	<i>V. unguiculata</i>	4	14	39	46	6.2	5.7
114		Type IIIa membrane protein cp-wap13	O24548	<i>V. unguiculata</i>	323	8	28	Putative reversibly glycosylatable polypeptide	Q9SR90	<i>A. thaliana</i>	3	10	39	46	6.2	5.6

123		Adenosylhomocysteinase	Q4H1G1	<i>B. vulgaris</i>	508	11	18	Adenosylhomocysteinase 1	O23255	<i>A. thaliana</i>	8	16	54	55	5.8	5.7
138		Betaine-aldehyde dehydrogenase	Q6DQ92	<i>M. acuminata</i>	66	1	5	Betaine-aldehyde dehydrogenase	Q6DQ92	<i>M. acuminata</i>	6	26	29	59	6.3	5.6
140		NI	-	-	-	-	-	Putative alpha-xylosidase	Q8VWV9	<i>P. pinaster</i>	2	31	101	104	6.3	5.6
141		Alpha-glucosidase 1	O22444	<i>A. thaliana</i>	59	1	1	Alpha-glucosidase 1	O22444	<i>A. thaliana</i>	1	1	101	105	5.6	5.7
142		Alpha glucosidase 1	O22444	<i>A. thaliana</i>	54	1	1	Alpha-glucosidase 1	O22444	<i>A. thaliana</i>	2	2	101	106	5.6	5.8
143		Cobalamin-independent methionine synthase	Q6KCR0	<i>A. thaliana</i>	131	2	3	Cobalamin-independent methionine synthase	Q6KCR0	<i>A. thaliana</i>	2	3	91	76	8.2	5.9
199		Putative adenosine kinase	Q6SV73	<i>P. tremula</i>	612	9	23	Putative adenosine kinase	Q6SV73	<i>P. alba</i>	8	38	25	42	6.0	5.1
201		S-adenosylmethionine synthetase	Q9FVG7	<i>P. contorta</i>	802	11	39	S-adenosylmethionine synthetase	Q9FVG7	<i>P. contorta</i>	14	43	43	47	5.6	5.4
Metabolismo energético																
41		Triosephosphate isomerase	P48496	<i>S. oleracea</i>	293	3	13	Triosephosphate isomerase	P48496	<i>S. oleraceae</i>	12	42	34	28	6.5	5.5
43		Triosephosphate isomerase	P48496	<i>S. oleracea</i>	106	1	4	Triosephosphate isomerase	P46225	<i>S. cereale</i>	2	8	34	29	6.5	5.3
58		Malate dehydrogenase	Q5QLS8	<i>O. sativa</i>	50	1	3	Malate dehydrogenase	Q42972	<i>O. sativa</i>	1	3	42	39	7.6	7.8

59		Nodule-enhanced malate dehydrogenase	Q81278	<i>G. max</i>	440	7	19	Nodule-enhanced malate dehydrogenase	Q81278	<i>G. max</i>	10	23	44	40	6.9	5.8
69		NADP specific isocitrate dehydrogenase	Q9ZW11	<i>D. carota</i>	646	10	23	NADP specific isocitrate dehydrogenase	Q9ZW11	<i>D. carota</i>	10	28	46	47	6.5	6.2
70		NADP specific isocitrate dehydrogenase	Q9ZW11	<i>D. carota</i>	440	8	18	NADP specific isocitrate dehydrogenase	Q9ZW11	<i>D. carota</i>	11	30	46	47	6.5	6.1
78		Enolase	Q42971	<i>O. sativa</i>	88	2	8	Enolase 1	Q9LEJ0	<i>H. brasiliensis</i>	2	6	48	53	5.4	5.7
79		UTP--glucose-1-phosphate uridylyltransferase	Q43772	<i>H. vulgare</i>	257	6	15	UTP--glucose-1-phosphate uridylyltransferase	Q9SDX3	<i>M. acuminata</i>	9	21	52	56	5.2	5.6
81		UDP-glucose pyrophosphorylase	Q8W557	<i>A. fruticosa</i>	182	3	7	UDP-glucose pyrophosphorylase	Q8W557	<i>A. fruticosa</i>	4	9	52	56	6.1	5.4
85		ATP synthase subunit beta	Q01859	<i>O. sativa</i>	259	5	13	ATP synthase subunit beta	P29685	<i>H. brasiliensis</i>	7	14	59	56	5.9	5.2
87		ATP synthase subunit beta	Q01859	<i>O. sativa</i>	370	6	17	ATP synthase subunit beta	P19023	<i>Z. mays</i>	8	17	59	56	5.9	5.1
111		Enolase	Q6WB92	<i>G. barbadense</i>	86	2	5	Enolase	Q43130	<i>M. crystallinum</i>	2	6	48	45	6.2	6.1
122		Enolase	Q42971	<i>O. sativa</i>	252	4	12	Enolase	Q42971	<i>O. sativa</i>	3	10	48	53	5.4	5.8
144		2,3-bisphosphoglycerate-independent phosphoglycerate mutase	P35493	<i>R. communis</i>	97	3	5	Phosphoglycerate mutase	Q9XE59	<i>S. tuberosum</i>	2	4	61	63	5.5	5.9
193		ATP synthase gamma chain	Q59153	<i>I. nil</i>	152	3	8	ATP synthase subunit gamma	P26360	<i>I. batatas</i>	3	10	36	34	9.0	4.2
198		Fructose-bisphosphate aldolase	Q9LLD7	<i>O. sativa</i>	132	3	8	Fructose-bisphosphate aldolase	Q53P96	<i>O. sativa</i>	2	6	42	38	9.0	5.6

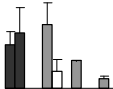
Reserva

1		Vicilin-like storage protein	Q8LKI7	<b>A. angustifolia</b>	587	10	20	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	9	23	53	13	7.7	4.6
3		Vicilin-like storage protein	Q8LKI7	<b>A. angustifolia</b>	998	18	22	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	9	24	53	13	7.7	4.8
4		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	552	8	17	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	8	20	53	12	7.7	4.9
5		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	518	9	18	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	10	26	53	13	7.7	5.0
6		NI	-	-	-	-	-	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	2	4	53	14	7.7	4.9
7		High molecular weight glutenin subunit 1By15	Q4JHY1	<i>T. aestivum</i>	111	2	2	High molecular weight glutenin y-subunit	Q7Y0S9	<i>T. elongatum</i>	1	2	78	14	8.6	4.9
8		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	508	10	18	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	6	14	53	15	7.7	4.8
9		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	433	8	15	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	8	20	53	15	7.7	4.6
18		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	1035	22	36	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	13	31	53	22	7.7	6.5
19		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	1100	24	35	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	11	27	53	22	7.7	6.8
27		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	905	21	32	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	9	24	53	22	7.7	6.4

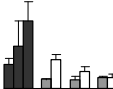
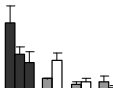
31		NI	-	-	-	-	-	Vicilin-like storage protein	Q40844	<i>P. glauca</i>	3	7	51	24	7.9	5.5
35		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	175	4	8	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	6	13	53	27	7.7	6.0
38		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	566	10	23	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	11	18	53	28	7.7	5.7
39		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	244	7	23	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	10	26	53	29	7.7	5.6
40		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	277	7	15	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	9	23	53	29	7.7	5.6
42		Vicilin	Q8LKI7	<i>A. angustifolia</i>	218	5	12	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	10	26	53	29	7.7	5.4
44		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	167	5	12	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	1	2	53	30	7.7	6.1
45		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	304	6	12	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	5	11	53	29	7.7	6.2
61		Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	727	13	27	Vicilin-like storage protein	Q8LKI7	<i>A. angustifolia</i>	7	15	53	41	7.7	6.9
Resposta ao estresse																
13		Peroxiredoxin	Q8S3L0	<i>P. tremula</i>	116	3	16	Peroxiredoxin	Q8S3L0	<i>P. tremula</i>	6	40	17	19	5.6	5.0

