

Seed protein profile of *Pinus greggii* and *Pinus patula* through functional genomics analysis

Perfiles proteómicos de *Pinus greggii* y *Pinus patula* a través de análisis de genómica funcional

Orlis B Alfonso ^a, David Ariza-Mateos ^{b*}, Guillermo Palacios-Rodríguez ^b, Alexandre Ginhás Manuel ^a, Francisco J Ruiz-Gómez ^b

^aJosé Eduardo dos Santos University, Agricultural Sciences Faculty, Department of Forest Engineering, Huambo, Angola.

*Corresponding author: ^bUniversity of Cordoba, Department of Forest Engineering, Laboratory of Dendrochronology, Silviculture and Global Change, DendrodatLab-ERSAF, Campus de Rabanales, Ctra. N. IV, 14071 Córdoba, Spain; Tel.: +34 957 218381, dariza@uco.es

SUMMARY

The present work was carried out with the aim of analyzing and describing the seed proteome of *Pinus patula* and *Pinus greggii*. The analysis was performed using the “shotgun” (“gel-free”) strategy. Proteins were extracted using the TCA/Phenol/Acetone protocol, subsequently separated by liquid chromatography and analyzed by mass spectrometry (nLC LTQ Orbitrap). Protein identification was performed by consulting the specific database for *Pinus* spp and functional classification taking into account the three functional terms (biological processes, cellular components and molecular functions) of Gene Ontology. To extract relevant Gene Ontology terms, a singular enrichment analysis (SEA) was performed, the terms were considered relevant for a minimum threshold of significance FDR < 0.05. After analyzing protein profiles, a total of 1091 proteins were identified, 362 proteins common in both species, 100 exclusives to *P. greggii* and 267 exclusive proteins to *P. patula*. The comparative analysis of the distribution of proteins as a function of the three functional terms reveals similarity between the two species. The most abundant proteins were associated with oxidation-reduction processes, in terms of cellular components; in both species integral membrane proteins predominate. Regarding molecular function, the predominant expression of proteins in both species was related to ATP synthesis. The metabolic pathway with the highest participation of proteins (17 in *P. greggii* and 19 in *P. patula*) was the glycolysis/gluconeogenesis pathway. These results demonstrate that there is little genetic variability between the two species, with proteins associated with metabolic processes prior to germination prevailing.

Key words: coniferous, genetic variability, molecular markers, tandem mass spectrometry.

RESUMEN

El presente trabajo se llevó a cabo con el objetivo de analizar, describir y comparar el proteoma de la semilla de *Pinus patula* y *Pinus greggii*. El análisis se realizó utilizando la estrategia “shotgun”. Las proteínas se extrajeron mediante el protocolo del ácido tricloroacético / Fenol / Acetona, se separaron por cromatografía líquida y se analizaron por espectrometría de masas. La identificación de proteínas se realizó consultando la base de datos específica para *Pinus* spp. y la clasificación funcional, teniendo en cuenta los tres términos funcionales (procesos biológicos, componentes celulares y funciones moleculares) de la Gene Ontology. Para extraer los términos relevantes de la Gene Ontology se realizó un análisis de enriquecimiento singular. Se identificaron 1091 proteínas, 362 proteínas comunes a ambas especies, 100 exclusivas de *Pinus greggii* y 267 exclusivas de *Pinus patula*. El análisis comparativo de la distribución de proteínas en función de los tres términos funcionales reveló la similitud entre las dos especies. Las proteínas más abundantes se asociaron a procesos de oxidación-reducción. En cuanto a los componentes celulares, en ambas especies predominan las proteínas integrales de membrana. En cuanto a la función molecular, la expresión mayoritaria de las proteínas en ambas especies estaba relacionada con la síntesis de ATP. La vía metabólica con mayor participación de proteínas fue la vía de la glicólisis/gluconeogénesis. Estos resultados demuestran que hay poca variabilidad genética entre las dos especies, debido a la similitud de sus perfiles proteicos, prevaleciendo las proteínas asociadas a los procesos metabólicos previos a la germinación.

