JOURNAL OF PROTEOMICS 72 (2009) 337–352

available at www.sciencedirect.com

www.elsevier.com/locate/jprot

Changes in the 2-DE protein profile during zygotic embryogenesis in the Brazilian Pine (*Araucaria angustifolia*)

Tiago S. Balbuena[a,*], *Vanildo Silveira*[b], *Magno Junqueira*[c], *Leonardo L.C. Dias*[a], *Claudete Santa-Catarina*[d], *Andrej Shevchenko*[c], *Eny I.S. Floh*[a]

[a]*Plant Cell Biology Laboratory, Institute of Biosciences, University of São Paulo, São Paulo, Brazil*
[b]*Biotechnology Laboratory, State University of North Fluminense, Rio de Janeiro, Brazil*
[c]*Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany*
[d]*Cellular and Tissue Research Laboratory, State University of North Fluminense, Rio de Janeiro, Brazil*

ARTICLE DATA

Article history:
Received 10 November 2008
Accepted 6 January 2009

Keywords:
LC-MS/MS
MS BLAST
Plant proteomics
Seed development
Somatic embryogenesis

ABSTRACT

Araucaria angustifolia is the only native conifer of economic importance in the Brazilian Atlantic Rainforest. Due to a clear-cutting form of exploitation this species has received the status of vulnerable. The aim of this work was to investigate and characterize changes in protein expression profile during seed development of this endangered species. For this, the proteome of developing seeds was characterized by 2-DE and LC-MS/MS. Ninety six proteins were confidently identified and classified according to their biological function and expression profile. Overaccumulated proteins in early seed development indicated a higher control on oxidative stress metabolism during this phase. In contrast, highly expressed proteins in late stages revealed an active metabolism, leading to carbon assimilation and storage compounds accumulation. Comprehensive protein expression profiles and identification of overaccumulated proteins provide new insights into the process of embryogenesis in this recalcitrant species. Considerations on the improvement and control of somatic embryogenesis through medium manipulation and protein markers screening using data generated are also discussed.

1. Introduction

Araucaria *angustifolia* (Bert) O. Ktze is the only native conifer of economic importance in the Brazilian Atlantic Rainforest. Originally, the forests of *A. angustifolia* covered an area of 20 million hectares in Brazil [1,2]. A clear-cutting form of exploitation and the spread of agriculture reduced their original area to 1–2% and drove this species to the status of vulnerable [2]. Seeds of *A. angustifolia* are recalcitrant, desiccation-sensitive, and non-dormant which currently hampers their storage and utilization in large-scale conservation programs [3,4].

Seed development is one the major subjects in plant physiological research. Increasing numbers of data are now retrievable from large-scale analyses of gene expression during this process. In the context of proteomics, however, most of available information is derived from studies of model or grain species, which presents a desiccation phase at the end of seed filling and a quiescent state at maturation [5–7]. The only attempt to study seed proteome in a recalcitrant tree species was carried out in *Fagus sylvatica*, a Fagaceae species, during dormancy breaking [8].

To date there have been few studies concerning to proteome analysis in conifers. Costa et al. [9] and Gion et al. [10] carried out

* *Corresponding author*. Universidade de São Paulo, Instituto de Biociências, Rua do Matão 277, Cidade Universitária, CP11461, 05422-970, São Paulo, SP, Brazil. Tel.: +55 11 30917556; fax: +55 11 30917547.
E-mail address: tsbalbuena@usp.br (T.S. Balbuena).

1874-3919/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2009.01.011

comparative analyses in wood forming tissue. Lippert et al. [11] and Wang et al. [12] studied plant host interactions in *Picea sitchensis* and *Pinus nigra*, while Fernando [13] and Wagner et al. [14] identified several proteins related to conifer sexual reproduction aspects. Dong et al. [15], studying the somatic embryogenesis of *Picea glauca*, reported the appearance of several new proteins in response to removal of phytohormones from the culture medium prior to abscisic acid stimulation during the maturation process in somatic embryos.

The only attempts to understand the changes in protein expression during the zygotic embryogenesis in conifer species were carried out in *P. glauca* [16], *Cupressus sempervirens* [17] and *A. angustifolia*, in which protein expression between the early and late stages of embryogenesis using 2-DE and PMF for protein identification was carried out [18] and patterns of chitinases and arabinogalactan proteins expression during seed development were also investigated [19]. As for protein changes on the whole level, there are no reports.

In the present study we conducted a proteomic analysis of *A. angustifolia* seed development and aimed to investigate and characterize changes in the protein expression patterns throughout the process studied. Protein expression profiling provided new insights into the process of embryogenesis in this and other recalcitrant seeds, and a background for future improvement and control of somatic embryogenesis for in vitro scale up propagation, through medium manipulation and protein markers screening.

2. Materials and methods

2.1. Plant material

Seeds of *A. angustifolia* were harvested in the Santa Catarina State, Brazil (27°47′S, 49°29′W), from December 2006 to May 2007 and embryos at the stages of proembryo (1), globular (2), torpedo (3), early-cotyledonary (4), late-cotyledonary (5) and mature (6) were morphologically determined according to Astarita et al. [20]. Due to the small size and the low protein recovery of proembryo, globular and torpedo embryos, these were together extracted with megagametophyte tissues. Megagametophytes and the zygotic embryos from stages 4, 5 and 6 were individualized under a dissecting microscope (Fig. 1). All materials were stored at −80 °C prior to analysis.

2.2. Protein extraction

Protein extracts were prepared in biological triplicates for each time point. Due to the changes in seed tissues fresh weight (FW) throughout seed development (Table 1S), each biological sample was prepared from a bulk of, at least, 15 seeds and grounded to a fine powder under liquid nitrogen. For protein extraction 300 mg of each grounded and mixed sample powders were transferred into clear 2 mL microtubes containing 1.5 mL of extraction buffer (7 M urea, 2 M thiourea, 1% DTT, 2% triton X-100, 0.5% pharmalyte (GE Healthcare, Freiburg, Germany), 1 mM PMSF, 5 μM pepstatin). All extracts were briefly vortexed and kept in the extraction buffer, standing on ice, for 30 min followed by centrifugation at 12,000 *g* for 5 min at 4 °C. The supernatants were transferred to clear microtubes and proteins were precipitated in ice, for 30 min, in trichloroacetic acid (10%) and washed three times with cold acetone. Finally, proteins were re-suspended and concentrated in 0.5 mL of the same extraction buffer, with addition of 0.5% immobilized pH gradient (IPG) buffer (pH 4–7) (GE Healthcare) instead of pharmalyte. Protein concentration was estimated by 2-D Quant Kit (GE Healthcare) and samples were stored at −20 °C until 2-DE.

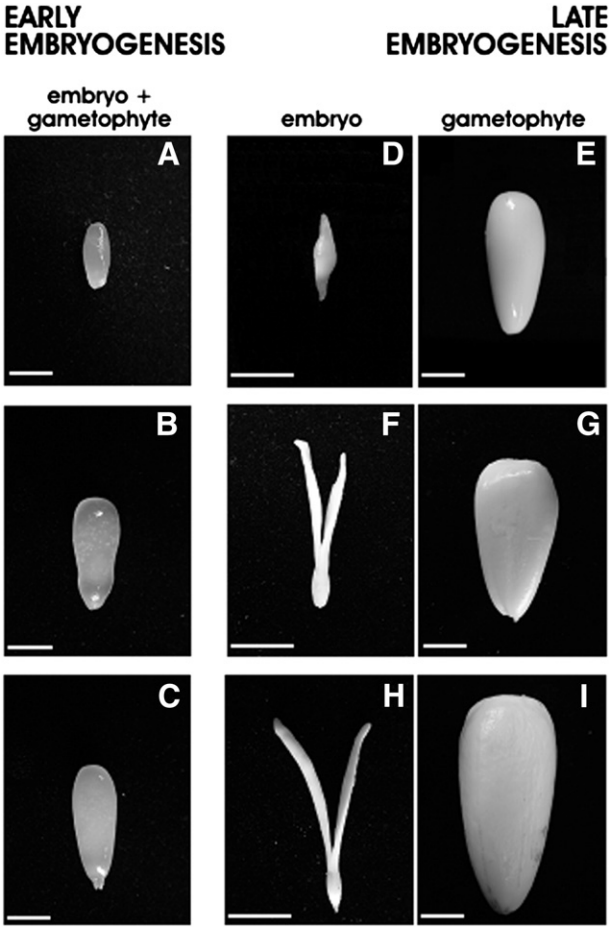

Fig. 1 – The development of *Araucaria angustifolia* tissues during zygotic embryogenesis. Proembryo (A), globular (B), torpedo (C), early-cotyledonary (D, E), late-cotyledonary (F, G) and mature (H, I) stages. Bars: 0.7 cm.

2.3. Two dimensional gel electrophoresis

Two 2D gels were performed to each biological sample. Sample aliquots containing 180 μg of proteins were used to 2-DE. Prior to this assay protein extracts were separated across a broad-range pH 3–10 IPG strip and resulted in low resolution gels, especially those from early seed development stages (Fig. 1S), containing spots preferentially distributed in the range of pH 4–7. Aiming a better 2-DE spot resolution for comparative analysis and gel excision for mass spectrometer analysis the use of IPGs with a linear separation range of pH 4–7 was chosen. Prior to loading in 18 cm IPG strips, small volumes of rehydration buffer (7 M urea, 2 M thiourea, 2% CHAPS, 0.5% IPG buffer (pH 4–7), 1% DTT, 0.002%

bromophenol blue) were added to the sample aliquots in order to achieve a final volume of 375 μL. After 12 h in gel rehydration, isoelectric focusing was performed on an IPGphor II apparatus (GE Healthcare) for a total of 35 kVh at 20 °C. IPG strips were then subjected to reduction, alkylation by 2×15 min incubations with buffer (50 mM Tris–HCl, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue) containing 125 mM DTT for the first incubation and 125 mM iodoacetamide for the second. Then, strips were applied to the top of a 12% polyacrylamide gel. Second dimension electrophoresis was carried out at 25 mA per gel in a Protean II apparatus (Bio-Rad, Hercules, US) and gels were silver-stained, according to Shevchenko et al. [21].