62		Putative aldose reductase	Q65WW3	<i>O. sativa</i>	107	3	5	Putative aldose reductase	Q8GXW0	<i>A. thaliana</i>	6	22	36	42	6.3	5.7
80		Putative mitochondrial aldehyde dehydrogenase ALDH2a	Q6YWQ9	<i>O. sativa</i>	57	1	3	Mitochondrial aldehyde dehydrogenase	Q8LST5	<i>S. bicolor</i>	1	2	46	56	5.9	5.5
98		Cytosolic ascorbate peroxidase 2	Q76LA6	<i>G. max</i>	55	1	7	Cytosolic ascorbate peroxidase	Q8H1K7	<i>R. raetam</i>	1	6	27	28	5.7	4.8
104		Ascorbate peroxidase	Q6RY58	<i>P. pinaster</i>	218	4	17	Ascorbate peroxidase	Q6RY58	<i>P. pinaster</i>	4	19	27	30	5.4	5.5
Sinalização																
99		14-3-3-like protein GF14 iota	Q8LEN1	<i>A. thaliana</i>	166	3	13	14-3-3-like protein D	Q96453	<i>G. max</i>	2	8	29	34	4.9	4.7
102		14-3-3-like protein GF14 iota	Q8LEN1	<i>A. thaliana</i>	113	2	10	14-3-3 protein	Q944P2	<i>F. cirrhosa</i>	2	12	29	34	4.9	4.8
211		14-3-3 protein	Q6PWL7	<i>S. chacoense</i>	306	6	22	14-3-3-like protein	Q9SP07	<i>L. longiflorum</i>	4	19	29	32	4.7	4.7
212		14-3-3-like protein GF14 iota	Q8LEN1	<i>A. thaliana</i>	161	3	13	14-3-3 protein	Q944P2	<i>F. cirrhosa</i>	3	17	29	33	4.9	4.7
Transporte																
163		NI	-	-	-	-	-	Patellin-3	Q56Z59	<i>A. thaliana</i>	2	5	56	113	5.2	4.7
164		NI	-	-	-	-	-	Patellin-3	Q56Z59	<i>A. thaliana</i>	2	5	56	113	5.2	4.7

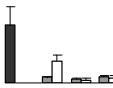
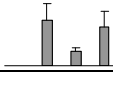
Ciclo celular

194		Proliferating cell nuclear antigen	P17070	<i>O. sativa</i>	376	9	37	Proliferating cell nuclear antigen large form	Q00265	<i>D. carota</i>	17	49	29	36	4.6	4.3
-----	---	------------------------------------	--------	------------------	-----	---	----	---	--------	------------------	----	----	----	----	-----	-----

Estrutural

68		Actin	Q9SP17	<i>P. rubens</i>	1288	23	62	Actin	P93485	<i>P. sativum</i>	18	55	83	46	6.0	5.2
197		Tubulin beta-2 chain	P18026	<i>Z. mays</i>	291	5	12	Tubulin beta-4 chain	Q41782	<i>Z. mays</i>	3	7	50	37	4.8	5.4

Classificação incerta

65		NI	-	-	-	-	-	F13E7.34 protein	Q9M8R4	<i>A. thaliana</i>	4	11	42	44	5.2	5.6
191		GTP-binding protein, ras-like	Q9FJH0	<i>A. thaliana</i>	368	7	31	GTP-binding protein, ras-like	Q9FJH0	<i>A. thaliana</i>	7	34	24	26	5.7	5.1

NI: proteína não identificada

a: números correspondem àqueles presentes na Figura 6

b: perfil de expressão durante o desenvolvimento da semente. Histogramas em preto, cinza e branco indicam acúmulo em megagametófitos contendo embriões; embriões; e megagametófitos, respectivamente

c: número de acesso no Swiss-Prot/TrEMBL.

d: escore total inferido por probabilidade MOWSE pelo MASCOT

e: número de sequências peptídicas únicas identificadas pelo MASCOT

f: porcentagem de cobertura prevista pelo MASCOT

g: número de alinhamentos identificados pelo MS BLAST com escore igual ou acima de 55

h: porcentagem de cobertura prevista pelo MS BLAST de acordo com a fórmula  $[\sum \text{alinhamentos positivos (aa)} \times 100] / \text{sequência prevista (aa)}$ .

A estratégia de identificação de proteínas por similaridade de sequências validou todas as identificações sugeridas pelo MASCOT, e permitiu a identificação de 8 novas proteínas (Tabela 4: “spots” 6, 31, 65, 140, 161, 163, 164 e 188). Vale ainda ressaltar que das 96 identificações, 38 encontravam-se presentes em mais de um “spot”, como a proteína vicilina (número de acesso Q8W557). A ocorrência de “spots” múltiplos já foi reportada em espécies de coníferas, como em *Pinus pinaster* (GION *et al.*, 2005) e *Picea sitchensis* (LIPPERT *et al.*, 2007). Sua presença pode ser consequência de modificações pós-traducionais, ou da existência de diferentes isoformas, derivadas de um único gene ou de variantes alélicas (LIPPERT *et al.*, 2007).

#### 4.2.2. Classificação funcional de proteínas

As 96 proteínas identificadas neste trabalho foram classificadas de acordo com sua função biológica. Para tanto, foi utilizada a classificação proposta pelo “Gene Ontology Annotation”, disponível no sítio [www.ebi.ac.uk/GOA/](http://www.ebi.ac.uk/GOA/). Como não existem catálogos funcionais delimitados para espécies de coníferas, foi utilizado um catálogo global, disponível no sítio [mips.gsf.de/project/funcat](http://mips.gsf.de/project/funcat) (RUEPP *et al.*, 2004). Sendo assim, as proteínas foram classificadas em 9 e 8 categorias para os estádios globular e maduro, respectivamente (Figura 7). As classes mais representativas em ambos os estádios foram as proteínas envolvidas no processamento protéico (ajuste conformacional e degradação de proteínas), metabolismo basal, metabolismo energético e de reserva.

A maior classe de proteínas detectadas durante a embriogênese de *A. angustifolia* é aquela referente à função biológica de processamento protéico, sendo as proteínas do tipo “heat shock” as mais abundantes dentro dessa classe. O segundo grupo mais representativo foi aquele formado por proteínas envolvidas no metabolismo basal, contendo 26 e 17% dos “spots” nos estádios globular e maduro, respectivamente (Figura 7). A classe das proteínas envolvidas no metabolismo energético, ou seja, relacionadas à glicólise, via do ácido tricarboxílico (TCA) e conversão de energia constituiu-se de 16 “spots”; entretanto, apenas 10 identificações foram observadas. Outro grupo com alta frequência de identificações redundantes é o das proteínas de reserva. Este grupo apresentou a maior variação entre os

estádios globular e maduro, sendo que no estágio maduro foi uma das classes mais abundantes. Sendo assim, verificou-se que o envolvimento de 2% dos “spots” no estágio globular e 25% no estágio maduro relacionavam-se aos processos de reserva.

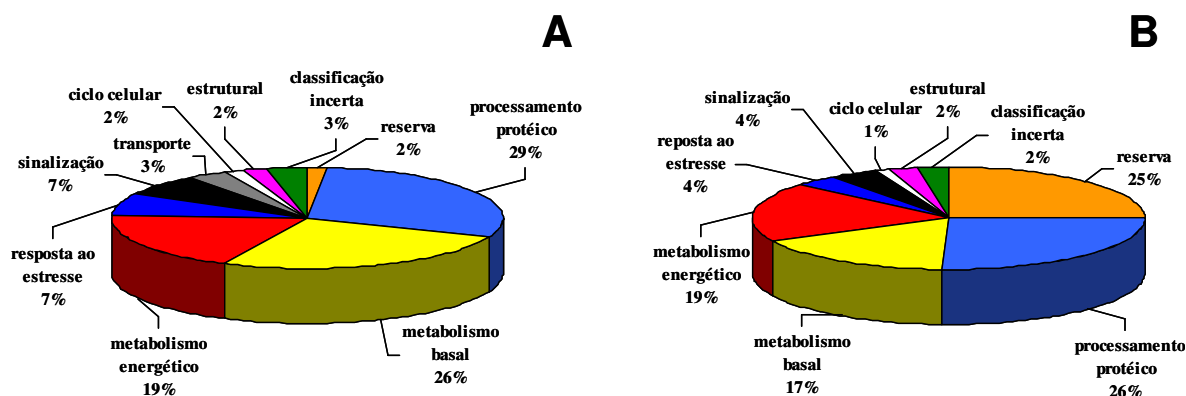


Figura 7: Classificação funcional dos “spots” identificados durante o desenvolvimento das sementes de *A. angustifolia*. **A**: estágio globular, **B**: estágio maduro.