Palabras clave: coníferas, variabilidad genética, marcadores moleculares, espectroscopia de masa tándem.

INTRODUCTION

Pinus spp. is considered an important pillar in forestry, not only for the number of species but also for their economic importance as endemic or introduced species, both in natural ecosystems and in forest plantations (Muniz *et al.* 2020). In the 1960s and 1970s, several species of *Pinus* were introduced into Angola, mostly from Mexico. *Pinus patula* Schl and Cham. and *Pinus greggii* Engelm. are two species of Mexican *Pinus* spp. modified in afforestation and restoration programs throughout the world, mainly because they are fast growing species and resistant to adverse environmental conditions, such as drought, shallow, eroded and poor in nutrients soils (Casique *et al.* 2019). Molecular phylogeny studies confirm that these species are genetically similar, however, despite their similarity, the *P. greggii* species differs from *P. patula*, because it has short, erect leaves, with superior thickness and epidermal thickness, while *P. patula* has a higher epidermal cell width in relation to *P. greggii* (Singh *et al.* 2019). Plant biotechnology has led to the emergence of new research areas in forestry in the last two decades. Advances in molecular biology resulted in the common use of methods for identification of species based on the use of molecular markers.

These methods have been proposed as valid tools for evaluating the productivity, growth and response to biotic and abiotic stresses in different forest species and are applicable to any type of vegetative material; therefore, permitting the correct identification of the species with a reasonably low effort (Rodríguez de Francisco *et al.* 2016).

When applied to molecular and genetic studies in forestry, the new techniques of molecular biology and -omic approaches allow us to establish the relationship between individual trees and progenies; and also, to look deeper into the population structure and biological diversity. Together with traditional biochemistry and genetics, these techniques (DNA molecular markers, transcriptomics, proteomics and metabolomics) have been used to characterize genetic variability in proteome of seeds of other *Pinus* species in *P. pinea* L. (Loewe *et al.* 2018), *P. occidentalis* (Rodríguez de Francisco *et al.* 2016) and *P. halepensis* (Taïbi *et al.* 2017). Molecular studies are effective tools to genetically characterize species, particularly considering the most likely effects of climate change on forests. The proteins profile analysis of different sample tissues, for example seeds, is very useful because it helps to detect the variability in type and protein concentrations and to identify proteins as putative phenotype markers, differentiating populations or elite phenotypes (Rey *et al.* 2019).

This study will allow i) to characterize, through molecular markers (proteins) of the seeds of these two *Pinus* species (*P. patula* and *P. greggii*) ii) to classify and compare the main biological categories (biological processes, cellular components and molecular function) and their metabolic profiles in which they are involved and iii) to compare the genetic variability between seeds of both species

for the selection of the appropriate forest material in the conservation and management programs for exotic species in Angola and, even more, in the current context of change, the projections of which are particularly more serious on the African continent.

METHODS

Plant material. The plant material used for the analysis were homogeneous seed lots of *Pinus patula* and *Pinus greggii* species, harvested in 2016. The certified seeds of both species came from controlled progenitors and were provided by the Seed Bank of the Mexico Forestry Genetic Center (CNRGM) (table S1, Supplementary Material).

Protein extraction. Protein extraction was done using the TCA/Phenol/Acetone methodology described by Wang *et al.* (2006). First, *Pinus patula* and *Pinus greggii* seeds were extracted from pine nuts and disinfected by submerging them in a 5 % Sodium Hypochlorite solution (NaClO) followed by rinsing them with regular tap water. Afterwards, the seeds were grinded in a mortar with the help of liquid nitrogen until a fine dust was obtained. Six dust samples were weighed (0.2 g) that represented the three biological replicates used for each species. Decanted proteins were re-suspended in a solubilization medium of urea 7M, thiourea 2M, CHAPS at 4 % per volume, Triton X-100 0.5 % (v/v) and DTT 100 nM. The resulting solution was centrifuged for 10 minutes at 1200 rpm. Finally, proteins were quantified by the Bradford method using bovine serum albumin (BSA) as standard (Maldonado *et al.* 2008). Extracted proteins were kept at -80 °C until the shotgun proteomic analysis was carried out.