2.4. *Spot matching*

Duplicated silver-stained 2D gels from each of the three biological samples were analyzed using the Image Master Platinum v. 6 software (GE Healthcare). The spot detection parameters were optimized by checking different protein spots in certain regions of the gel and then automatically detected, followed by visual inspection for removal or addition of undetected spots. For 2-DE characterization representative in silico images were obtained, containing only spots that were detected in all six repetitions. For this, the processed gels were automatically matched in order to attribute a common spot identity for the same spot derived from different gels and visually inspected for improper spot matches. In silico images were then used for all 2-DE gels characterization.

2.5. *In gel digestion of proteins*

In-gel protein digestion was performed as described in Shevchenko et al. [22]. Individual protein spots were excised from a gel slab using a clean scalpel, cut into ca. 1×1 mm pieces and placed into 600 μL microtubes. Gel pieces were dehydrated in 500 μL acetonitrile for 10 min and then reduced in 50 μL of a DTT solution (10 mM DTT in 100 mM ammonium bicarbonate) for 30 min at 56 °C, followed by a dehydration step in acetonitrile for 10 min. For alkylation, 50 μL of 55 mM iodoacetamide in 100 mM ammonium bicarbonate was added and gel pieces incubated at room temperature in the dark. Prior to addition of sequencing grade porcine trypsin (Promega, Madison, USA) at 16 ng μL^{-1}, gel pieces were dehydrated by acetonitrile. After 120 min, samples were placed into an air circulation thermostat and incubated overnight at 37 °C. Upon in-gel digestion, gel pieces were saturated with 100 μL of extraction buffer 5% formic acid (FA): acetonitrile (1:2, v/v) and incubated for 15 min at 37 °C in shaker. Supernatants were then collected , pooled together and ... dried down in a vacuum centrifuge. For LC-MS/MS analyses 0.05% trifluoroacetic acid (TFA) (10 μL) was added into the dried tubes, incubated for 2 min and vortexed for 10 min at 7 *g*.

2.6. *LC-MS/MS*

LC-MS/MS was performed on an Ultimate 3000 nanoLC system (Dionex, Sunnyvale, USA), which was interfaced to a LTQ Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) via a robotic nanoflow ion source TriVersa (Advion BioSciences, NY, USA). Ionization voltage was set to 1.7 kV and the capillary transfer temperature was set at 180 °C. For each round of analysis 4 μL of the tryptic peptides were loaded onto 5 mm×300 μm id trapping column packed with C18 PepMAP100 5 μm particles (Dionex) in 0.05% TFA at the flow rate of 20 μL min^{-1}. After 6 min peptides were eluted into 15 cm×75 μm id nanocolumn packed with C18 PepMAP100 3 μm particles (Dionex) at the flow rate of 200 nL min^{-1} and separated using the following mobile phase gradient: from 5 to 20% of solvent B in 20 min, 20 to 50% B in 16 min, 50 to 100% B in 5 min, 100% B during 10 min, and back to 5% B in 10 min. Solvent A was 95:5 water:acetonitrile (v/v) containing 0.1% FA; solvent B was 20:80 water:acetonitrile (v/v) containing 0.1% FA. MS data were acquired in data-dependent acquisition (DDA) mode controlled by Xcalibur 2.0 software (Thermo Fisher Scientific). The automated gain control (AGC) was set to 5×10^5 charges for survey scan on the Orbitrap and 5×10^4 charges for MS/MS on the ion trap analyzers. Typical data-dependent acquisition (DDA) cycle consisted of a survey scan within *m/z* 300 to 1600 performed at the Orbitrap analyzer under the target mass resolution of 60,000 (FWHM, full width at half maximum) followed by MS/MS fragmentation of the four most abundant precursor ions under the normalized collision energy of 35% in the linear trap. Singly charged ions were excluded from MS/MS experiments, and *m/z* of fragmented precursor ions were dynamically excluded for further 90 s. MS/MS spectra were exported as .dta files, using BioWorks 3.1 software (Thermo Fisher Scientific).

2.7. *MS data analysis*

MS/MS spectra queries were filtered against a library of more than 15,000 non-annotated background spectra using EagleEye software as described in Junqueira et al. [23]. Background spectra recognized by EagleEye were removed, while the remaining genuine spectra were merged into a single .mgf (MASCOT generic format) file and searched against a MSDB database (2.344.227 sequence entries; updated April, 2006) by MASCOT v. 2.1 software (Matrix Science Ltd., London, UK) installed on a local 2 CPU server. Tolerances for precursor and

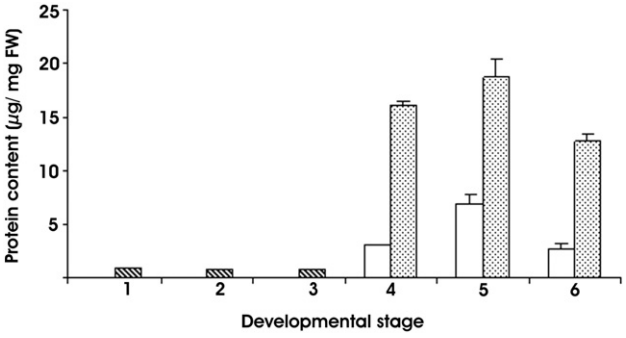

Fig. 2 – Protein content from zygotic embryogenesis of *Araucaria angustifolia* at different developmental stages. Stage of development: 1 = proembryo, 2 = globular, 3 = torpedo, 4 = early-cotyledonary, 5 = late-cotyledonary, 6 = mature. Crossed boxes indicate intact megagametophyte containing embryo, the dashed and blank boxes indicate isolated embryo and megagametophyte tissues, respectively.

fragment masses were set at 10 parts per million (ppm) and 0.6 Da, respectively. Up to 2 missed cleavages were allowed and the following parameters were used for database searches: instrument profile: ESI-Trap; fixed modification: carbamidomethyl (cysteine); variable modification: oxidation (methionine) and acetylation of the N-terminal peptide of protein sequence entry were set as variable modifications. MASCOT identifications of proteins were considered confident if hits were produced by matching of at least three MS/MS spectra with peptide ions scores above 20. For hits matched by one or two spectra it was required that at least one spectrum should be matched with score of 50 or better [24].

In parallel to stringent database search, de novo sequencing was performed as described by Waridel et al. [25]. After EagleEye spectra filtering the remaining MS/MS spectra (represented by .dta files) obtained in each LC-MS/MS experiment were interpreted de novo by a modified version of PepNovo software [26]. For each spectrum, PepNovo reported the expected confidence of produced sequence candidates by assigning a quality score, which stands for the expected number of confidently called amino acid residues. Unless specified otherwise, in de novo sequencing experiments, we only considered sequence candidates having the score of 6.0 or above [27]. For MS BLAST searches, all selected peptide sequence candidates obtained by PepNovo sequencing of all peptide precursors, were merged into a single MS BLAST query string and searches were performed against nr database at the web-accessible server at http://genetics.bwh.harvard.edu/msblast, applying LC-MS/MS presets as described by Junqueira et al. [28].

2.8. *Protein expression profiling*

To access information on similarities among protein expression profiles, we used the hierarchical clustering software EPCLUST, available at URL: http://www.bioinf.ebc.ee/EP/EP/EPCLUST/. For this, normalized spot volumes from representative in silico images from the same protein accession within

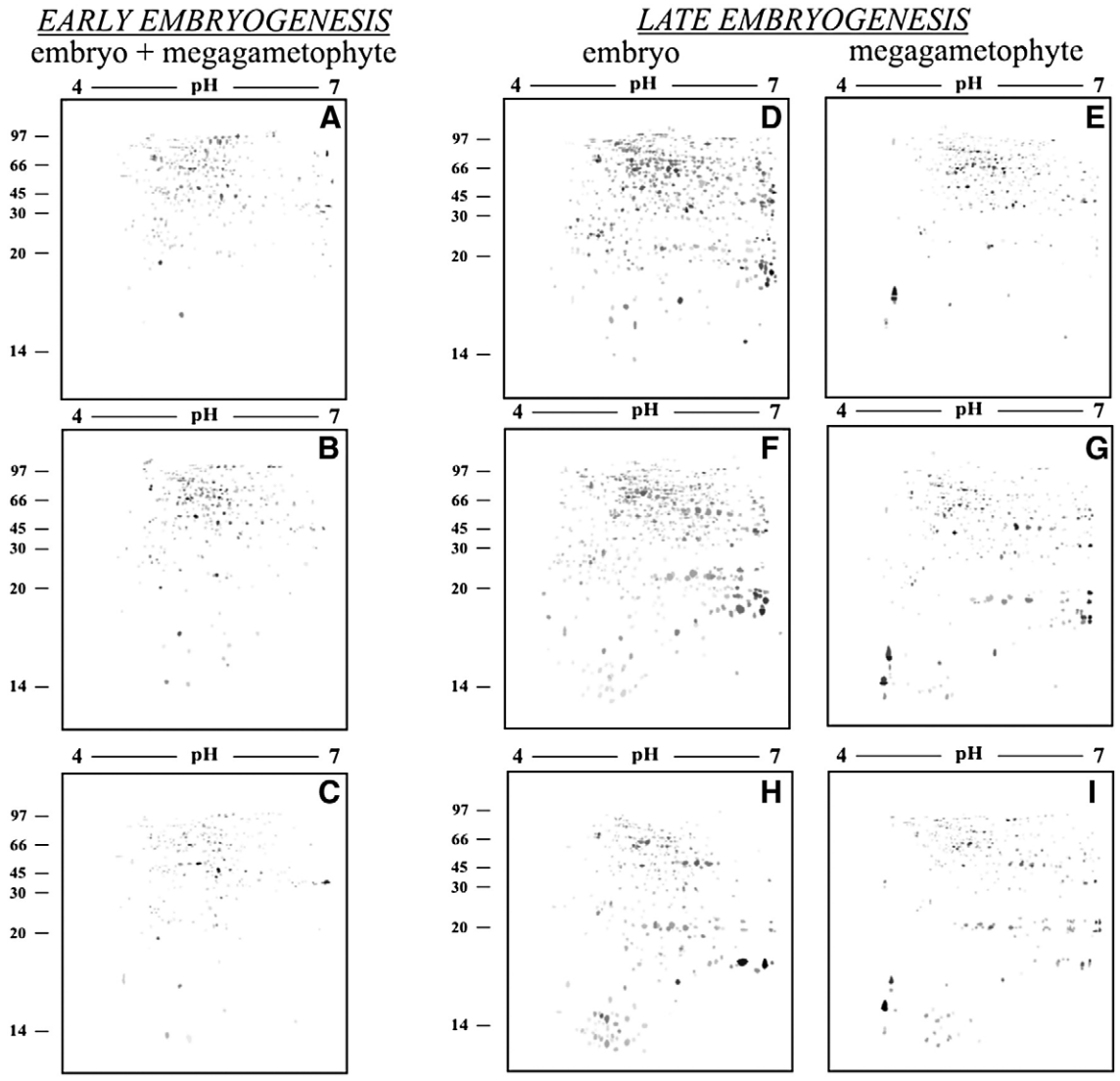

Fig. 3 – In silico 2-DE images of proteins extracted during *Araucaria angustifolia* seed development. Images represent the expression of protein profiles in proembryo (A), globular (B), torpedo (C), early-cotyledonary (D, E), late-cotyledonary (F, G), mature (H, I) stages. The pH gradient and position of Mr are indicated at the top and sides of gel images, respectively.