#### 4.2.3. Dinâmica de acúmulo das proteínas identificadas

Em análises do proteoma por 2-DE, a expressão de um único gene pode resultar na detecção de vários “spots” protéicos (AGRAWAL *et al.*, 2008). Além disso, *A. angustifolia* é uma espécie cujo genoma é desconhecido, dificultando a caracterização e identificação das funções biológicas desenvolvidas pelas diferentes isoformas. Portanto, na caracterização da dinâmica do perfil de detecção das proteínas identificadas foram consideradas todas as variações por acesso entre os 96 “spots” identificados.

Sendo assim, os 66 acessos protéicos foram agrupados de acordo com a abundância e o perfil de detecção observado durante da embriogênese zigótica (Figura 8). O agrupamento A é caracterizado por proteínas abundantes durante todo o desenvolvimento embrionário e é formado por apenas 7 acessos protéicos, como a actina e a proteína ligadora à RuBisCo (RBP). O agrupamento B é formado por 26 proteínas com baixo acúmulo durante a embriogênese; enquanto que o agrupamento C é formado por 25 proteínas abundantes nos estádios finais da embriogênese. Finalmente, o agrupamento D, formado

por apenas 6 proteínas, é caracterizado por alto acúmulo no início do desenvolvimento embrionário.

Visando uma melhor compreensão das alterações no proteoma e na ativação de diferentes rotas metabólicas durante o desenvolvimento da semente, algumas proteínas de maior abundância no início e final da embriogênese serão discutidas. Essas proteínas, além de fornecer informações sobre mudanças no metabolismo da embriogênese zigótica, são potenciais candidatas a marcadores protéicos neste processo de desenvolvimento.

#### 4.2.4. Proteínas de maior abundância durante o início do desenvolvimento das sementes

A ascorbato peroxidase (APX) é uma das enzimas mais importantes do sistema oxidativo, atuando no controle e modulação dos níveis intracelulares do peróxido de hidrogênio ( $H_2O_2$ ) (ISHIKAWA e SHIGEOKA, 2008). O  $H_2O_2$  possui um efeito celular deletério em altas concentrações; entretanto, ele pode atuar na defesa contra organismos invasores durante o final do desenvolvimento da semente e na oxidação de inibidores da germinação em angiospermas (OGAWA e IWABUCHI, 2001; SCHOPFER *et al.*, 2001). Atualmente, existem poucos trabalhos relacionados à síntese de espécies reativas de oxigênio (ROS) durante o desenvolvimento das sementes recalcitrantes. Por outro lado, vários estudos têm reportado a produção de ROS durante o acúmulo de proteínas e dessecação em espécies ortodoxas (MCDONALD, 1999; PUKACKA e RATAJCZAK, 2005).

Embora apenas um acesso protéico de APX (Q76LA6) tenha sido incorporado ao agrupamento D, o outro (Q6RY58) também apresentou maior acúmulo no estágio globular (Figura 8). A identificação dessa enzima, em abundância nos estádios pró-embrionário, globular e torpedo, sugere um elevado metabolismo oxidativo durante o início do desenvolvimento embrionário, quando a APX pode atuar na regulação e controle da atividade das ROS. Tal fato está de acordo com a grande quantidade de identificações funcionalmente agrupadas na categoria 'resposta ao estresse' durante o início de desenvolvimento das sementes de *A. angustifolia*, correspondente ao estágio globular (Figura 7A).

O início da embriogênese em coníferas é caracterizado por diversos ciclos de divisão nuclear, seguido de celularização (CAIRNEY e PULLMAN, 2007). A divisão celular nas plantas ocorre pela formação da placa celular e é caracterizada por eventos de mobilização e transporte de membranas (BEDNAREK e FALBEL, 2002). A patelina (“spots” 163 e 164), que foi observada em pequena quantidade e identificada principalmente no estágio globular (Tabela 4 e Figura 8), é uma proteína carregadora recrutada durante o desenvolvimento da placa celular, sendo a sua estrutura constituída por um domínio de ligação lipídica e um domínio de interação com o aparato golgiano (PETERMAN *et al.*, 2004). Além de atuar nos processos de transporte celular envolvidos na citocinese, ela também participa da biossíntese de fosfolípidos (PETERMAN *et al.*, 2004). O aumento nos níveis de patelina nos estádios globular e torpedo, seguido de redução durante o final do desenvolvimento das sementes, indicam uma possível redução nas taxas de divisão celular durante o acúmulo de proteínas e maturação da semente de *A. angustifolia*. Tal fato está de acordo com a ausência de alterações e transições morfológicas, nos embriões zigóticos e megagametófitos, a partir do estágio torpedo (ASTARITA *et al.*, 2003) e podem indicar uma redução na taxas de diferenciação e perda de competência celular para a formação de pró-embriões à partir do estágio cotiledonar inicial.

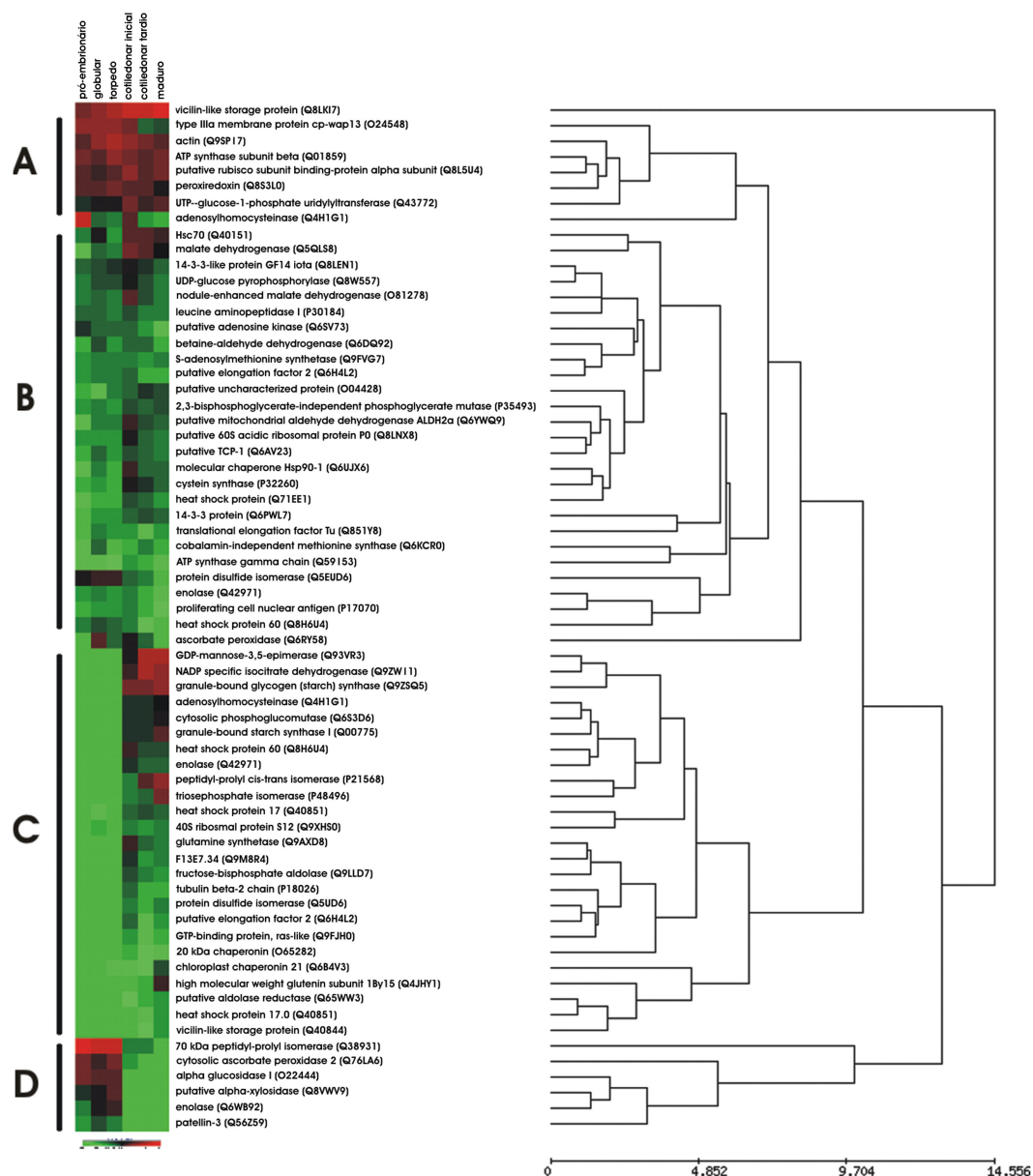


Figura 8: Abundância e perfil de detecção das 66 proteínas únicas identificadas durante o desenvolvimento da semente de *A. angustifolia*.