Identification of proteins by nLC-MS MS.

Sample preparation: The shotgun gel-free protein analysis was carried out in the Proteomics Facility at the Research Support Central Service (SCAI) University of Cordoba (Spain). Proteins extracts were cleaned-up in 1D SDS-PAGE at 10 % polyacrylamide. Samples were loaded onto the stacking gel and a stable electrical potential of 100 V was applied until the electrophoresis front reached the resolving gel, coinciding when the protein extract had entered 1 cm into the resolving gel. The gel was stained with Commasie Blue and protein bands were cut off, diced, and kept in water until digestion.

Protein digestion: Briefly, gel dices were destained in 200 mM ammonium bicarbonate (AB)/50 % acetonitrile for 15 min followed by 5 min in 100 % Acetonitrile. Proteins were reduced by adding 20 mM dithiothreitol in 25 mM AB and incubated for 20 min at 55 °C. The mixture was cooled down to room temperature, followed by alkylation of free thiols by adding 40 mM iodoacetamide in 25 mM AB, and placed in the dark for 20 min. The gel pieces were

then washed twice in 25 mM AB. Proteolytic digestion was performed by adding Trypsin (Promega Madison, WI) at 12.5 ng/μl of enzyme in 25 mM AB and incubated at 37 °C overnight. Protein digestion was stopped by adding trifluoroacetic acid at 1 % final concentration and the digested samples were finally dried in Speedvac (Eppendorf Vacufuge Concentrator, Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA).

nLC-MS2 analysis: The peptide mixture resulting from the digestion was cleaned to remove the salts in an Acclaim Pepmap precolumn of 300 μm x 5 mm (Thermo Fisher Scientific Inc.) with 2 % (ACN) / 0.05 % TFA for 5 min. at 5 μl/min. The separation of nLC peptides was performed with the Dionex Ultimate 3000 nano UPLC (Thermo Fisher Scientific Inc.) with a C18 analytical column 75 μm x 50 mm Acclaim Pepmap precolumn (Thermo Fisher Scientific Inc.) at 40 °C for all analyses. Mobile phase buffer A was composed of water, 0.1 % formic acid. Mobile phase B was composed of 20 % acetonitrile, 0.1 % formic acid. In the elution, the following gradients of 4-35 % of phase B were used for 60 minutes, 35-55 % of B for 3 minutes, 55-90 % of B for 3 minutes followed by washing 8 minutes to 90 % of B and a rebalancing of 12 minutes at 4 % of B. The total chromatography time was 85 min.

Peptides elution results were converted into gas-phase ions by Nano electrospray ionization and analyzed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo Fisher Scientific, Inc.) mass spectrometer operated in positive mode. Survey scans of peptide precursors from 400 to 1500 m/z were performed at 120K resolution (at 200 m/z) with a 4 × 10⁵ ion count target. Tandem MS was performed by isolation at 1.2 Da with the quadrupole, CID fragmentation with normalized collision energy of 35, and a rapid scan MS analysis in the ion trap. The automated gain control (AGC) ion count target was set to 2 × 10³ and the max injection time was 300 ms; only those precursors with charge between 2–5 were sampled for MS². The dynamic exclusion duration was set to 15 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top 30 mode with 3 s cycles, meaning the instrument would continuously perform MS² events until a maximum of top 30 non-excluded precursors or 3 s, whichever was shorter.

Finally, the spectra obtained from the mass analysis were analyzed with the software Proteome Discoverer v. 2.1.0.81 (Thermo Fisher Scientific Inc.) and the SEQUEST/HT search engine using a pine protein database inferred from the existing transcript (35,374) collection in *Pinus*, SustainPineDB v3.0 available on the Sustain Pine Project Platform PLE2009-0016 (Fernández *et al.* 2011). Protein sequences were extracted from the transcripts once the correct reading frame was identified after using blastx against Acrogymnospermae (ancestor). For those transcripts that did not present homology, the highest ORF (Open Reading Frame) was selected from all those that

could be obtained by theoretically translating into the six possible reading frames.