Table 1 – Number of common spots detected in different stage gels during zygotic embryogenesis in *Araucaria angustifolia*.

Stage of development[a, b, c]	Common spots[b]		
	EM	E	M
Early embryogenesis			
1 (EM: 426)–2 (EM: 425)	203	–	–
2 (EM: 425)–3 (EM: 353)	179	–	–
Late embryogenesis			
4 (E: 955, M: 545)–5 (E: 773, M: 473)	–	414	167
5 (E: 773, M: 473)–6 (E: 496, M: 427)	–	235	175

Spots presented in each in silico images were detected in all six gels repetitions.
[a] 1=proembryo, 2=globular, 3=torpedo, 4=early-cotyledonary, 5=late-cotyledonary, 6=mature.
[b] Number of detected spots in each in silico image is presented in parenthesis.
[c] EM = embryo and megagametophyte, E = embryo, M = megagametophyte.

and between embryo and megagametophyte tissues were summed and converted to $\log_2$ transformed data. The Euclidian distances and UPGMA algorithm were used for the analysis.

3. Results and discussion

3.1. Protein content and two-dimensional gels characterization

Measurements of extracted protein content during *A. angustifolia* seed development revealed an increase of almost 17-fold in protein content (Fig. 2). The highest amount of protein was detected during seed filling, when embryos reached the late-cotyledonary stage (19 μg mg^{-1} FW). Although this species produce starchy seeds, Panza et al. [3] demonstrated the presence of protein storage vacuoles in *A. angustifolia* mature cotyledons and embryo axis, suggesting that embryos accumulate storage proteins. Due to this feature and the absence of a quiescence stage in this species [4] a plateau in protein content was expected after seed filling; however, protein profile showed a decrease in protein levels in mature seeds, indicating that the protein burst observed in late-cotyledonary stage may be mostly due to an increase in translational processes, and possible of metabolic activity, that takes place during seed filling. In accordance with this result, total protein spot detection analysis of representative in silico images (Fig. 3) showed that spot number increased almost three folds between torpedo tissues (353 spots) and early-cotyledonary stage embryo (955 spots) gels (Table 1).

3.2. Protein identification using combined MASCOT and MS BLAST searches

To ensure a good overview, in terms of pH and molecular weight, and to assess changes in metabolic protein profiles during *A. angustifolia* seed development, the 155 most abundant spots from globular and mature 2-DE gels were excised and in gel digested using trypsin (Fig. 4). For protein identification, MASCOT stringent database search and MS BLAST driven sequence similarity database search were used in a combined approach [28]. This combined strategy of peptide sequence analysis resulted in the identification of 56 spots in globular stage and 81 mature stage gels, from which 15 and 40 spots were exclusively detected in globular and mature gel stages, respectively (Table 2). This represents a

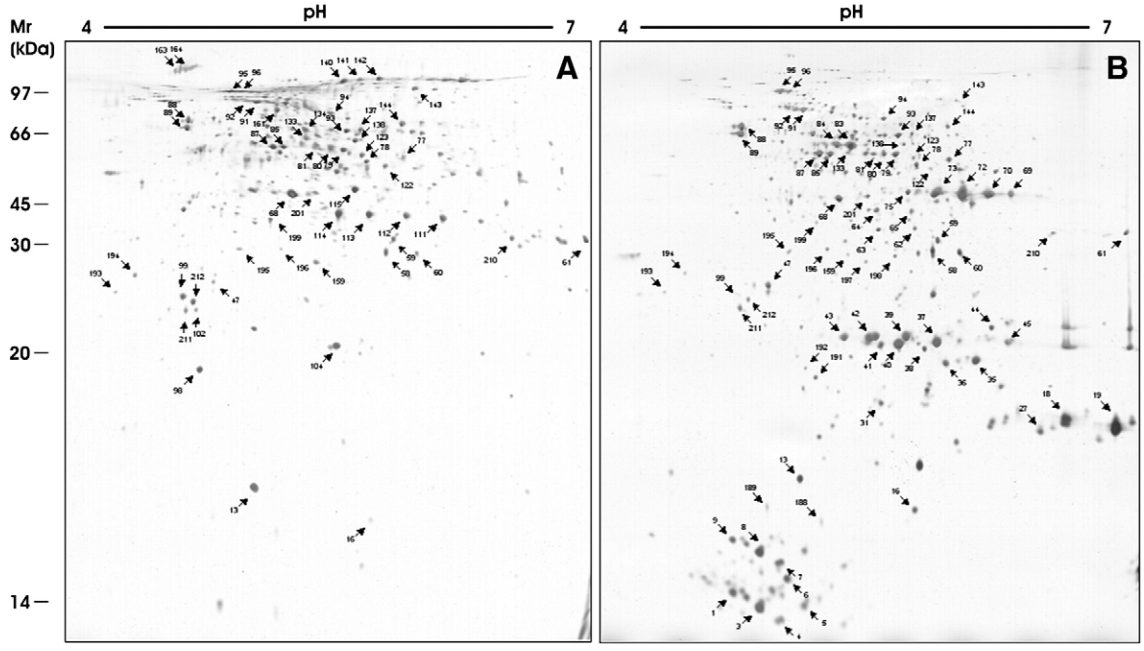

Fig. 4 – Reference 2-DE maps for globular (embryo+megagametophyte) and mature (embryo) tissues of *Araucaria angustifolia* seed development (pH 4–7 linear gradient). Proteins that were identified are marked with arrows and numbered in following Table 2.

Table 2 – List of identified proteins during seed development in *Araucaria angustifolia* via combined MASCOT stringent search and MS BLAST sequence-similarity search.

Spot[a]	Stage[b]	MASCOT search						MS BLAST search					MW (kDa)		pI	
		Protein name	Accession number[c]	Organism	Score[d]	Unique peptides[e]	Coverage[f] (%)	Protein name	Accession number[c]	Organism	Matched queries[g]	Coverage[h] (%)	Theory	Observation	Theoretical	Experimental
Protein fate																
16	G, M	Heat shock protein 17.0	Q40851	*P. glauca*	94	2	11	Heat shock protein 17.0	Q40851	*P. glauca*	3	23	17	17	5.8	5.7
36	M	Chloroplast chaperonin 21	Q6B4V3	*V. vinifera*	70	1	11	Putative chaperonin 21	Q6Y679	*P. vittata*	1	17	13	27	4.6	5.84
37	M	Peptidyl-prolyl cis–trans isomerase	P21568	*S. lycopersicum*	74	1	8	Peptidyl-prolyl cis–trans isomerase CYP20-2	Q9ASS6	*A. thaliana*	3	15	18	29	8.8	5.8
63	M	Putative elongation factor 2	Q6H4L2	*O. sativa*	164	3	3	OSJNBa0020P07.3 protein	Q7XTK1	*O. sativa*	4	6	94	41	5.9	5.5
75	M	Protein disulfide isomerase	Q5EUD6	*Z. mays*	146	3	4	Putative disulfide-isomerase	Q75M08	*O. sativa*	3	9	40	47	6.3	5.6
88	G, M	Putative rubisco subunit binding-protein alpha subunit	Q8L5U4	*A. thaliana*	236	4	7	RuBisCO large subunit-binding protein subunit alpha	P21238	*A. thaliana*	10	19	62	63	5.1	4.7
89	G, M	Putative rubisco subunit binding-protein alpha subunit	Q8L5U4	*A. thaliana*	391	5	9	RuBisCO large subunit-binding protein subunit alpha	P21238	*A. thaliana*	21	43	62	61	5.1	4.7
91	G, M	Hsc70	Q40151	*L. esculentum*	454	8	20	Heat shock protein 70	Q40693	*O. sativa*	6	11	72	70	5.2	5.0
92	G, M	Hsc70	Q40151	*L. esculentum*	318	6	14	Heat shock protein 70	Q40693	*O. sativa*	4	8	72	70	5.2	4.9
93	G, M	Heat shock protein 60	Q8H6U4	*P. dulcis*	485	7	13	Heat shock protein 60	Q8H6U4	*P. dulcis*	16	33	58	61	5.3	5.6
95	G, M	Molecular chaperone Hsp90-1	Q6UJX6	*N. benthamiana*	975	20	22	Heat shock protein 83	P51819	*I. nil*	19	30	80	80	4.9	4.9
96	G, M	Heat shock protein	Q71EE1	*H. brasiliensis*	913	18	19	Heat shock protein 81-1	Q0J4P2	*O. sativa*	19	30	80	80	5.0	4.9
115	G	Protein disulfide isomerase	Q5EUD6	*Z. mays*	65	2	6	Probable protein disulfide-isomerase	P38661	*M. sativa*	3	8	40	52	6.3	5.6
133	G, M	Leucine aminopeptidase 1	P30184	*A. thaliana*	245	5	10	Leucine aminopeptidase 1	P30184	*A. thaliana*	9	20	55	56	5.7	5.3
134	G	Heat shock protein 60	Q8H6U4	*P. dulcis*	116	3	4	Heat shock protein 60	Q8H6U4	*P. dulcis*	5	10	58	70	5.3	5.4
137	G, M	Putative TCP-1	Q6AV23	*O. sativa*	432	8	17	Putative TCP-1	Q6AV23	*O. sativa*	16	34	58	61	5.6	5.7
159	G, M	Putative elongation factor 2	Q6H4L2	*O. sativa*	164	3	3	OSJNBa0020P07.3 protein	Q7XTK1	*O. sativa*	4	6	94	38	5.9	5.3
161	G	NI	–	–	–	–	–	70 kDa peptidyl-prolyl isomerase	Q38931	*A. thaliana*	3	5	62	82	5.2	5.2
188	M	NI	–	–	–	–	–	40S ribosomal protein S12	Q9XHS0	*H. vulgare*	5	41	15	16	5.4	5.1
189	M	Heat shock protein 17.0	Q40851	*P. glauca*	122	3	15	Heat shock protein 17.0	Q40851	*P. glauca*	4	32	17	17	5.8	4.8
192	M	20 kDa chaperonin	O65282	*A. thaliana*	110	3	6	Chloroplast chaperonin 21	Q6B4V4	*V. vinifera*	8	36	27	26	8.9	5.1
195	G, M	Putative 60S acidic ribosomal protein P0	Q8LNX8	*Z. elegans*	241	5	36	60S acidic ribosomal protein P0	P50345	*L. luteus*	12	39	22	39	9.6	4.9
196	G, M	Putative 60S acidic ribosomal protein P0	Q8LNX8	*Z. elegans*	255	6	38	60S acidic ribosomal protein P0	O24573	*Z. mays*	12	44	22	39	9.6	5.2
210	G, M	Translational elongation factor Tu	Q851Y8	*O. sativa*	503	9	22	Translational elongation factor Tu	Q8W2C4	*O. sativa*	12	27	48	41	6.0	6.5