A proteína antígeno nuclear de proliferação celular (PCNA) possui um papel importante na duplicação e reparo do DNA (HERRERA *et al.*, 2000), tendo sido observada em sementes de milho (HERRERA *et al.*, 2000). Em *A. angustifolia*, uma espécie recalcitrante, foi observada uma redução nos níveis de PCNA durante o final do desenvolvimento, indicando a necessidade de síntese *de novo* dessa proteína durante a maturação das sementes.

Processos biológicos que envolvem altas taxas de divisão celular requerem a expressão e acúmulo de proteínas envolvidas na biossíntese e degradação de membrana e parede celular, como as xilosidases (“spot” 140) e as proteínas de membrana cp-wap 13 (“spots” 112, 113 e 114), essenciais na expansão celular e tecidual de tecidos jovens (SANCHEZ *et al.*, 2003; WALD *et al.*, 2003; NAKAI *et al.*, 2007).

Poucas proteínas envolvidas no metabolismo do carbono e de reserva durante os primeiros estádios da embriogênese de *A. angustifolia*. Destaca-se a identificação da alfa-glucosidase do tipo I (“spots” 141 e 142), a qual hidrolisa, preferencialmente, substratos heterogêneos como a sacarose (NAKAI *et al.*, 2007). É interessante notar que enquanto os níveis de alfa-glucosidase diminuem, os níveis da enzima amido sintetase (“spots” 83 e 84) aumentam durante o desenvolvimento da semente. Esta situação sugere que, durante o início do desenvolvimento embrionário, a sacarose seja hidrolisada pela alfa-glucosidase I, liberando alfa-D-glucose que é utilizada como substrato na biossíntese do amido, que é o metabólito de reserva mais abundante em sementes maduras de *A. angustifolia* (CORDENUNSI *et al.*, 2004).

#### 4.2.5. Proteínas de maior abundância durante o final do desenvolvimento das sementes

As chaperoninas são proteínas que desempenham papel fundamental no processamento protéico. Atualmente, elas são divididas em duas classes de acordo com sua localização celular: a) as chaperoninas do tipo I que estão presentes no estroma do cloroplasto, matriz das mitocôndrias e em eubactérias; e b) as chaperoninas do tipo II que são encontradas no citossol de eucariotos e em arqueobactérias (KOUmoto *et al.*, 2001). Uma das principais diferenças entre as duas classes é que as chaperoninas do grupo I têm sua atividade dependente de co-chaperoninas (KOUmoto *et al.*, 2001). Durante o desenvolvimento das sementes de *A. angustifolia*, as chaperoninas não apresentaram acúmulo preferencial em nenhum estágio de desenvolvimento. Entretanto, as chaperoninas de acesso protéico igual a O65282 e Q6B4V3, ambas apresentando 20 kDa, foram detectadas apenas no final do processo estudado (Figura 8).

As chaperoninas cpn60 e a co-chaperonina cpn10 já foram extensivamente estudadas em *Escherichia coli* (BONSHTIEN *et al.*, 2007). Em plantas, um homólogo desta co-chaperonina foi identificado em *Pisum sativum* e *Spinacia oleracea*, possuindo o dobro de sua massa e cuja sequência indicou ser formada por dois homólogos da cpn10, sendo assim denominada co-chaperonina cpn20 (BERTSCH *et al.*, 1992; DICKSON *et al.*, 2000). Estudos *in vitro* mostraram que a cpn20 participa, em conjunto com a cpn60, no processamento da RuBisCo (BERTSCH *et al.*, 1992; VIITANEN *et al.*, 1995). Embora a RBP (Q8L5U4) tenha sido detectada nas fases iniciais e finais do desenvolvimento de *A. angustifolia*, a co-chaperonina cpn20 não foi detectada durante o início da embriogênese (Figura 8). No presente trabalho verificou-se a presença das co-chaperoninas cpn20 somente a partir do estágio torpedo. Esta situação indica que a diferenciação e desenvolvimento do aparelho fotossintético inicia-se a partir do desenvolvimento cotiledonar, podendo atuar no fornecimento de energia, através do processo fotossintético.

Durante os estádios iniciais da embriogênese, a concentração de ATP é baixa, já que o O<sub>2</sub> disponível é principalmente convertido para superóxido (ROLLETSCHEK *et al.*, 2004). Posteriormente, a partir da diferenciação do aparelho fotossintético e expressão e diferenciação de enzimas das vias glicolíticas e do ciclo do TCA, os níveis de ATP aumentam. Embora a enzima enolase tenha sido detectada em todos os estádios de desenvolvimento da semente de *A. angustifolia* e, conseqüentemente, tenha sido localizada nos agrupamentos B, C e D, outras enzimas envolvidas na glicólise e na via do TCA foram identificadas somente nas fases finais do desenvolvimento. Dentre elas destacam-se a trifosfato isomerase, a frutose bifosfato aldolase e a isocitrato desidrogenase. O grande acúmulo dessas enzimas nos estádios finais da embriogênese sugere uma alteração no metabolismo de estresse oxidativo. Com uma maior disponibilidade de O<sub>2</sub>, como acceptor final de hidrogênio, e aumento das taxas de respiração na semente madura. A D-glucose é outro substrato envolvido na cadeia respiratória e é possivelmente encontrada em grandes quantidades no final da embriogênese devido à hidrólise da sacarose pela ação da alfa-glucosidade do tipo I no início do desenvolvimento do embrião. Outra evidência de grande acúmulo de D-glucose no final do desenvolvimento das sementes refere-se à alta expressão

da enzima amido sintetase (Q9ZSQ5 e Q00775), que utiliza a D-glucose como substrato para a biossíntese do amido.

O final do desenvolvimento embrionário é caracterizado por um aumento na quantidade de metabólitos de reserva e acúmulo de enzimas relacionadas, como por exemplo a amido sintetase em sementes ricas em amido. Embora as sementes de *A. angustifolia* sejam amiláceas, “spots” da proteína vicilina foram identificados em grande quantidade na semente madura. A detecção de múltiplos “spots” relacionados com essa identificação sugerem a existência de alterações pós traducionais e expressão de isoformas protéicas (LIPPERT *et al.*, 2007). No presente trabalho não foram verificadas enzimas proteolíticas, embora em *A. bidwillii* foi identificada a ação de aminopeptidases em sementes maduras (GALLESCHI *et al.*, 2002). A identificação dessas mesmas endopeptidases em sementes maduras de *A. cunninghamii* (GALLESCHI *et al.*, 2002), uma espécie ortodoxa, indica que a presença de enzimas proteolíticas é independente da existência de um desenvolvimento contínuo do embrião, sem a interpolação de uma etapa de baixa atividade metabólica, como observado em sementes ortodoxas. Portanto, alternativamente, o fenômeno de identificações múltiplas de proteínas do tipo vicilina pode ser resultado da digestão enzimática dessas proteínas de reserva, gerando peptídeos de diferentes tamanhos, pls, e diferentes padrões eletroforéticos.