Peptides were generated by theoretical tryptic digestion. Carbamidomethylation of cysteine was considered as a fixed modification, and variable modification was methionine oxidation, error tolerance for precursor ions of 10 ppm and product (fragmented) ions of 0.1 Da. Peptide spectral matches (PSMs) were validated with the Percolator algorithm (included in the Proteome Discoverer software), using a filter based on values of 1 % of the false discovery rate FDR (False Discovery Rate).

Protein functional classification. The proteins identified by the nLC LTQ Orbitrap analysis were functionally annotated using the Blast2GO software (v.4.1.9) against the nr-NCBI database, conserving up to 20 blast “hits” with an e-value threshold of 1e⁻⁵. Likewise, metabolic domains and pathways were annotated by exploring the IPRScan (protein domains identifier) and the Kyoto Encyclopedia of Genes and Genomes database (KEGGs). The Blast Description Annotator algorithm was used to assign a description to each sequence. For those sequences that remained without any information in this phase (10 sequences), the threshold was relaxed to an e-value 1e⁻³ and tried to find similarities against nt-NCBI, to assign a description to these sequences; although this result was not used to annotate them functionally due to their low confidence. The classification for the functional profiles was prepared considering the terms of Gene Ontology, taking into account the three present categories: biological processes, cellular components and molecular functions.

Statistical analysis. To classify the main functional processes in which the most abundant proteins and their metabolic associated profiles for each species may be involved, a functional enrichment analysis (SEA) was used to determine the most relevant Gene Ontology terms associated with the proteins identified for each species. This analysis consists in transforming the two lists of proteins to be used in comparing their Gene Ontology annotations, using the corresponding protein-functional term association table. Subsequently, a Fisher’s exact test with a 2 x 2 contingency table was used to verify if there are statistically significant overrepresented Gene Ontology annotations in any of the two species. A multiple testing correction was applied to each term to correct the departure from the simultaneously multiple hypothesis tested. The significance of terms was determined by using statistical tests with a minimum significance level of FDR < 0.05.

RESULTS

Protein profile analysis. A total of 1091 proteins were identified: 629 in seeds of *P. patula* and 462 in *P. greggii*. Once proteins profiles were compared, 362 common protein species were identified in both *Pinus* species. A total

of 100 and 267 unique proteins were identified for *P. greggii*, and *P. patula* respectively, with 362 proteins present in both species. Most identifications were matched by homology for *Picea sitchensis* annotations ([figure S1](#) and [table S2](#), Supplementary Material).

Classification of functional profiles according to Gene Ontology terms for biological processes, cellular components and molecular functions. The comparative analysis between species of proteins distribution according to the biological process revealed similarities between the two species ([figure 1](#) and [table S2](#) Supplementary Material), with higher predominance of proteins involved in oxidation-reduction processes (78 and 104 proteins in *P. greggii* and *P. patula*, respectively), followed by the processes of monocarboxylic acid metabolism, translation, ribosomal biogenesis, and catabolism of organic substances.

When proteins were classified regarding the different cellular components, integral or intrinsic membrane proteins predominated in both species (33 and 40 in *P. greggii* and *P. patula*, respectively) ([figure 2](#)). Most identified proteins belong to components of the membrane, including transport proteins, plasma membrane proteins and catalytic proteins of the cell wall and other inner organelles. Other important proteins were the cytosolic ribosomal units and proteins involved in nuclear activity, such as gene expression. When proteins were grouped by their molecular function, ([figure 3](#)), ATP binding, with 48 and 62 proteins in *P. greggii* and *P. patula* respectively, was the function with the larger number of proteins identified, followed by the structural constitution of the ribosome with 48 and 59 proteins respectively. In general, the results of the analysis of the functional profiles indicate that no enrichment was detected in any Gene Ontology term considering $FDR < 0.05$; since most of these terms (biological processes, cellular components and molecular function) had a similar representation in the two species.