Metabolism																
47	G, M	Putative uncharacterized protein	O04428	*C. paradisi*	291	5	17	Putative uncharacterized protein	O04428	*C. paradisi*	10	34	33	34	5.5	4.8
60	G, M	Cysteine synthase	P32260	*S. oleracea*	131	4	8	Cysteine synthase	Q9XEA6	*O. sativa*	7	7	41	38	6.8	5.9
64	M	Glutamine synthetase	Q9AXD8	*A. marina*	498	10	17	Glutamine synthetase	Q9AXD8	*A. marina*	12	36	39	44	6.2	5.4
72	M	GDP-mannose-3, 5-epimerase	Q93VR3	*A. thaliana*	143	3	7	GDP-mannose 3, 5-epimerase	Q93VR3	*A. thaliana*	1	4	43	47	5.9	5.9
73	M	GDP-mannose-3, 5-epimerase	Q93VR3	*A. thaliana*	395	7	20	GDP-mannose-3, 5-epimerase	Q93VR3	*A. thaliana*	4	14	43	47	5.9	5.8
77	G, M	Adenosylhomocysteinase	Q4H1G1	*B. vulgaris*	211	5	13	Adenosylhomocysteinase	P93253	*M. crystallinum*	3	6	54	55	5.8	5.9
83	M	Granule-bound glycogen (Starch) synthase	Q9ZSQ5	*A. membranaceus*	213	4	6	Granule-bound glycogen (Starch) synthase	Q9ZSQ5	*A. membranaceus*	8	16	67	58	7.6	5.3
84	M	Granule-bound starch synthase 1	Q00775	*S. tuberosum*	60	1	3	Granule-bound starch synthase 1	Q43092	*P. sativum*	2	4	67	58	6.9	5.2
94	G, M	Cytosolic phosphoglucomutase	Q6S3D6	*P. tomentosa*	459	6	14	Phosphoglucomutase	P93804	*Z. mays*	15	27	63	66	5.5	5.5
112	G	Type IIIa membrane protein cp-wap13	O24548	*V. unguiculata*	124	3	8	Type IIIa membrane protein cp-wap13	O24548	*V. unguiculata*	3	10	39	46	6.2	5.9
113	G	Type IIIa membrane protein cp-wap13	O24548	*V. unguiculata*	358	8	30	Type IIIa membrane protein cp-wap13	O24548	*V. unguiculata*	4	14	39	46	6.2	5.7
114	G	Type IIIa membrane protein cp-wap13	O24548	*V. unguiculata*	323	8	28	Putative reversibly glycosylatable polypeptide	Q9SR90	*A. thaliana*	3	10	39	46	6.2	5.6
123	G, M	Adenosylhomocysteinase	Q4H1G1	*B. vulgaris*	508	11	18	Adenosylhomocysteinase 1	O23255	*A. thaliana*	8	16	54	55	5.8	5.7
138	G, M	Betaine-aldehyde dehydrogenase	Q6DQ92	*M. acuminata*	66	1	5	Betaine-aldehyde dehydrogenase	Q6DQ92	*M. acuminata*	6	26	29	59	6.3	5.6
140	G	NI	–	–	–	–	–	Putative alpha-xylosidase	Q8VWV9	*P. pinaster*	2	31	101	104	6.3	5.6
141	G	Alpha-glucosidase 1	O22444	*A. thaliana*	59	1	1	Alpha-glucosidase 1	O22444	*A. thaliana*	1	1	101	105	5.6	5.7
142	G	Alpha glucosidase 1	O22444	*A. thaliana*	54	1	1	Alpha-glucosidase 1	O22444	*A. thaliana*	2	2	101	106	5.6	5.8
143	G, M	Cobalamin-independent methionine synthase	Q6KCR0	*A. thaliana*	131	2	3	Cobalamin-independent methionine synthase	Q6KCR0	*A. thaliana*	2	3	91	76	8.2	5.9
199	G, M	Putative adenosine kinase	Q6SV73	*P. tremula*	612	9	23	Putative adenosine kinase	Q6SV73	*P. alba*	8	38	25	42	6.0	5.1
201	G, M	S-adenosylmethionine synthetase	Q9FVG7	*P. contorta*	802	11	39	S-adenosylmethionine synthetase	Q9FVG7	*P. contorta*	14	43	43	47	5.6	5.4
Energy																
41	M	Triosephosphate isomerase	P48496	*S. oleracea*	293	3	13	Triosephosphate isomerase	P48496	*S. oleraceae*	12	42	34	28	6.5	5.5
43	M	Triosephosphate isomerase	P48496	*S. oleracea*	106	1	4	Triosephosphate isomerase	P46225	*S. cereale*	2	8	34	29	6.5	5.3
58	G, M	Malate dehydrogenase	Q5QLS8	*O. sativa*	50	1	3	Malate dehydrogenase	Q42972	*O. sativa*	1	3	42	39	7.6	7.8
59	G, M	Nodule-enhanced malate dehydrogenase	O81278	*G. max*	440	7	19	Nodule-enhanced malate dehydrogenase	O81278	*G. max*	10	23	44	40	6.9	5.8
69	M	NADP specific isocitrate dehydrogenase	Q9ZWI1	*D. carota*	646	10	23	NADP specific isocitrate dehydrogenase	Q9ZWI1	*D. carota*	10	28	46	47	6.5	6.2
70	M	NADP specific isocitrate dehydrogenase	Q9ZWI1	*D. carota*	440	8	18	NADP specific isocitrate dehydrogenase	Q9ZWI1	*D. carota*	11	30	46	47	6.5	6.1
78	G, M	Enolase	Q42971	*O. sativa*	88	2	8	Enolase 1	Q9LEJ0	*H. brasiliensis*	2	6	48	53	5.4	5.7
79	G, M	UTP-glucose-1-phosphate uridylyltransferase	Q43772	*H. vulgare*	257	6	15	UTP-glucose-1-phosphate uridylyltransferase	Q9SDX3	*M. acuminata*	9	21	52	56	5.2	5.6
81	G, M	UDP-glucose pyrophosphorylase	Q8W557	*A. fructicosa*	182	3	7	UDP-glucose pyrphosphorylase	Q8W557	*A. fructicosa*	4	9	52	56	6.1	5.4
85	G, M	ATP synthase subunit beta	Q01859	*O. sativa*	259	5	13	ATP synthase subunit beta	P29685	*H. brasiliensis*	7	14	59	56	5.9	5.2
87	G, M	ATP synthase subunit beta	Q01859	*O. sativa*	370	6	17	ATP synthase subunit beta	P19023	*Z. mays*	8	17	59	56	5.9	5.1
111	G	Enolase	Q6WB92	*G. barbadense*	86	2	5	Enolase	Q43130	*M. crystallinum*	2	6	48	45	6.2	6.1
122	G, M	Enolase	Q42971	*O. sativa*	252	4	12	Enolase	Q42971	*O. sativa*	3	10	48	53	5.4	5.8

(continued on next page)

Table 2 (*continued*)