O evento mais importante da maturação é o acúmulo de proteínas de reserva, a partir de aminoácidos livres, no final da embriogênese. Em *Pinus taeda* e *Pinus pinaster* os resíduos de aminoácidos mais frequentes são a arginina, a glutamina e o ácido glutâmico (ALLONA *et al.*, 1992; RODRIGUEZ *et al.*, 2006). As plantas utilizam nitrato, amônio e o gás N<sub>2</sub> como as principais fontes externas de nitrogênio para a biossíntese de aminoácidos e compostos orgânicos (MASCLAUX-DAUBRESSE *et al.*, 2006). O nitrato é o último composto inorgânico da via de assimilação do nitrogênio, o qual é incorporado pela ação da enzima glutamina sintetase (GS). Posteriormente, a glutamina, na presença de 2-oxoglutarato, é convertida em glutamato por atividade da glutamato sintase (GOGAT) nos cloroplastos. O glutamato pode ser convertido em outro aminoácido, a glicina, por atividade da glutamato:glioxalato aminotransferase (GGAT), nos peroxissomos, produzindo uma

molécula de 2-oxoglutarato que pode ser reciclada na forma de glutamato (Figura 9). Alternativamente, o glutamato pode também ser convertido em 2-oxoglutarato, pela atividade da enzima glutamato desidrogenase (GDH), produzindo amônio (Figura 9).

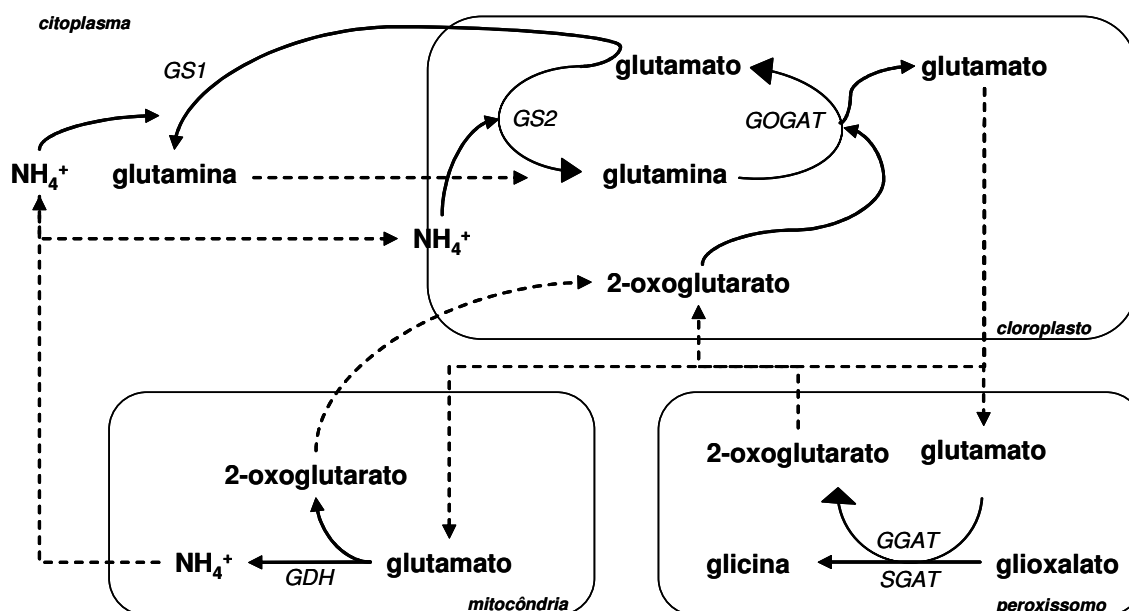


Figura 9: Esquema simplificado da assimilação e metabolismo do nitrogênio. Enzimas catalizadoras das reações: glutamina sintetase (**GS**), glutamato sintase (**GOGAT**), glutamina:glicoxalato aminotransferase (**GGAT**), serino:glicoxalase aminotransferase (**SGAT**), glutamato desidrogenase (**GDH**). Esquema adaptado de MASCLAUX-DAUBRESSE *et al.* (2006).

As reações catabolizadas pela enzima GS podem ocorrer no citoplasma e no cloroplasto pela atividade da GS1 e GS2, respectivamente; entretanto, a enzima GS2 ainda não foi detectada em coníferas (CANOVAS *et al.*, 1998). Sabe-se que a GS1 possui duas isoformas que apresentam propriedades estruturais e cinéticas distintas. A GS1a é induzida pela luz, enquanto a GS1b apresenta expressão constitutiva e está associada aos tecidos vasculares (MASCLAUX-DAUBRESSE *et al.*, 2006). Um estudo comparativo da expressão dos genes GS1a e GS1b, em embriões zigóticos e somáticos de *Pinus pinaster*, mostrou que a expressão do gene GS1b ocorre principalmente nas células do procâmbio dos embriões (RODRIGUEZ *et al.*, 2006). Entretanto, não foi observada expressão do GS1a em

embriões zigóticos, enquanto que seus transcritos foram detectados somente a partir do estágio cotiledonar nos embriões somáticos (RODRIGUEZ *et al.*, 2006).

Embora pouco se saiba à respeito do metabolismo do nitrogênio nos tecidos embrionários de coníferas, o suprimento e balanço correto desse composto é crítico no desenvolvimento embrionário (RODRIGUEZ *et al.*, 2006). Sabe-se que o ciclo GS / GOGAT é a via preferencial de assimilação do N durante a embriogênese somática em *Picea glauca* (JOY *et al.*, 1997). No presente trabalho, foi observado um grande acúmulo da enzima GS, a qual cataliza a conversão entre o ácido glutâmico e glutamina, no estágio cotiledonar inicial. Alterações no acúmulo dessa enzima, durante o processo de maturação, podem indicar que a biossíntese e interconversão da glutamina e do ácido glutâmico é ativa durante o início da maturação de coníferas. Em *A. angustifolia*, verificou-se que o ácido glutâmico é aminoácido livre mais abundante em sementes maduras (ASTARITA *et al.*, 2003). Além disso, a composição de aminoácidos da proteína vicilina (Q8LKI7) indica que mais de 10% de sua sequência é composta por ácido glutâmico. Sendo assim, a expressão da enzima GS em tempos específicos da embriogênese pode ser determinante para a obtenção de embriões maduros na espécie.

## **5. Considerações finais**

A eletroforese bidimensional e a espectrometria de massas mostraram-se eficientes na caracterização do proteoma de *A. angustifolia*, durante o desenvolvimento das sementes. Estudos de expressão temporal das identificações aqui obtidas, principalmente daquelas relacionadas às proteínas do tipo vicilinas, poderão fornecer informações importantes sobre a dinâmica da expressão dos genes e compartimentalização das proteínas correlatas. Além disso, proteínas estágio-específicas podem ser utilizadas como marcadores do processo de maturação em culturas embriogênicas de *A. angustifolia* e servir de base para estudos genéticos, fisiológicos e de metabolismo sobre o desenvolvimento embrionário em espécies de coníferas.

