Journal of Proteomics 130 (2016) 180–189

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Quantitative proteomic analysis of *Araucaria angustifolia* (Bertol.) Kuntze cell lines with contrasting embryogenic potential

André Luis Wendt dos Santos [a,1], Paula Elbl [a,1], Bruno Viana Navarro [a], Leandro Francisco de Oliveira [a], Fernanda Salvato [b], Tiago Santana Balbuena [c], Eny Iochevet Segal Floh [a,*]

[a] Department of Botany, Institute of Bioscience, University of São Paulo, São Paulo, São Paulo, Brazil
[b] Department of Plant Biology, Institute of Biology, State University of Campinas, Campinas, São Paulo, Brazil
[c] College of Agricultural and Veterinary Sciences, São Paulo State University, Jaboticabal, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 6 May 2015
Received in revised form 26 August 2015
Accepted 22 September 2015
Available online 3 October 2015

Keywords:
Quantitative proteomics
Somatic embryogenesis
Conifer
Embryogenic potential
Molecular markers

ABSTRACT

GeLC–MS/MS based label free proteomic profiling was used in the large scale identification and quantification of proteins from Brazilian pine (*Araucaria angustifolia*) embryogenic cell (EC) lines that showed different propensities to form somatic embryos. Using a predicted protein sequence database that was derived from *A. angustifolia* RNA-Seq data, 2398 non-redundant proteins were identified. The $\log_2$ of the spectral count values of 858 proteins of these proteins showed a normal distribution, and were used for statistical analysis. Statistical tests indicated that 106 proteins were significantly differentially abundant between the two EC lines, and that 35 were more abundant in the responsive genotype (EC line SE1) and 71 were more abundant in the blocked genotype (EC line SE6). An increase in the abundance of proteins related to cell defense, anti-oxidative stress responses, and storage reserve deposition was observed in SE1. Moreover, in SE6 we observed an increased abundance of two proteins associated with seed development during the embryogenic cell proliferation stage, which we suggest is associated with genotypes showing a low responsiveness to embryo formation. Differences in protein abundance between the EC lines are discussed in terms of carbohydrate metabolism, cell division, defense response, gene expression, and response to reactive oxygen species.

1. Introduction

Asexual plant propagation *via* somatic embryogenesis (SE) has proven to be effective in large-scale clonal propagation, as well as for the development of stable transgenic varieties [1]. SE has also been used as an experimental system to better understand totipotency in later diverging land plants [2] and in association with cryopreservation can provide biotechnological tools to the *ex situ* conservation of endangered plant species [3,4]. Although SE protocols have been developed for a number of species [1,2,4], the limitations imposed by suboptimal conditions during *in vitro* growth of somatic embryos have restricted its use in others [5]. The establishment of a successful SE protocol requires the development of molecular markers that can be used to select cell lines with highly embryogenic potential [6,7]. Conservation strategies based on SE associated with cryopreservation are being developed for *Araucaria angustifolia*, an endangered conifer that is native to Brazil [7,8]. SE induction in *A. angustifolia* was first described in Ref. [9], and since then several reports have outlined attempts to develop a SE protocol for this species, but to date successful plant regeneration has not been possible [6,7,10].

One approach to generating such markers and to better understanding the regulation of *in vitro* embryogenesis is to use global transcript expression or protein profiling [11,12]. Indeed, proteome profiling has already been used to study somatic embryo formation and development in several plant species [2,12,13], including *A. angustifolia* [7,8]. Moreover, proteome analysis represents a promising approach for biomarker discovery, since it is now possible to identify and quantify proteins on a large scale [14,15]. However, protein identification is generally dependent on the existence of high quality DNA sequence databases, derived from genome sequences or expressed sequence tags (ESTs), which are by definition not available for non-model species [16]. However, next-generation sequencing technologies and their applications, such as RNA-sequencing (RNA-Seq), are substantially expanding the number of plant species with high coverage sequence databases, thereby improving the potential for proteomic analysis of species for which a genome sequence is not available [17]. This has the potential to pave the way for studies of the molecular events associated with embryogenesis [18]. Recently, Elbl *et al.* [10] developed an extensive *A. angustifolia* sequence dataset by *de novo* sequencing of the transcriptomes of *A. angustifolia* seeds at different stages of development, as well as embryogenic cell lines with different embryogenic potential: cell line SE1,