Spot[a]	Stage[b]	MASCOT search						MS BLAST search					MW (kDa)		pI	
		Protein name	Accession number[c]	Organism	Score[d]	Unique peptides[e]	Coverage[f] (%)	Protein name	Accession number[c]	Organism	Matched queries[g]	Coverage[h] (%)	Theory	Observation	Theoretical	Experimental
Energy																
144	G, M	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	P35493	*R. communis*	97	3	5	Phosphoglycerate mutase	Q9XE59	*S. tuberrosum*	2	4	61	63	5.5	5.9
193	G, M	ATP synthase gamma chain	Q59I53	*I. nil*	152	3	8	ATP synthase subunit gamma	P26360	*I. batatas*	3	10	36	34	9.0	4.2
198	M	Fructose-bisphosphate aldolase	Q9LLD7	*O. sativa*	132	3	8	Fructose-bisphosphate aldolase	Q53P96	*O. sativa*	2	6	42	38	9.0	5.6
Storage process																
1	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	587	10	20	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	9	23	53	13	7.7	4.6
3	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	998	18	22	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	9	24	53	13	7.7	4.8
4	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	552	8	17	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	8	20	53	12	7.7	4.9
5	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	518	9	18	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	10	26	53	13	7.7	5.0
6	M	NI	–	–	–	–	–	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	2	4	53	14	7.7	4.9
7	M	High molecular weight glutenin subunit 1By15	Q4JHY1	*T. aestivum*	111	2	2	High molecular weight glutenin y-subunit	Q7Y0S9	*T. elongatum*	1	2	78	14	8.6	4.9
8	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	508	10	18	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	6	14	53	15	7.7	4.8
9	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	433	8	15	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	8	20	53	15	7.7	4.6
18	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	1035	22	36	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	13	31	53	22	7.7	6.5
19	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	1100	24	35	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	11	27	53	22	7.7	6.8
27	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	905	21	32	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	9	24	53	22	7.7	6.4
31	M	NI	–	–	–	–	–	Vicilin-like storage protein	Q40844	*P. glauca*	3	7	51	24	7.9	5.5
35	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	175	4	8	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	6	13	53	27	7.7	6.0
38	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	566	10	23	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	11	18	53	28	7.7	5.7
39	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	244	7	23	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	10	26	53	29	7.7	5.6
40	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	277	7	15	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	9	23	53	29	7.7	5.6
42	M	Vicilin	Q8LKI7	*A. angustifolia*	218	5	12	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	10	26	53	29	7.7	5.4
44	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	167	5	12	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	1	2	53	30	7.7	6.1
45	M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	304	6	12	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	5	11	53	29	7.7	6.2
61	G, M	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	727	13	27	Vicilin-like storage protein	Q8LKI7	*A. angustifolia*	7	15	53	41	7.7	6.9
Stress response and detoxification																
13	G, M	Peroxiredoxin	Q8S3L0	*P. tremula*	116	3	16	Peroxiredoxin	Q8S3L0	*P. tremula*	6	40	17	19	5.6	5.0
62	M	Putative aldose reductase	Q65WW3	*O. sativa*	107	3	5	Putative aldose reductase	Q8GXW0	*A. thaliana*	6	22	36	42	6.3	5.7

80	G, M	Putative mitochondrial aldehyde dehydrogenase ALDH2a	Q6YWQ9	*O. sativa*	57	1	3	Mitochondrial aldehyde dehydrogenase	Q8LST5	*S. bicolor*	1	2	46	56	5.9	5.5
98	G	Cytosolic ascorbate peroxidase 2	Q76LA6	*G. max*	55	1	7	Cytosolic ascorbate peroxidase	Q8H1K7	*R. raetam*	1	6	27	28	5.7	4.8
104	G	Ascorbate peroxidase	Q6RY58	*P. pinaster*	218	4	17	Ascorbate peroxidase	Q6RY58	*P. pinaster*	4	19	27	30	5.4	5.5
Cellular signaling																
99	G, M	14-3-3-like protein GF14 iota	Q8LEN1	*A. thaliana*	166	3	13	14-3-3-like protein D	Q96453	*G. max*	2	8	29	34	4.9	4.7
102	G	14-3-3-like protein GF14 iota	Q8LEN1	*A. thaliana*	113	2	10	14-3-3 protein	Q944P2	*F. cirrhosa*	2	12	29	34	4.9	4.8
211	G, M	14-3-3 protein	Q6PWL7	*S. chacoense*	306	6	22	14-3-3-like protein	Q9SP07	*L. longiflorum*	4	19	29	32	4.7	4.7
212	G, M	14-3-3-like protein GF14 iota	Q8LEN1	*A. thaliana*	161	3	13	14-3-3 protein	Q944P2	*F. cirrhosa*	3	17	29	33	4.9	4.7
Cellular transport																
163	G	NI	–	–	–	–	–	Patellin-3	Q56Z59	*A. thaliana*	2	5	56	113	5.2	4.7
164	G	NI	–	–	–	–	–	Patellin-3	Q56Z59	*A. thaliana*	2	5	56	113	5.2	4.7
Cell cycle and DNA processing																
194	G, M	Proliferating cell nuclear antigen	P17070	*O. sativa*	376	9	37	Proliferating cell nuclear antigen large form	Q00265	*D. carota*	17	49	29	36	4.6	4.3
Structure																
68	G, M	Actin	Q9SP17	*P. rubens*	1288	23	62	Actin	P93485	*P. sativum*	18	55	83	46	6.0	5.2
197	M	Tubulin beta-2 chain	P18026	*Z. mays*	291	5	12	Tubulin beta-4 chain	Q41782	*Z. mays*	3	7	50	37	4.8	5.4
Unclassified																
65	M	NI	–	–	–	–	–	F13E7.34 protein	Q9M8R4	*A. thaliana*	4	11	42	44	5.2	5.6
191	M	GTP-binding protein, ras-like	Q9FJH0	*A. thaliana*	368	7	31	GTP-binding protein, ras-like	Q9FJH0	*A. thaliana*	7	34	24	26	5.7	5.1

NI: non-identified protein.

[a] Spot numbers correspond to the numbers indicated in Fig. 4.

[b] Stage of development in which the corresponding spot was detected.

[c] Accession number in Swiss-Prot/TrEMBL.

[d] Probability based MOWSE score of MASCOT software for the hit.

[e] Number of unique peptide sequences identified via MASCOT.

[f] Percentage of predicted protein sequence covered by matched peptides via MASCOT.

[g] Number of unique peptide sequences identified via MSBLAST that had a significant sequence-alignment score superior or equal to 55.

[h] Percentage of predicted protein sequence covered by matched query sequence via MSBLAST according to [Σpositive queries (aa)×100]/predicted protein (aa).

very successful result considering that *A. angustifolia* genome has not been sequenced and, currently, only sequences of 6 unique proteins are available in NCBI database.

Stringent database searches enable precise protein identification relying upon a few fragmented peptide precursors. However, any peptide sequence discrepancy precludes stringent matching and, consequently, makes protein identification impossible. In *A. angustifolia* seed development, automated de novo sequencing and MS BLAST search protocol has proven its efficiency. This strategy confirmed all *A. angustifolia* cross species MASCOT identifications. Additionally, it enabled the identification of eight proteins (Table 2: spots 6, 31, 65, 140, 161, 163, 164 and 188) that were not identified by MASCOT.

From the 96 identified proteins, 38 appeared as double or triple spots. Multiple vicilin-like protein (acc. number *Q8LKI7*) spots were identified in mature stage gels (Table 2). Similarly high spot redundancy was recently found in other conifer species, such as in *Pinus pinaster* [10] and *P. sitchensis* [11]. Multiple spots for a given protein mostly originate from post-translational modifications, but can also be attributed to differentially spliced forms or allelic variants [11].

3.3. *Functional classification of proteins*

Biological processes, in which the 96 identified proteins are involved, were defined according to the Gene Ontology Annotation database (http://www.ebi.ac.uk/GOA/). Since no functional classification has been defined in conifers, it was based on the comprehensive catalogue proposed by Ruepp et al. [29] and available at http://mips.gsf.de/projects/funcat.

To assess differences in the protein profile during *A. angustifolia* seed development, proteins were classified into nine and eight categories in globular and mature stages, respectively (Fig. 5). The most representative classes of proteins observed in globular and mature stages were protein fate (folding, modification and protein destination), metabolism, energy and storage proteins (Fig. 5).

Protein fate class, which includes proteins related to the folding, assembly and modification processes, constitutes 17 and 21 spots in globular and mature stages, respectively. Within this group, HSPs are the most abundant as 6 spots were identified in both globular and mature stages. Second most representative proteins during seed development were the proteins involved in basic metabolic processes, as they accounted for 26 and 17% of identified proteins in globular and mature stage gels, respectively. Energy protein class constitutes of proteins related to the glycolysis, tricarboxylic acid-pathway and energy conversion and although sixteen spots were detected and classified within this group, it comprises only ten unique identifications. Another group containing high levels of redundant identifications was the storage protein class. They showed major percentage of variation between globular and mature stages, representing the most abundant class of proteins identified in mature stage. In globular tissues, this group represented only 2% of total identified proteins, however in mature stage 19 spots matched its sequence with vicilin-like storage proteins, representing 25% of total identified spots.

3.4. *Protein expression profiling during seed development*

It is well known that a single gene can manifest itself as multiple protein spots on a 2-D gel due to alternative splicing or posttranslational modifications [5]. Although it is possible to detect isoforms in 2-D gels, *A. angustifolia* is a species with unknown genome and protein identification mostly relies on limited sequence coverage. Both of these features hamper revealing the molecular and biological function of different isoforms. Thus, for the analysis of changes in the metabolic processes that take place during seed development, we have only considered changes in protein accessions from all the 96 spots identified.

The 66 unique proteins were clustered within 4 groups (Fig. 6). Cluster A contained 7 unique proteins, most of which highly accumulated throughout seed development, such as actin (Q9SPI7) and the rubisco binding protein (RBP) (Q8L5U4). Cluster B contained 26 proteins showing low accumulation throughout seed development. Cluster C contained 25 proteins overaccumulated in late stages, whereas cluster D contained proteins overaccumulated in early seed development. Aiming a better understanding in changes of the proteome and activation of different metabolic pathways throughout seed development, in the following paragraphs we will discuss some characteristic overaccumulated proteins in early and late stages and their possible function in the process studied. These

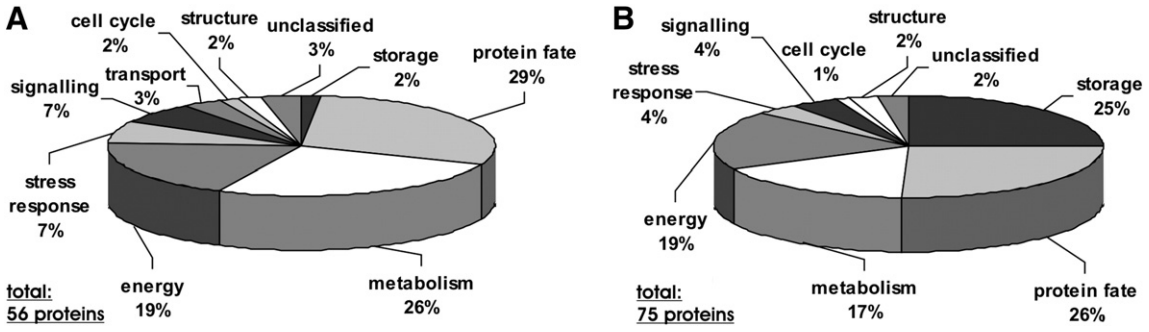

Fig. 5 – Functional distribution of the identified spot during *Araucaria angustifolia* seed development. A: globular stage; B: mature stage.