## 6. Referências bibliográficas

- Agrawal, G. K., M. Hajduch, K. Graham, J. J. Thelen. In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. Plant Physiology, v.148, p.504-518. 2008.
- Allona, I., R. Casado, C. Aragoncillo. Seed Storage Proteins from *Pinus pinaster* Ait - homology of major components with 11s proteins from angiosperms. Plant Science, v.87, p.9-18. 1992.
- Anguelova-Merhar, V. S., C. Calistru, P. Berjak. A study of some biochemical and histopathological responses of wet-stored recalcitrant seeds of *Avicennia marina* infected by *Fusarium moniliforme*. Annals of Botany, v.92, p.401-408. 2003.
- Astarita, L. V., E. I. S. Floh, W. Handro. Free amino acid, protein and water content changes associated with seed development in *Araucaria angustifolia*. Biologia Plantarum, v.47, p.53-59. 2003.
- Bednarek, S. Y., T. G. Falbel. Membrane trafficking during plant cytokinesis. Traffic, v.3, p.621-629. 2002.
- Bertsch, U., J. Soli, R. Seetharam, P. V. Viitanen. Identification, characterization, and DNA-sequence of a functional double groes-like chaperonin from chloroplasts of higher-plants. Proceedings of the National Academy of Sciences of the United States of America, v.89, p.8696-8700. 1992.
- Bewley, J. D. Seed germination and dormancy. Plant Cell, v.9, p.1055-1066. 1997.
- Bonshtien, A. L., C. Weiss, A. Vitlin, A. Niv, G. H. Lorimer, A. Azem. Significance of the N-terminal domain for the function of chloroplast cpn20 chaperonin. Journal of Biological Chemistry, v.282, p.4463-4469. 2007.
- Cairney, J., G. S. Pullman. The cellular and molecular biology of conifer embryogenesis. New Phytologist, v.176, p.511-536. 2007.
- Canovas, F. M., F. R. Canton, A. Garcia-Gutierrez, F. Gallardo, R. Crespillo. Molecular physiology of glutamine and glutamate biosynthesis in developing seedlings of conifers. Physiologia Plantarum, v.103, p.287-294. 1998.
- Chen, S. X., A. C. Harmon. Advances in plant proteomics. Proteomics, v.6, p.5504-5516. 2006.
- Cordenunsi, B. R., E. W. De Menezes, M. I. Genovese, C. Colli, A. G. A. De Souza, F. M. Lajolo. Chemical composition and glycemic index of Brazilian pine (*Araucaria angustifolia*) seeds. Journal of Agricultural and Food Chemistry, v.52, p.3412-3416. 2004.
- Dickson, R., C. Weiss, R. J. Howard, S. P. Aldrick, R. J. Ellis, G. Lorimer, A. Azem, P. V. Viitanen. Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. Journal of Biological Chemistry, v.275, p.11829-11835. 2000.
- Dos Santos, A. L. W., N. Wietholter, N. E. El Gueddari, B. M. Moerschbacher. Protein expression during seed during in *Araucaria angustifolia*: transient accumulation of class

- IV chitinases and arabinogalactan proteins. Physiologia Plantarum, v.127, p.138-148. 2006.
- Flinn, B. S., D. R. Roberts, D. T. Webb, B. C. S. Sutton. Storage protein changes during zygotic embryogenesis in interior spruce. Tree Physiology, v.8, p.71-81. 1991.
- Frank, A., P. Pevzner. PepNovo: De novo peptide sequencing via probabilistic network modeling. Analytical Chemistry, v.77, p.964-973. 2005.
- Galleschi, L., A. Capocchi, S. Ghiringhelli, F. Saviozzi. Some properties of proteolytic enzymes and storage proteins in recalcitrant and orthodox seeds of *Araucaria*. Biologia Plantarum, v.45, p.539-544. 2002.
- Gion, J. M., C. Lalanne, G. Le Provost, H. Ferry-Dumazet, J. Paiva, P. Chaumeil, J. M. Frigerio, J. Brach, A. Barre, A. De Daruvar, S. Claverol, M. Bonneau, N. Sommerer, L. Negrone, C. Plomion. The proteome of maritime pine wood forming tissue. Proteomics, v.5, p.3731-3751. 2005.
- Habermann, B., J. Oegema, S. Sunyaev, A. Shevchenko. The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches. Molecular & Cellular Proteomics, v.3, p.238-249. 2004.
- Herrera, I., M. D. Sanchez, J. Molina, J. Plasencia, J. M. Vazquez-Ramos. Proliferating cell nuclear antigen expression in maize seed development and germination: Regulation by phytohormones and its association with putative cell cycle proteins. Physiologia Plantarum, v.110, p.127-134. 2000.
- Ishikawa, T., S. Shigeoka. Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. Bioscience Biotechnology and Biochemistry, v.72, p.1143-1154. 2008.
- Joy, R. W., H. J. Vogel, T. A. Thorpe. Inorganic nitrogen metabolism in embryogenic white spruce cultures. A nitrogen 14/15 NMR study. Journal of Plant Physiology, v.151, p.306-315. 1997.
- Koumoto, Y., T. Shimada, M. Kondo, I. Hara-Nishimura, M. Nishimura. Chloroplasts have a novel cpn10 in addition to cpn20 as co-chaperonins in *Arabidopsis thaliana*. Journal of Biological Chemistry, v.276, p.29688-29694. 2001.
- Lippert, D., S. Chowrira, S. G. Ralph, J. Zhuang, D. Aeschliman, C. Ritland, K. Ritland, J. Bohlmann. Conifer defense against insects: proteomic analysis of Sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or feeding by white pine weevils (*Pissodes strobi*). Proteomics, v.7, p.248-270. 2007.
- Masclaux-Daubresse, C., M. Reisdorf-Cren, K. Pageau, M. Lelandais, O. Grandjean, J. Kronenberger, M. H. Valadier, M. Feraud, T. Jougllet, A. Suzuki. Glutamine synthetase-glutamate synthase pathway and glutamate dehydrogenase play distinct roles in the sink-source nitrogen cycle in tobacco. Plant Physiology, v.140, p.444-456. 2006.

- Mcdonald, M. B. Seed deterioration: physiology, repair and assessment. Seed Science and Technology, v.27, p.177-237. 1999.
- Nakai, H., S. Tanizawa, T. Ito, K. Kamiya, Y. M. Kim, T. Yamamoto, K. Matsubara, M. Sakai, H. Sato, T. Imbe, M. Okuyama, H. Mori, Y. Sano, S. Chiba, A. Kimura. Function-unknown glycoside hydrolase family 31 proteins, mRNAs of which were expressed in rice ripening and germinating stages, are alpha-glucosidase and alpha-xylosidase. Journal of Biochemistry, v.142, p.491-500. 2007.
- Ogawa, K., M. Iwabuchi. A mechanism for promoting the germination of *Zinnia elegans* seeds by hydrogen peroxide. Plant and Cell Physiology, v.42, p.286-291. 2001.
- Panza, V., V. Lainez, H. Maroder, I. Prego, S. Maldonado. Storage reserves and cellular water in nature seeds of *Araucaria angustifolia*. Botanical Journal of the Linnean Society, v.140, p.273-281. 2002.
- Pawlowski, T. A. Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: Influence of abscisic and gibberellic acids. Proteomics, v.7, p.2246-2257. 2007.
- Peterman, T. K., Y. M. Ohol, L. J. McCreynolds, E. J. Luna. Patellin1, a novel Sec14-like protein, localizes to the cell plate and binds phosphoinositides. Plant Physiology, v.136, p.3080-3094. 2004.
- Pukacka, S., E. Ratajczak. Production and scavenging of reactive oxygen species in *Fagus sylvatica* seeds during storage at varied temperature and humidity. Journal of Plant Physiology, v.162, p.873-885. 2005.
- Roberts, E. H. Predicting the storage life of seeds. Seed Science and Technology, v.1, p.499-514. 1973.
- Rodriguez, M. J. P., M. F. Suarez, R. Heredia, C. Avila, D. Breton, J. F. Trontin, L. Filonova, P. Bozhkov, S. Von Arnold, L. Harvengt, F. M. Canovas. Expression patterns of two glutamine synthetase genes in zygotic and somatic pine embryos support specific roles in nitrogen metabolism during embryogenesis. New Phytologist, v.169, p.35-44. 2006.
- Rolletschek, H., W. Weschke, H. Weber, U. Wobus, L. Borisjuk. Energy state and its control on seed development: starch accumulation is associated with high ATP and steep oxygen gradients within barley grains. Journal of Experimental Botany, v.55, p.1351-1359. 2004.
- Ruepp, A., A. Zollner, D. Maier, K. Albermann, J. Hani, M. Mokrejs, I. Tetko, U. Guldener, G. Mannhaupt, M. Munsterkotter, H. W. Mewes. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Research, v.32, p.5539-5545. 2004.
- Sallandrouze, A., M. Faurobert, M. El Maataoui. Characterization of the developmental stages of cypress zygotic embryos by two-dimensional electrophoresis and by cytochemistry. Physiologia Plantarum, v.114, p.608-618. 2002.