Functional classification of proteins by exploring the KEGGs database. The results of the functional classification of proteins in the KEGGs database showed that the proteins identified in the study participate in a total of 99 metabolic pathways between the two species, 90 common on both species and 9 exclusives for *P. patula* ([table S2](#) Supplementary Material). Those exclusive pathways were correspondent to the biotin metabolism, lipid metabolism, flavonoid biosynthesis, metabolism of glycerophospholipids, linoleic acid metabolism, phosphatidylinositol signaling system, porphyrin and chlorophyll metabolism, riboflavin metabolism and streptomycin biosynthesis.

The common metabolic pathways accounting for the high number of proteins among the identified ones were the same for both species: glycolysis/gluconeogenesis ([table 1](#)), antibiotic biosynthesis, fatty acid biosynthesis, lipid metabolism, pyruvate metabolism, nucleotide metabolism, carbon fixation in photosynthetic organisms and oxidative phosphorylation.

Regarding the glycolysis/gluconeogenesis pathway, 19 of the identified proteins were related to enzymes involved in this metabolic pathway ([figure 4](#)). Most of these proteins (16) are common in both species ([table 1](#)), although three were only identified in one of the species: acetyltransferase, identified only in *P. greggii*, and Diphosphate-fructose-6-phosphate 1-phosphotransferase (PF6) and Phosphoenolpyruvate carboxykinase ATP (PEPCK), only identified in *P. patula*.

DISCUSSION

Identified proteins and data analysis. In this work, the analysis of the protein profile was based on the shotgun proteomic approach, considered by many researchers as an effective tool for high performance proteomic studies (Woo *et al.* 2019). However, due to the sensitivity of peptide identification in this approach, authors such as Romero *et al.* (2014) recommend the use of specific proteins from similar organisms to improve the sensitivity of peptide identification when working with non-model species. In this case, a database of sequences specific to the genus *Pinus*, SustainPineDB v3.0 (Fernández *et al.* 2011), was used, generated from gene predictions of the latest assembly version of *Pinus pinaster* specific genome.

The species that contributed the most to protein annotation was *Picea sitchensis*, a genomically highly studied species and for which there is a large amount of information available in databases. Authors such as De la Torre *et al.* (2019) state that there are low levels of genetic variation between these genera. However, previous genetic studies by Kovačević *et al.* (2013) concluded that the *Picea* genus is clearly separated from the *Pinus* genus, being closer to the *Abies* genus.

Gene ontology and protein expression of P. patula and P. greggii seeds. The oxidation-reduction processes accounted for the highest number of proteins identified in the two *Pinus* species analyzed. Similar results were obtained by other authors studying the proteomic profile of forest species seeds through the “shotgun” strategy. Pullman *et al.* (2015) have shown that oxidation-reduction processes can improve embryogenic tissue initiation (ET) in conifers, early embryonic development and germination in these species. On the other hand, most of the proteins were located in or related to the cell membrane for both *P. patula* and *P. greggii*. According to Rosental *et al.* (2014) the cell membrane is directly involved in the process of germination, during which it experiences a specific and functional reorganization by actively participating in the imbibition of water by the seed.

Regarding the expression, the proteins overexpressed for the three Gene Ontology were those involved in storage, carbohydrate metabolism and energy synthesis, processes associated with seed preparation for germination. These results agree with Morel *et al.* (2014) reported in *Pinus pinaster* seeds. It is considered, also, that proteins