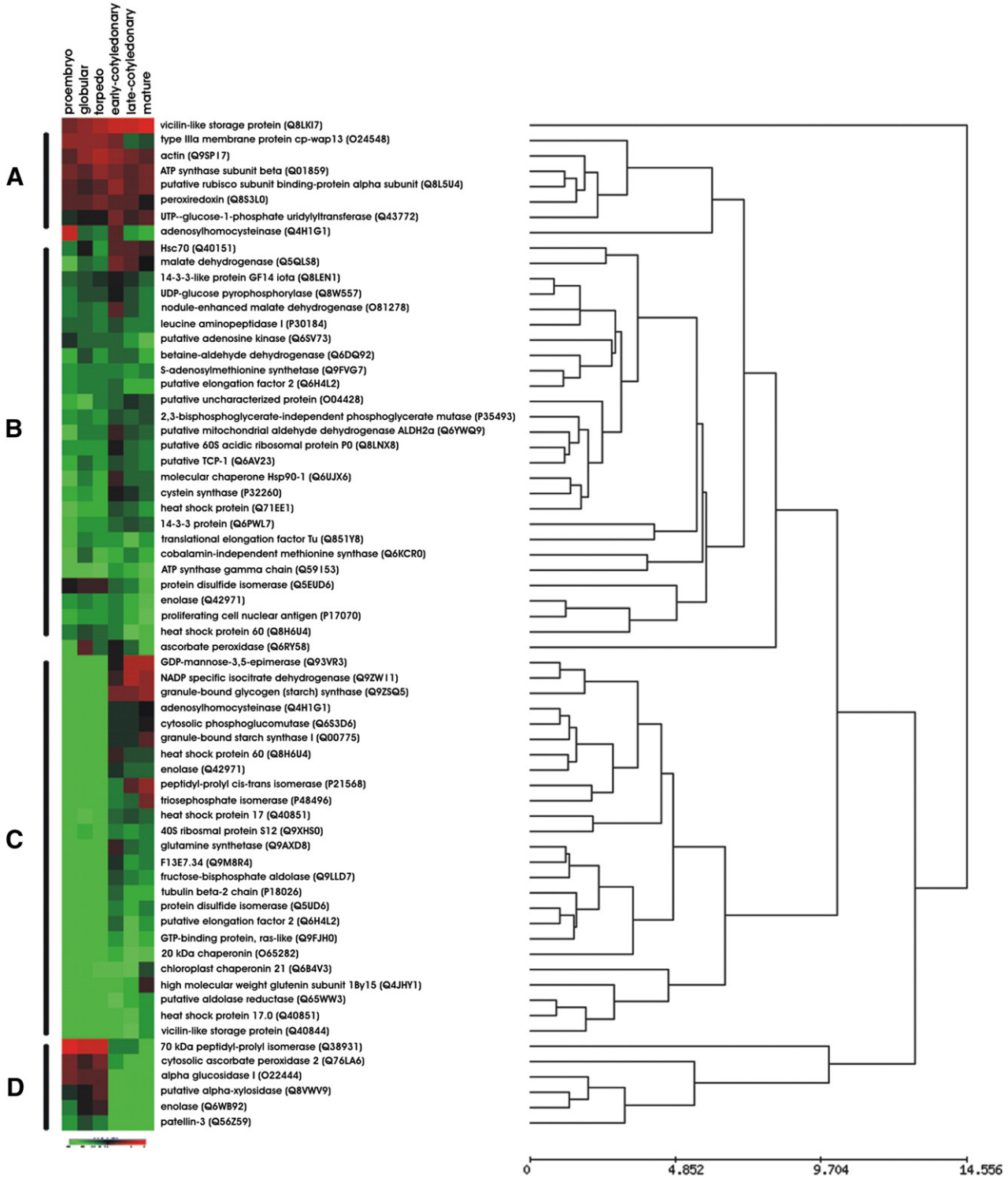

Fig. 6 – Hierarchical clustering analysis of the 66 unique proteins identified during *Araucaria angustifolia* seed development. Dataset clustering was carried out through the sum of all normalized spot volumes from the same protein accession within and between embryo and megagametophyte tissues.

proteins may give insights into specific or preferentially metabolic pathways during the studied process and may also be potential candidates for markers of this process.

3.4.1. *Overaccumulated proteins in early stages*

Many biological events result in cellular oxidative stress, which is mainly associated with accumulation of reactive

oxygen species (ROS) and/or a ROS-scavenging system, like peroxidases. Ascorbate peroxidase (APX) is one of the most important enzymes involved in the oxidative defense system, controlling and modulating hydrogen peroxide (H_2O_2) cellular levels [30]. Although H_2O_2 has a harmful effect on cells at high concentrations; it is an important molecule at the end of seed development and early germination due to its protecting effect against invasion by parasitic organisms and, although not yet described in Gymnosperms, by oxidizing germination inhibitors presented in the pericarp of Angyosperms [31,32]. Few works have been developed relating the ROS production during seed development, specially in recalcitrant seeds. Contrarily, several studies have documented the production of ROS during seed storage in the dry state [33,34]. Although only one APX accession (Q76LA6) was clustered within cluster D, the other showed higher expression in early stages of seed development and, as predicted, was not found in mature stage (Fig. 6). The identification of APXs overaccumulated in proembryo, globular and torpedo stages indicates a higher oxidative stress metabolism in early stages of seed development and the consequent higher accumulation of cytosolic APX for controlling and maintenance of oxidative stress generated by a high ROS activity, which is in agreement with the higher number of identified proteins classified into the stress response class during early seed development (Fig. 5). Over-accumulation of peroxiredoxin in early seed development contributes to this hypothesis (Table 2S), as in dormant seeds and orthodox seeds it shows to play a part in inhibition of germination under unfavorable conditions [35].

In early stages of conifer zygotic embryogenesis, the zygote undergoes several rounds of nuclear duplication without cytokinesis, followed by a cellularization phase, where high rates of cell divisions are then achieved [36]. Plant cells partition their cytoplasm during cytokinesis by building a cell plate structure from the inside out between the two sets of daughter chromosomes, which is dominated by membrane-trafficking events [37]. Patellin (spots 163, 164), which was low expressed and identified especially in globular stage (Fig. 6), is a carrier protein that is recruited from the cytoplasm to the expanding and maturing cell plate. High levels of patellin in globular and torpedo stages also indicate a decrease in the mitotic activity during seed filling and maturation. This is in accordance with the absence of morphological transitions in zygotic embryos from torpedo to mature stage, as observed by Astarita et al. [20], and may also indicate a decrease in cell differentiation and lost of cell competency from torpedo/early cotyledonary stages. To our knowledge, this is the first report on patellin accumulation during seed development.

Proliferating cell nuclear antigen (PCNA) plays an important role in DNA duplication and repair [38]. Despite the recalcitrant behavior of *A. angustifolia* seeds, PCNA levels decreased in late embryogenesis, as observed in maize mature seeds [38]. This suggests the necessity of de novo synthesis of this protein during late germination and early seedling development independently of the recalcitrant nature of seeds.

Biological processes with high rates of cell division should also require the expression of cell wall degradation, loosening and biosynthesis proteins, such as alpha-xylosidase (spot 140) and type IIIa membrane protein cp-wap13 (spots 112, 113 and 114) [39–41], which are essential for growth and tissue expansion in fast growth phases, usually observed in young tissues.

As expected few proteins related to carbon and storage metabolism were detected in early embryogenesis, such as alpha-glucosidase type I (spots 141, 142), which hydrolyzes preferentially heterogeneous substrates, like sucrose [40]. It is interesting to notice that alpha-glucosidase levels decrease during seed development, whereas starch synthase enzyme levels (spots 83, 84) increase in late stages of seed development (Table 2S, Fig. 6), probably indicating that in early stages sucrose is hydrolyzed by alpha-glucosidase I, releasing alpha-D-glucose that is used as substrate for starch synthesis, the major storage compound in *A. angustifolia* mature seeds [42].

3.4.2. *Overaccumulated proteins in late stages*

Chaperonins did not present a preferential accumulation throughout seed developmental stages. However, the 20 kDa accessions (O65282, Q6B4V3) were only identified in late embryogenesis. Chaperonins are proteins that play a vital role in protein folding in eukaryotic and prokariotic cells. They are generally divided into two groups, according to the cellular localization: (a) group I chaperonins are localized in the stroma of chloroplasts, the matrix of mitochondria and eubacteria, and (b) group II chaperonins are found in eukaryotic cytosol and archaebacteria [43]. One of the differences between these groups is that group I work together with co-chaperonins, whereas group II work alone [43].

The chaperonin cpn60 and co-chaperonin cpn10 have been characterized in detail in *Escherichia coli* [44]. An homologue of co-chaperonin cpn 10 was identified in both pea and spinach as being twice the size of the classic mammalian cpn10 and sequence analysis showed that the protein was composed of two linked cpn10 homologs [45,46]. In vitro studies showed that this cpn20 is capable of assisting chloroplast cpn60 in refolding denatured rubisco [45,47]. Interestingly, in *A. angustifolia* seed development, although RBP (Q8L5U4) was detected in early and late stages, the co-chaperonin cpn20 was not identified in early stages (Fig. 6). During *Daucus carota* somatic embryogenesis the expression of photosynthesis-associated mRNAs and differentiation of photosynthetic embryos was observed in late torpedo-shaped embryos [48]. In *A. angustifolia* seed development, cpn20/21 expression was only observed from torpedo stages indicating that photosynthetic apparatus differentiation may start from this stage on (Fig. 6).