- Sanchez, M., C. Gianzo, J. Sampedro, G. Revilla, I. Zarra. Changes in alpha-xylosidase during intact and auxin-induced growth of pine hypocotyls. Plant and Cell Physiology, v.44, p.132-138. 2003.
- Schopfer, P., C. Plachy, G. Frahy. Release of reactive oxygen intermediates (superoxide radicals, hydrogen peroxide, and hydroxyl radicals) and peroxidase in germinating radish seeds controlled by light, gibberellin, and abscisic acid. Plant Physiology, v.125, p.1591-1602. 2001.
- Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen, M. Mann. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature Protocols, v.1, p.2856-2860. 2006.
- Shevchenko, A., M. Wilm, O. Vorm, M. Mann. Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. Analytical Chemistry, v.68, p.850-858. 1996.
- Silveira, V., C. Santa-Catarina, T. S. Balbuena, F. M. S. Moraes, C. A. O. Ricart, M. V. Sousa, M. P. Guerra, W. Handro, E. I. S. Floh. Endogenous abscisic acid and protein contents during seed development of *Araucaria angustifolia*. Biologia Plantarum, v.52, p.101-104. 2008.
- Viitanen, P. V., M. Schmidt, J. Buchner, T. Suzuki, E. Vierling, R. Dickson, G. H. Lorimer, A. Gatenby, J. Soll. Functional-Characterization of the Higher-Plant Chloroplast Chaperonins. Journal of Biological Chemistry, v.270, p.18158-18164. 1995.
- Wald, F. A., R. Kissen, P. Du Jardin, S. Moreno. Characterization of UDP-glucose : protein transglucosylase genes from potato. Plant Molecular Biology, v.52, p.705-714. 2003.

## **CAPÍTULO 5**

## CAPÍTULO 5. Considerações gerais, implicações e perspectivas

No presente trabalho foi caracterizado o proteoma do desenvolvimento da semente de *Araucaria angustifolia*. As características deste sistema vegetal, associado ao número limitado de estudos com esse material exigiram a adequação de metodologias iniciais.

Sementes de *A. angustifolia* são ricas em amido, os quais podem bloquear os poros dos géis de acrilamida, causando precipitação, aumento no tempo de focalização, rastros e perda de proteínas (CARPENTIER *et al.*, 2005). Portanto, foi necessária a seleção da metodologia de extração ideal de proteínas para análises em eletroforese bidimensional. Dentre os métodos avaliados, a extração de proteínas em solução de uréia e tiouréia resultou em melhor fracionamento protéico e obtenção de géis bidimensionais (2-D) com maior resolução e maior número de “spots”.

Pelo fato de *A. angustifolia* ser uma espécie cujo genoma ainda não é conhecido. A identificação de suas proteínas baseia-se principalmente na homologia com outras espécies vegetais. Adicionalmente, os alinhamentos obtidos por buscas restritivas, via MASCOT, são de pequena extensão e, portanto, faz-se necessária a utilização de estratégias alternativas de validação das identificações obtidas em bancos de dados. As buscas por similaridade de sequências são eficientes estratégias de identificação de proteínas de organismos não sequenciados (LISKA *et al.*, 2004; WARINDEL *et al.*, 2007). Entretanto, por basear-se em alinhamentos entre sequências degeneradas, repetitivas e redundantes, muitos alinhamentos não correspondem àqueles das proteínas alvo, aumentando o tempo de processamento computacional e a probabilidade de identificações falso-positivas. Sendo assim, foi desenvolvido um programa de seleção de espectros para remoção daqueles considerados como interferentes, sem a necessidade de prévia identificação dos mesmos via MASCOT ou MS BLAST. A plataforma de seleção do programa Eagle Eye é baseada em valores estatísticos de dissimilaridade entre espectros e, portanto, não depende de informações referentes à fragmentação dos peptídeos e/ou sequência de aminoácidos. Os resultados obtidos mostraram que o Eagle Eye é eficiente na remoção de espectros interferentes, ou seja, relacionados a identificações de queratina e tripsina, diminuindo o

tempo de busca substancialmente e aumentando a sensibilidade na identificação de proteínas alvo. Ressalta-se que, devido à possibilidade de carregamento de diferentes bibliotecas de interferentes, o Eagle Eye apresenta potencial de uso em outras estratégias proteômicas, como na identificação de interatores protéicos em análises do tipo “shotgun”, já que populações de proteínas podem ser co-isoladas junto ao complexo em estudo independente da especificidade entre isca e presa.

Similaridades morfológicas, fisiológicas e bioquímicas entre os processos de embriogênese somática e zigótica foram demonstradas para diferentes sistemas vegetais. Uma melhor compreensão dos fatores associados à embriogênese somática pode ser obtida na interface entre os modelos de embriogênese zigótica e somática. Em *Pinus*, foi demonstrado que além das semelhanças morfológicas, similaridades estão presentes no padrão de aminoácidos nos diferentes estádios da embriogênese (PULLMAN *et al.*, 2003). Mais recentemente, foram observadas similaridades nos processos de embriogênese somática e zigótica em *Cyclamem persicum*, através da proteômica comparativa (WINKELMANN *et al.*, 2006).

Embora muitos estudos abordando a indução e evolução de embriões somáticos de *A. angustifolia* tenham sido realizados, apenas as fases iniciais do processo de embriogênese somática nessa espécie encontram-se relativamente bem estabelecidos (DOS SANTOS *et al.*, 2002; SILVEIRA *et al.*, 2006; STEINER *et al.*, 2007). O estudo do desenvolvimento embrionário em sementes, além de permitir um melhor entendimento das bases moleculares e fisiológicas envolvidas no processo embriogenético, podem gerar informações importantes para a manipulação e otimização da técnica de multiplicação *in vitro* por embriogênese somática (SILVEIRA *et al.*, 2008).

Para *A. angustifolia*, as etapas de rearranjo morfológico e metabólico seguem aquelas descritas para as gimnospermas (CAIRNEY e PULLMAN, 2007). Segundo observado no Capítulo 4, as alterações no perfil de acúmulo de proteínas indicam que a embriogênese zigótica de *A. angustifolia* pode ser dividida em embriogênese inicial e tardia, sendo que o rearranjo metabólico que ocorre durante a formação dos cotilédones parece ser crucial para o correto desenvolvimento embrionário.

Na embriogênese somática de coníferas, embriões imaturos têm sido especificamente utilizados na indução de novas culturas *in vitro*. Além disso, foi observado que o estágio de desenvolvimento do inóculo afeta a taxa de indução observada (PARK *et al.*, 1993; LI e HUANG, 1996; SALAJOVA *et al.*, 1999). Na embriogênese somática de *A. angustifolia*, a escolha do estágio de desenvolvimento correto do explante é crucial na indução de novas culturas embriogênicas, tendo sido obtidos ótimos resultados quando inoculados embriões zigóticos nos estádios torpedo e cotiledonar inicial (DOS SANTOS *et al.*, 2002; SILVEIRA *et al.*, 2002). A restringência no uso de embriões jovens em *A. angustifolia* e em outras espécies de coníferas indica que a reprogramação do desenvolvimento embrionário é difícil em culturas *in vitro* (STASOLLA e YEUNG, 2003). Considerando os resultados obtidos no presente trabalho, pode ser inferido que o sucesso na obtenção de pró-embriões somáticos utilizando inóculos embrionários no estágio torpedo e/ou cotiledonar inicial possivelmente deve-se a um aumento nos níveis de expressão gênica, tradução de diversas classes protéicas e consequente ativação de diferentes vias metabólicas, durante esses estádios. Esse fato refletiu-se no grande número de “spots” observados nos géis do estágio cotiledonar inicial, seguido de redução no estágio maduro, caracterizando o final de uma etapa de intensa divisão celular para uma etapa de diferenciação e determinação celular.