* Corresponding author at: Department of Botany, Institute of Bioscience, University of São Paulo, CEP 05508-090 São Paulo, São Paulo, Brazil.
E-mail address: enyfloh@usp.br (E.I.S. Floh).
[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jprot.2015.09.027
1874-3919/

which is responsive to somatic embryo development, and cell line SE6 in which somatic embryo development is blocked. This analysis provided insights into putative genetic determinants that contribute to the embryogenesis potential of cell lines, somatic embryo initiation, and differences in gene expression between the initial stages of somatic and zygotic embryogenesis.

In the current study, an *A. angustifolia* protein database, derived from RNA-Seq data, was generated in order to conduct a large scale proteomic analysis of the SE1 and SE6 embryogenic cell lines. Label-free proteome analysis, involving gel electrophoresis and liquid chromatography–tandem mass spectrometry (GeLC–MS/MS), was used to identify and quantify proteins extracted from the *A. angustifolia* cell lines. Proteins that were detected in both cell lines were annotated by Gene Ontology (GO) analysis and their annotations indicated a relationship with the GO categories 'gene expression', 'carbohydrate metabolism', 'response to reactive oxygen species', 'cell division' and 'defense response'. In addition to the potential for improving SE protocols by revealing molecular markers for early detection of cell lines with high embryogenic capacity, this study further advances the current understanding of molecular regulation during conifer embryogenesis.

2. Experimental procedures

2.1. Plant material

Two *A. angustifolia* embryogenic cell (EC) lines were used, SE1 and SE6. These cell lines (induced from zygotic embryos of different mother trees) were previously selected as described by Jo *et al.* [7], based on different responses under maturation conditions. Initially, these two cell lines were referred to as R and B cell lines [7]; however, during the transcriptome analysis of the R and B cell lines [10] they were renamed SE1 and SE6, respectively. The SE1 cell culture produces pre-cotyledonary embryos (85 ± 19 [Mean ± S.D.] embryos produced per gram fresh weight of embryonal mass) (de Oliveira LF and Floh EIS, unpublished data) in medium supplemented with osmotic agents and abscisic acid (maturation medium), which reach the early cotyledonary stage, while further development of the SE6 cell line is blocked and somatic embryos do not develop [7,10]. For proteomic analyses, the SE1 and SE6 cultures were grown for 21 days on MSG proliferation medium [10] in the dark, at 25 °C ± 1 before harvesting.

2.2. Protein database

Protein sequence database was derived from the *A. angustifolia* transcriptome datasets [10] using Transdecoder software (http://transdecoder.sourceforge.net/) from the Trinity suite, with default parameters. Partial (derived from internal, 5′ and 3′ cDNA sequences) and complete open reading frames (ORFs) were identified. The ORF annotations were assigned based on the previously annotated *Araucaria* transcriptome [10] and were coupled to the domain identification by analysis with HMMER using Pfam-B non-model species [19] in order to further characterize the predicted proteome. An overview of the *Araucaria* protein sequence database is provided in Supplemental material 1.