In early stages, due to low levels of O_2, mainly converted to superoxide, embryos are ATP limited, responding mainly by fermentation, and during further photosynthetic apparatus differentiation and expression of glycolytic and tricarboxylic acid (TCA) enzymes, ATP levels increase [49]. Although enolase was identified in all stages of seed development and, consequently, clustered within clusters B, C and D, other enzymes involved in the glycolytic and TCA pathway were only identified in late embryogenesis: triosephosphate isomerase, fructose-bisphosphate aldolase and isocitrate dehydrogenase. High expression of these enzymes in late developmental stages indicates a shift in the stress related metabolism, as discussed above, resulting in availability of the molecule O_2 as a final hydrogen acceptor and, thus, increasing respiratory levels in mature seed. D-glucose is the other substrate involved in the respiratory chain and is probably found in abundance during late embryogenesis due to sucrose

hydrolyzation by alpha-glucosidase I activity in early stages. Another evidence of high D-glucose levels in late stages of seed development is the abundant expression of granule-bound starch synthase (Table 2S, Fig. 6), which utilizes D-glucose for starch synthesis.

Final seed development is usually marked by increase in storage compounds and onset of related enzymes, like starch synthase in starchy seeds. Although *A. angustifolia* is one of these seeds, vicilin-like proteins were identified in large amounts in late seed development. Also, multiple spots were correlated to this identification and are probably due to post-translational modifications or different splice forms derived from single gene or the existence of allelic variants [11]. Alternatively, as a storage protein, it could also correspond to fragments released by specific proteolysis even before germination starts, as also observed in Arabidopsis seeds [50] and, thus, seems to be unrelated to desiccation responses.

Although very little is known about N metabolism in embryogenic tissues of conifers, a balanced N supply and metabolism seem to be critical for plant embryogenesis [51]. A major metabolic event during seed maturation is the accumulation of reserves at late embryogenesis. According to previous works, the major amino acid residues of seed storage proteins in *Pinus taeda* and *P. pinaster* are arginine, glutamine and glutamate [51,52]. In the present study the enzyme glutamine synthase (GS) expression was overaccumulated in early-cotyledonary stage (Table 2S, Fig. 6). Changes in GS expression during the maturation process indicate that the biosynthesis and interconversion of glutamine/glutamic acid is active during conifer maturation. Astarita et al. [20] observed that glutamic acid is the major free amino acid accumulated in mature seeds. Moreover, amino acid composition of *A. angustifolia* vicilin storage protein (Q8LKI7) indicates that more than 10% of its sequence is composed by glutamic acid. Thus, the time-specific expression of GS may be critical for obtaining fully mature embryos and may be considered another important protein marker of embryo maturation in this species.

3.5. *Proteomics approach in somatic embryogenesis*

Despite the attempts to develop a protocol for inducing somatic embryogenesis in *A. angustifolia*, only somatic embryos in early developmental stages were obtained [53,54].

Seed development studies have been useful to better understand the molecular and physiological basis of embryogenesis and system manipulation for in vitro multiplication via somatic embryogenesis [18].

Basically, *A. angustifolia* seed development can be divided into two distinct phases according to its protein expression profile: early and late embryogenesis. The metabolic arrangement that takes place during cotyledon formation seems to be crucial as the protein content and 2-DE maps presented a significant change during cotyledon differentiation and growth. In conifers somatic embryogenesis, immature embryos have been extensively used for embryogenic culture initiation, with the developmental explant stage affecting the induction rates [55–57].

In the somatic embryogenesis of *A. angustifolia*, the choice of the correct developmental stage of zygotic embryos is a critical factor in the induction of embryogenic cultures, with the best results when pre-cotyledonary zygotic embryos are used as explants [58,59]. This strict requirement of juvenile explants is common in conifer species and indicates that re-direction of developmental programs in this plant group is difficult to achieve in culture [60]. Taking into account the results obtained in the present work, it can be inferred that the successful conversion rates of embryogenic cultures when using torpedo/early-cotyledonary zygotic embryos may be mostly due to a boom in the gene expression levels and activation of different biochemical pathways, which lead to the detection of a great number of 2-DE spots and marks the transition from a growth and cell division phase to a highly determined and differentiated phase, the maturation.

Proteins involved in energetic, metabolic and protein fate processes are essential across *A. angustifolia* seed development. As predicted, metabolic proteins were the most abundant during early seed development as this stage is marked by intensive cell division and growth. In the *Araucaria* genus, both recalcitrant and orthodox seed species may be found. Due to its elevated moisture content and little desiccation tolerance, *A. angustifolia* is considered as a recalcitrant species [3]. The high number of energy and metabolic related spots in late zygotic embryogenesis of *A. angustifolia* indicates an active metabolic state during the late stages of seed development. Contrary to orthodox, the recalcitrant seeds are shed at high water contents and are also metabolically active at shedding [61]. Although dos Santos et al. [19] observed some similarities of protein electrophoretic pattern with orthodox seeds, the high number of identified metabolic and energy proteins in the maturation phase of seed development is in accordance with its recalcitrant nature and indicates a strategy of continuous development without the interposition of a dry and quiescent state, as suggested by Panza et al. [3]. These observations may be of a great value for the establishment of somatic embryogenesis protocols, as little is known about protein patterns during the maturation phase of recalcitrant coniferous species and most of the somatic embryogenesis protocols, including those for *A. angustifolia* [1,54] are based on the metabolic pattern changes in orthodox type seeds.

Furthermore, in the present work a similar expression profile, especially concerned to the highest abundant proteins, between mature embryos and megagametophyte was observed. This may reflect the mutual conversion of proteins and the signaling events taking place between the embryo and the surrounding tissue. Analysis of late-cotyledonary and mature seeds indicates that storage proteins, which are further used for plantlet growth, are accumulated in both embryo and megagametophyte tissues. Thus, the addition of maturation agents that induce protein expression in the culture medium are expected for successful somatic embryo maturation. Here we also suggest the addition of glutamic acid in maturation culture medium, as amino acid composition of vicilin storage protein indicates a preferential use of this amino acid in its primary structure. Additionally, vicilin-like proteins were far the most abundant identification during embryo maturation, reflecting their importance in late stage of *A. angustifolia* zygotic embryogenesis and potential use as protein markers for embryo maturation, attending on the development of somatic embryo protocols that may utilize

these and the other suggested spot markers for establishing a correct quality control and maximal plant conversion rates from somatic embryos.

4. Concluding remarks

The present work illustrates the robustness of the proteomic approach based on the use of 2-DE gels followed by protein identification via MS/MS analyses for a comprehensive investigation on the metabolic changes during *A. angustifolia* seed development. Further studies addressing temporal and spatial transcription of vicilin genes are being carried out and constitute a potential target for functional genomics studies. Furthermore, specific identified proteins may be used as markers for embryo maturation through in vivo imaging using fluorescent protein fusions. Therefore, the identified protein list presented in this work gives the foundation for future studies on the genetic, physiology and metabolism of the developmental embryo process in this and other coniferous species. Additionally, the coordination of this knowledge may give insight in future studies addressing the optimization of the somatic embryogenesis protocols for mass propagation and conservation strategies applied to *A. angustifolia*.

Acknowledgements

This research was carried out with financial support from the State of São Paulo Research Foundation (FAPESP) and the National Council for Scientific and Technological Development (CNPq) to E.I.S.F. This research was supported by the doctoral and postdoctoral FAPESP scholarships to T.S.B. and V.S., respectively, and by a short-term travel grant by the Deutscher Akademischer Austauschdienst (DAAD) to T.S.B. Work in Shevchenko lab was, in part, supported by 1R01GM070986-01A1 grant from NIH NIGMS. The authors thank Prof. Dr. Miguel P. Guerra and M.Sc. Neusa Steiner for kindly providing *A. angustifolia* seeds.

The authors have declared no conflicts of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jprot.2009.01.011.

REFERENCES

[1] Guerra MP, Silveira V, Santos ALW, Astarita LV, Nodari RO. Somatic embryogenesis in *Araucaria angustifolia* (Bert) O. Kuntze. In: Jain SM, Gupta PK, R.J. Newton, editors. Somatic embryogenesis in woody plants. Dordrecht: Kluwer Academic Publishers; 2000. p. 457–78.

[2] Guerra MP, Silveira V, Reis MS, Schneider L. Exploração, manejo e conservação da Araucária (*Araucaria angustifolia*). In: Simões LL, Lino CF, editors. Sustentável Mata Atlântica: a exploração de seus recursos florestais. São Paulo: Editora SENAC; 2002. p. 85–101.

[3] Panza V, Lainez V, Maroder H, Prego I, Maldonado S. Storage reserves and cellular water in mature seeds of *Araucaria angustifolia*. Bot J Linn Soc 2002;140:273–81.

[4] Salmen–Espindola L, Noin M, Corbineau F, Côme D. Cellular and metabolic damage induced by desiccation in recalcitrant *Araucaria angustifolia* embryos. Seed Sci Res 1994;4:193–201.

[5] Agrawal GK, Hajduch M, Graham K, Thelen JJ. In-depth investigation of the soybean seed-filling proteome and comparision with a parallel study of rapessed. Plant Physiol 2008;148:504–18.

[6] Finnie C, Malchior S, Roepstorff P, Svensson B. Proteome analysis of grain filling and seed maturation in barley. Plant Physiol 2002;129:1308–19.

[7] Gallardo K, le Signor C, Vandekerckhove J, Thompson RD, Burstin J. Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol 2003;133:664–82.

[8] Pawlowski TA. Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: influence of abscisic and giberellic acids. Proteomics 2007;7:2246–57.

[9] Costa P, Pionneau C, Bauw G, Dubos C, Bahrmann N, Kremer A, et al. Separation and characterization of needle and xylem pine proteins. Electrophoresis 1999;20:1098–108.

[10] Gion JM, Lalanne C, le Provost G, Ferry-Dumazet H, Paiva J, Chaumeil P, et al. The proteome of marine pine wood forming tissue. Proteomics 2005;5:3731–51.

[11] Lippert D, Chowrira S, Ralph SG, Zhuang J, Aeschliman D, Ritland C, et al. Conifer defense against insects: proteome analysis of sitka spruce (*Picea sitchensis*) bark induced by mechanical wounding or feeding by white pine weevils (*Pissodes strobi*). Proteomics 2007;7:248–70.

[12] Wang D, Eyles A, Mandich D, Bonello P. Systemic aspects of host–pathogen interactions in Austrian pine (*Pinus nigra*): a proteomics approach. Physiol Mol Plant P 2006;68:149–57.