Proteínas envolvidas no metabolismo basal, energético e processamento protéico são essenciais no desenvolvimento da semente de *A. angustifolia*. As proteínas do metabolismo basal são abundantes durante o início do desenvolvimento embrionário, ou seja, nos estádios caracterizados por intensa divisão celular e crescimento. Em *Araucaria*, existem espécies ortodoxas, como a *A. cunninghamii*, e recalcitrantes, como a *A. bidwillii* (GALLESCHI *et al.*, 2002). Devido ao elevado conteúdo de água e baixa tolerância à desidratação, as sementes de *A. angustifolia* são recalcitrantes (ESPINDOLA *et al.*, 1994; PANZA *et al.*, 2002). A grande quantidade de “spots” relacionados às identificações do metabolismo basal e energético no embrião maduro indica a existência de um metabolismo ativo no embrião maduro. Contrariamente às ortodoxas, as sementes recalcitrantes possuem alto conteúdo de água e são metabolicamente ativas quando maduras (PAMMENTER e

BERJAK, 2000). Embora tenham sido verificados padrões eletroforéticos similares ao observado para sementes ortodoxas, a grande quantidade de proteínas envolvidas no metabolismo basal e energético no estágio maduro corrobora a natureza recalcitrante da espécie e sugere uma estratégia de desenvolvimento contínuo sem a interpolação de uma fase de baixa atividade metabólica (PANZA *et al.*, 2002). Tais constatações podem ser muito importantes no estabelecimento e otimização de protocolos de embriogênese somática na espécie, já que pouco se sabe sobre o perfil protéico de coníferas recalcitrantes e grande quantidade de protocolos de embriogênese *in vitro*, incluindo-se aquele desenvolvido para *A. angustifolia* (GUERRA *et al.*, 2002; STEINER *et al.*, 2007), são baseados no perfil de acúmulo de proteínas em espécies ortodoxas.

Além disso, no presente trabalho, foram observados perfis eletroforéticos semelhantes de distribuição e acúmulo de proteínas entre embriões e megagametófitos, principalmente nos tecidos maduros. Tal fato pode refletir uma sinalização, e conversão mútua de proteínas, entre o embrião e o tecido que o envolve. A análise dos géis dos estádios cotiledonar tardio e maduro indicam que as proteínas de reserva, utilizadas para o crescimento da plântula, acumulam-se tanto no embrião, como no megagametófito. Sendo assim, a adição de agentes promotores de maturação, que induzam a síntese de proteínas de reserva do tipo vicilina, é esperada para a obtenção de embriões somáticos maduros. A adição de ácido glutâmico é sugerida, já que a análise da sequência primária da vicilina indica uma predominância desse resíduo de aminoácido. Além disso, as identificações do tipo vicilina foram as mais abundantes no presente trabalho. Esta situação reflete sua importância nos estádios finais do desenvolvimento embrionário de *A. angustifolia* e o seu potencial como marcador protéico no estabelecimento e controle da maturação de embriões somáticos.

### **Referências bibliográficas**

Cairney, J.,G. S. Pullman. The cellular and molecular biology of conifer embryogenesis. New Phytologist, v.176, p.511-536. 2007.

- Carpentier, S. C., E. Witters, K. Laukens, P. Deckers, R. Swennen, B. Panis. Preparation of protein extracts from recalcitrant plant tissues: and evaluation of different methods for two-dimensional gel electrophoresis. Proteomics, v.5, p.2497-2507. 2005.
- Dos Santos, A. L. W., V. Silveira, N. Steiner, M. Vidor, M. P. Guerra. Somatic embryogenesis in Parana Pine (*Araucaria angustifolia* (Bert.) O. Kuntze). Brazilian Archives of Biology and Technology, v.45, p.97-106. 2002.
- Espindola, L. E., M. Noin, F. Corbineau, D. Côme. Cellular and metabolic damage induced by desiccation in recalcitrant *Araucaria angustifolia* embryos. Seed Science Research, v.4, p.193-201. 1994.
- Galleschi, L., A. Capocchi, S. Ghiringhelli, F. Saviozzi. Some properties of proteolytic enzymes and storage proteins in recalcitrant and orthodox seeds of *Araucaria*. Biologia Plantarum, v.45, p.539-544. 2002.
- Guerra, M. P., V. Silveira, M. S. Dos Reis, L. Schneider. Exploração, manejo e conservação da araucária (*Araucaria angustifolia*). In: L. L. Simões, Lino, C. F. (Ed.). Sustentável Mata Atlântica: a exploração se seus recursos florestais. São Paulo: Editora Senac, 2002. Exploração, manejo e conservação da araucária (*Araucaria angustifolia*), p.85-101.
- Li, X. Y., F. H. Huang. Induction of somatic embryogenesis in loblolly pine (*Pinus taeda* L.). In Vitro Cellular & Developmental Biology-Plant, v.32, p.129-135. 1996.
- Liska, A. J., A. V. Popov, S. Sunyaev, P. Coughlin, B. Habermann, A. Shevchenko, P. Bork, E. Karsenti, A. Shevchenko. Homology-based functional proteomics by mass spectrometry: Application to the *Xenopus* microtubule-associated proteome. Proteomics, v.4, p.2707-2721. 2004.
- Pammenter, N. W., P. Berjak. Evolutionary and ecological aspects of recalcitrant seed biology. Seed Science Research, v.10, p.301-306. 2000.
- Panza, V., V. Lainez, H. Maroder, I. Prego, S. Maldonado. Storage reserves and cellular water in nature seeds of *Araucaria angustifolia*. Botanical Journal of the Linnean Society, v.140, p.273-281. 2002.
- Park, Y. S., S. E. Pond, J. M. Bonga. Initiation of somatic embryogenesis in white spruce (*Picea glauca*) - genetic-control, culture treatment effects, and implications for tree breeding. Theoretical and Applied Genetics, v.86, p.427-436. 1993.
- Pullman, G. S., S. Johnson, G. Peter, J. Cairney, C. Xu. Improving loblolly pine somatic embryo maturation: comparison of somatic and zygotic embryo morphology, germination, and gene expression. Plant Cell Reports, v.21, p.747-758. 2003.
- Salajova, T., J. Salaj, A. Kormutak. Initiation of embryogenic tissues and plantlet regeneration from somatic embryos of *Pinus nigra* Arn. Plant Science, v.145, p.33-40. 1999.
- Silveira, V., C. Santa-Catarina, T. S. Balbuena, F. M. S. Moraes, C. A. O. Ricart, M. V. Sousa, M. P. Guerra, W. Handro, E. I. S. Floh. Endogenous abscisic acid and protein

- contents during seed development of *Araucaria angustifolia*. Biologia Plantarum, v.52, p.101-104. 2008.
- Silveira, V., C. Santa-Catarina, N. N. Tun, G. F. E. Scherer, W. Handro, M. P. Guerra, E. I. S. Floh. Polyamine effects on the endogenous polyamine contents, nitric oxide release, growth and differentiation of embryogenic suspension cultures of *Araucaria angustifolia* (Bert.) O. Ktze. Plant Science, v.171, p.91-98. 2006.
- Silveira, V., N. Steiner, A. L. W. Dos Santos, R. O. Nodari, M. P. Guerra. Biotechnology tools in *Araucaria angustifolia* conservation and improvement: inductive factors affecting somatic embryogenesis. Crop Breeding and Applied Biotechnology, v.2, p.463-470. 2002.
- Stasolla, C., E. C. Yeung. Recent advances in conifer somatic embryogenesis: improving somatic embryo quality. Plant Cell Tissue and Organ Culture, v.74, p.15-35. 2003.
- Steiner, N., C. Santa-Catarina, V. Silveira, E. I. S. Floh, M. P. Guerra. Polyamine effects on growth and endogenous hormones levels in *Araucaria angustifolia* embryogenic cultures. Plant Cell, Tissue and Organ Culture, v.89, p.55-62. 2007.
- Warindel, P., A. Frank, H. Thomas, V. Surendranath, S. Sunyaev, P. Pevzner, A. Shevchenko. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated de novo sequencing. Proteomics, v.7, p.2318-2329. 2007.
- Winkelmann, T., D. Heintz, A. Van Dorsselaer, M. Serek, H. P. Braun. Proteomic analyses of somatic and zygotic embryos and endosperm tissue of *Cyclamen persicum*. XXI International Eucarpia Symposium: ISHS Acta Horticulturae, 2006.