2.3. Protein extraction and gel electrophoresis

Proteins were extracted from EC lines frozen in liquid nitrogen, and ground to a fine powder in a pre-chilled mortar. Aliquots of 0.5 g of powder were further ground in extraction buffer (proportion 1:3 powder:buffer, *w*/*v*) comprising 50 mM Tris–HCL (pH 8.0), 2% (*w*/*v*) SDS, 1 mM phenylmethanesulfonylfluoride, 1 mM ethylenediaminetetraacetic acid and 50 mM DL-Dithiothreitol After centrifugation at 10,000 × *g* for 25 min at 25 °C, the supernatants were collected and 200 μL aliquots stored −20 °C. The protein concentrations in the supernatants were determined using the 2D Quant kit (GE Healthcare Life Sciences), following the manufacturer's instructions. Protein extracts were prepared from four biological replicates for each EC line. Prior to SDS-PAGE [20], sample aliquots containing 10 μg of proteins were mixed with an equal volume of loading buffer (0.5 M Tris [pH 6.8], 20% [*v*/*v*] glycerol, 2% [*w*/*v*] SDS, 5% [*v*/*v*] 2-mercaptoethanol and traces of bromophenol blue) and incubated for 5 min at 99 °C. Gel electrophoresis was performed using 1 mm thick 12% (*w*/*v*) acrylamide slab gels for 100 min using 10 mA constant current. After electrophoresis, gels were fixed for 1 h in a solution containing 40% (*v*/*v*) ethanol and 10% (*v*/*v*) acetic acid and stained overnight with colloidal Coomassie blue, as previously described [21].

2.4. In-gel trypsin digestion

Prior to protein digestion, the gel lane corresponding to each biological replicate was excised from the gel and divided into five equally sized segments. These were further cut into approximately 1 mm cubes with a clean scalpel, which were then transferred into 1.5 mL sterile polypropylene tubes for gel trypsin digestion, as described by Shevchenko *et al.* [22]. Protein digestion was performed by the addition of 200 μL (5 ng/μL) of sequencing grade porcine trypsin (Promega, Madison, WI) in digestion solution (100 mM ammonium bicarbonate/10% acetonitrile [ACN]). Samples were digested overnight at 37 °C and gel pieces were then saturated with 400 μL of extraction buffer (5% formic acid [FA]/ACN [1:2, *v*/*v*]) and incubated for 30 min at 37 °C. Supernatants were collected and then dried in a vacuum centrifuge and stored at −80 °C until used for LC–MS/MS analyses.

2.5. LC–MS/MS analyses

The trypsinized protein samples were dissolved in 0.1% FA and subjected to online nanoflow liquid chromatography tandem mass spectrometry (LC–MS/MS) using a Q-Exactive quadrupole Orbitrap mass spectrometer (Thermo Scientific, Germany). The peptides were fractionated using the EASY-nLC 1000 system (Thermo Scientific, Denmark) and a reverse phase analytical column (10 cm, 75 μm ID, 120 Å, C18-AQ, Thermo Scientific). The flow rate was set to 200 nL/min, and peptides were separated with a linear ACN gradient from 5 to 35% for 45 min followed by a linear increase to 80% ACN over a period of 5 min. After 5 min at 80% ACN, column re-equilibration was achieved by a reduction to 5% ACN for 3 min. Nanospray peptide ionization was carried out using a voltage equal to 2 kV. The MS/MS scans were acquired in data dependent mode. Survey scans (*m*/*z* 400–2000) were acquired with the resolution adjusted to 70,000 FWHM (full width at half maximum). The 10 most intense ions with a charge state ≥2 were isolated and fragmented in the collision cell by higher-energy collisional dissociation (HCD) at 35 eV. Analyses of MS/MS scans were performed with FWHM set at 35,000. The dynamic exclusion parameter was set to 60 s.

2.6. Identification and quantification of proteins

Raw MS data were processed using MaxQuant 1.4.1.2 software [23], supported by the Andromeda search engine [24] for peptide identification. MS/MS spectra were searched using Andromeda against a concatenated database consisting of the 48,246 EST sequences in the *A. angustifolia* transcriptome database [10] in both forward and reverse orientations, supplemented with frequently observed contaminants. MS/MS peak lists were filtered to leave the 10 peaks with highest abundance per 100 Da intervals with mass tolerances of 10 ppm for the precursor ions and 0.02 Da for the fragment ions. Cysteine carbamidomethylation (Cys + 57.021464 Da) was used as a fixed modification and oxidized methionine (+15.994915 Da) as a variable modification. Up to two missing cleavages were allowed and a minimum of two peptides were required for protein identification. Peptide matches were assembled into protein groups using the Occam's razor principle,