[13] Fernando DD. Characterization of pollen tube development in *Pinus strobus* (Eastern white pine) through proteomic analysis of differentially expressed proteins. Proteomics 2005;5:4917–26.

[14] Wagner RE, Mugnaini S, Sniezko R, Hardie D, Poulis B, Nepi M, et al. Proteomics evaluation of gymnosperm pollination drop proteins indicates highly conserved and complex biological functions. Sex Plant Reprod 2007;20:181–9.

[15] Dong J-Z, Bock CA, Dunstan DI. Influences of altered phytohormone use on endogenous ABA and mRNA populations during white spruce (*Picea glauca*) somatic embryo culture. Tree Physiol 1997;17:53–7.

[16] Flinn BS, Roberts DR, Webb DT, Sutton BCS. Storage protein changes during zygotic embryogenesis in interior spruce. Tree Physiol 1991;8:71–81.

[17] Sallandrouze A, Faurobert M, el Maataoui M. Characterization of the developmental stages of cypress zygotic embryos by two-dimensional electrophoresis and by cytochemistry. Physiol Plant 2002;114:608–18.

[18] Silveira V, Santa-Catarina C, Balbuena TS, Moraes FMS, Ricart CAO, Sousa MV, et al. Endogenous abscisic acid and protein contents during seed development of *Araucaria angustifolia*. Biol Plantarum 2008;52:101–4.

[19] dos Santos ALW, Wiethölter N, Gueddari NE, Moerschbacher BM. Protein expresssion during seed development in *Araucaria angustifolia*: transient accumulation of class IV chitinases and arabinogalactan proteins. Physiol Plant 2006;127:138–48.

[20] Astarita LV, Floh EIS, Handro W. Free amino acid, protein and water changes associated with seed development in *Araucaria angustifolia*. Biol Plantarum 2003;47:53–9.

[21] Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 1996;68:850–8.
[22] Shevchenko A, Thomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc 2007;1:2856–60.
[23] Junqueira M, Spirin V, Balbuena TS, Waridel P, Surendranath V, Kryukov G, et al. Separating the wheat from the chaff: unbiased filtering of background tandem mass spectra improves protein identification. J Proteome Res 2008;7:3382–95.
[24] Charneau S, Junqueira M, Costa CM, Pires DL, Fernandes ES, Bussacos AC, et al. The saliva proteome of the blood-feeding insect *Triatoma infestans* is rich in platelet-aggregation inhibitors. Int J Mass Spectrom 2007;268:265–76.
[25] Warindel P, Frank A, Thomas H, Surendranath V, Sunyaev S, Pevzner P, et al. Sequence similarity-driven proteomics in organisms with unknown genomes by LC MS/MS and automated de novo sequencing. Proteomics 2007;7:2318–29.
[26] Frank A, Pevzner P. PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal Chem 2005;77:964–73.
[27] Wielsch N, Thomas H, Surendranath V, Waridel P, Frank A, Pevzner P, et al. Rapid validation of protein identifications with the borderline statistical confidence via de novo sequencing and MS BLAST searches. J Proteome Res 2006;5:2448–56.
[28] Junqueira M, Spirin V, Balbuena TS, Thomas H, Adzhubei I, Sunyaev, et al. Protein identification pipeline for the homology-driven proteomics. J Proteomics 2008;71:346–56.
[29] Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, et al. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucl Acids Res 2004;32:5539–45.
[30] Ishikawa T, Shigeoka S. Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. Biosci Biotechnol Biochem 2008;72:1143–54.
[31] Ogawa K, Iwabuchi M. A mechanism for promoting the germination of *Zinnia elegans* seeds by hydrogen peroxide. Plant Cell Physiol 2001;42:286–91.
[32] Schopfer P, Plachy C, Frahry G. 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 Physiol 2001;125:1591–602.
[33] McDonald MB. Seed deterioration: physiology, repair and assessment. Seed Sci Technol 1999;27:177–237.
[34] Pukacka S, Ratajczak E. Production and scavenging of reactive oxygen species in *Fagus sylvatica* seeds during storage at varied temperature and humidity. J Plant Physiol 2005;162:873–85.
[35] Haslekas C, Viken MK, Grini PE, Nygaard V, Nordgard SH, Meza TJ, et al. Seed 1-cysteine peroxiredoxin antioxidants are not involved in dormancy, but contribute to inhibition of germination during stress. Plant Physiol 2003;133:1148–57.
[36] Cairney J, Pullman G. The cellular and molecular biology of conifer embryogenesis. New Phytol 2007;176:511–36.
[37] Bednarek SY, Falbel TG. Membrane trafficking during plant cytokinesis. Traffic 2002;3:621–9.
[38] Herrera I, Sánchez MP, Molina J, Plasencia J, Vázquez-Ramos JM. Proliferating cell nuclear antigen expression in maize seed development and germination: regulation by phytohormones and its association with putative cell cycle proteins. Physiol Plant 2000;110:127–34.
[39] Wald FA, Kissen R, du Jardin P, Moreno S. Characterization of UDP-glucose: protein transglucosylase genes from potato. Plant Mol Biol 2003;52:705–14.
[40] Nakai H, Tanizawa S, Ito T, Kamiya K, Kim YM, Yamamoto T, et al. Function-unknown glycoside hydrolase family 31 proteins, mRNAs of which were expressed in rice ripening and germinating stages, are α-glucosidase and α -xylosidase. J Biochem 2007;142:491–500.
[41] Sánchez M, Gianzo C, Sampedro J, Revilla G, Zarra I. Changes in α -xylosidase during intact and auxin-induced growth of pine hypocotyls. Plant Cell Physiol 2003;44:132–8.
[42] Cordenunsi BR, de Menezes EW, Genovese MI, Colli C, do Souza AGA, Lajolo FM. Chemical composition and glycemic index of Brazilian pine (*Araucaria angustifolia*) seeds. J Agric Food Chem 2004;52:3412–6.
[43] Kuomoto Y, Shimada T, Kondo M, Hara-Nishimura I, Nishimura M. Chloroplasts have a novel cpn10 in addition to cpn20 as co-chaperonins in *Arabidopsis thaliana*. J Biol Chem 2001;276:29688–94.
[44] Bonshtien AL, Weiss C, Vitlin A, Niv A, Lorimer GH, Azem A. Significance of the n-terminal domain for the function of chloroplast cpn20 chaperonin. J Biol Chem 2007;282:4463–9.
[45] Bertsch U, Soll J, Seetharam R, Viitanen PV. Identification, characterization, and DNA-sequence of a functional double groes-like chaperonin from chloroplasts of higher-plants. Proc Natl Acad Sci U S A 1992;89:8696–700.
[46] Dickson R, Weiss C, Howard RJ. Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. J Biol Chem 2000;275:11829–35.
[47] Viitanen PV, Schmidt M, Buchner J, Suzuki T, Vierling E, Dickson R, et al. Functional-characterization of the higher-plant chloroplast chaperonins. J Biol Chem 1995;270:18158–64.
[48] Sato-Nara K, Demura T, Fukuda H. Expression of photosynthesis-related genes and their regulation by light during somatic embryogenesis in *Daucus carota*. Planta 2004;219:23–31.
[49] Rolletschek H, Weber H, Borisjuk L. Energy status and its control on embryogenesis of legumes. Embryo photosynthesis contributes to oxygen supply and is coupled to biosynthetic fluxes. Plant Physiol 2003;132:1196–206.
[50] Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, et al. Proteomic analysis of *Arabidopsis* seed germination and priming. Plant Physiol 2001;126:835–48.
[51] Rodríguez MJP, Suárez MF, Heredia R, Ávila C, Breton D, Trontin JF, et al. Expression patterns of two glutamine synthetase genes in zygotic and somatic pine embryos support specific roles in nitrogen metabolism during embryogenesis. New Phytol 2006;169:35–44.
[52] Allona I, Casado R, Aragoncillo C. Seed storage proteins from *Pinus pinaster* Ait.: homology of major components with 11S proteins from angiosperms. Plant Sci 1992;87:9–18.
[53] Silveira V, Santa-Catarina C, Tun NN, Scherer GFE, Handro W, Guerra MP, et al. Polyamine effects on the endogenous polyamine contents, nitric oxide release, growth and differentiation of embryogenic suspension cultures of *Araucaria angustifolia* (Bert.) O. Ktze. Plant Sci 2006;171:91–8.
[54] Steiner N, Santa-Catarina C, Silveira V, Floh EIS, Guerra MP. Polyamine effects on growth and endogenous hormones levels in *Araucaria angustifolia* embryogenic cultures. Plant Cell Tissue Organ Cult 2007;99:52–62.
[55] Park YS, Pond SE, Bonga JM. Initiation of somatic embryogenesis in white spruce (*Picea glauca*): genetic control, culture treatment effects, and implications for tree breeding. Theor Appl Genet 1993;86:427–36.
[56] Salajova T, Salaj J, Kormutak A. Initiation of embryogenic tissues and plantlet regeneration from somatic embryos of *Pinus nigra* Arn. Plant Sci 1999;145:33–40.
[57] Li XY, Huang FH. Induction of somatic embryogenesis in loblolly pine (*Pinus taeda* L.). In Vitro Cell Dev Biol Plant 1996;32:129–35.

[58] dos Santos ALW, Silveira V, Steiner N, Vidor M, Guerra MP. Somatic embryogenesis in Parana Pine (*Araucaria angustifolia* (Bert.) O. Kuntze). Braz Arch Biol Technol 2002;45:97–106.
[59] Silveira V, Steiner N, Santos ALW, Nodari RO, Guerra MP. Biotechnology tolls in *Araucaria angustifolia* conservation and improvement: inductive factors affecting somatic embryogenesis. Crop Breed Appl Biotechnol 2002;2:463–70.
[60] Stasolla C, Yeung EC. Recent advances in conifer somatic embryogenesis: improving somatic embryo quality. Plant Cell Tissue Organ Cult 2003;74:15–35.
[61] Pammenter NW, Berjak P. Evolutionary and ecological aspects of recalcitrant seed biology. Seed Sci Res 2000;10:301–6.