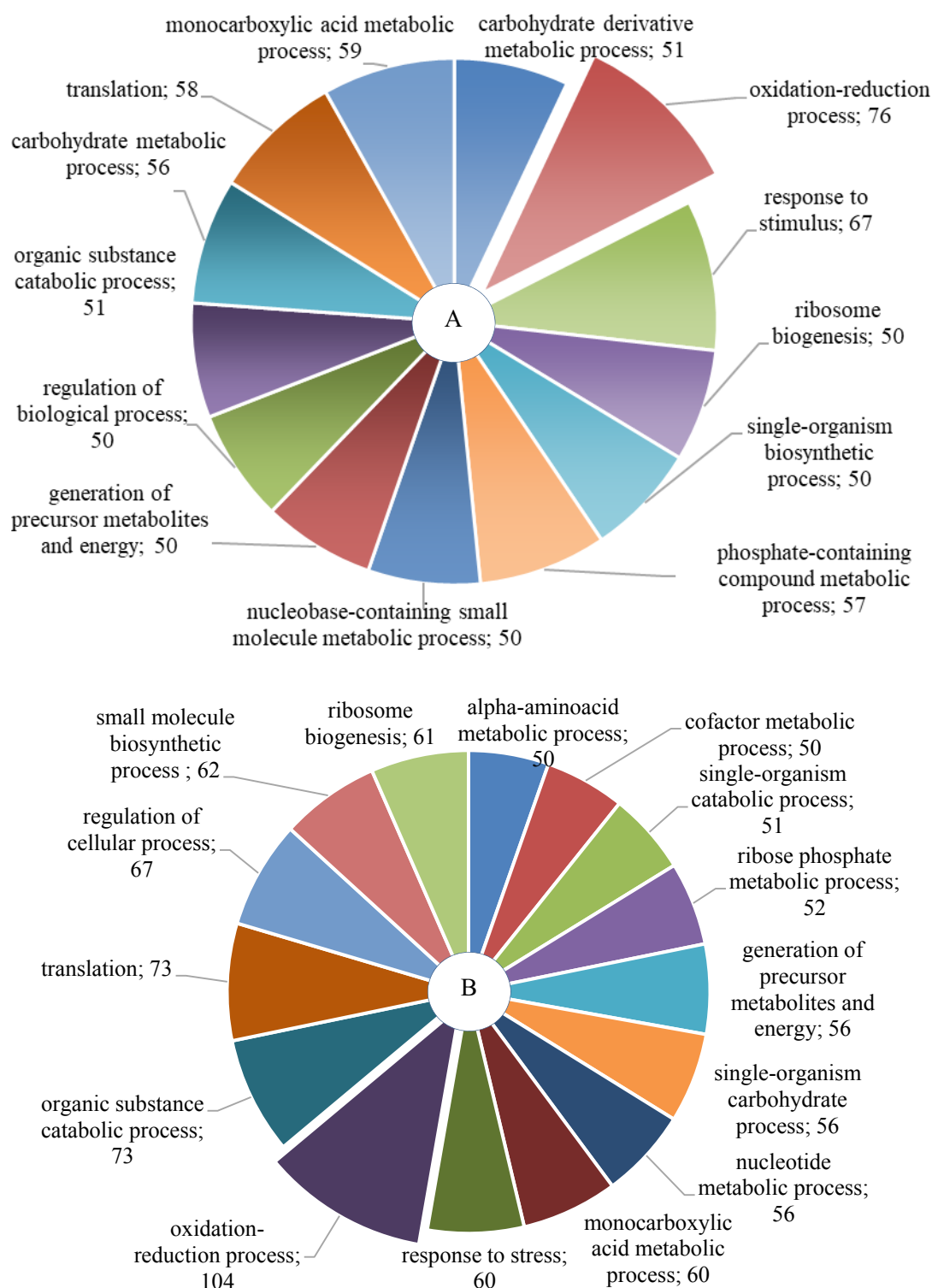


Figure 1. Graphic representation of the distribution of proteins by Biological Process (BP) Gene Ontology, present in the seeds of the two species studied: (A) *Pinus greggii*, (B) *Pinus patula*. Only those processes involving at least 50 proteins are represented. The numbers indicate the amount of proteins that participate in each of the processes represented.

Representación gráfica de la distribución de proteínas por Proceso Biológico (BP) GO, presente en las semillas de las dos especies estudiadas. (A) *Pinus greggii*, (B) *Pinus patula*. Solo están representados aquellos procesos que involucran al menos 50 proteínas. Los números indican la cantidad de proteínas que participan en cada uno de los procesos representados.

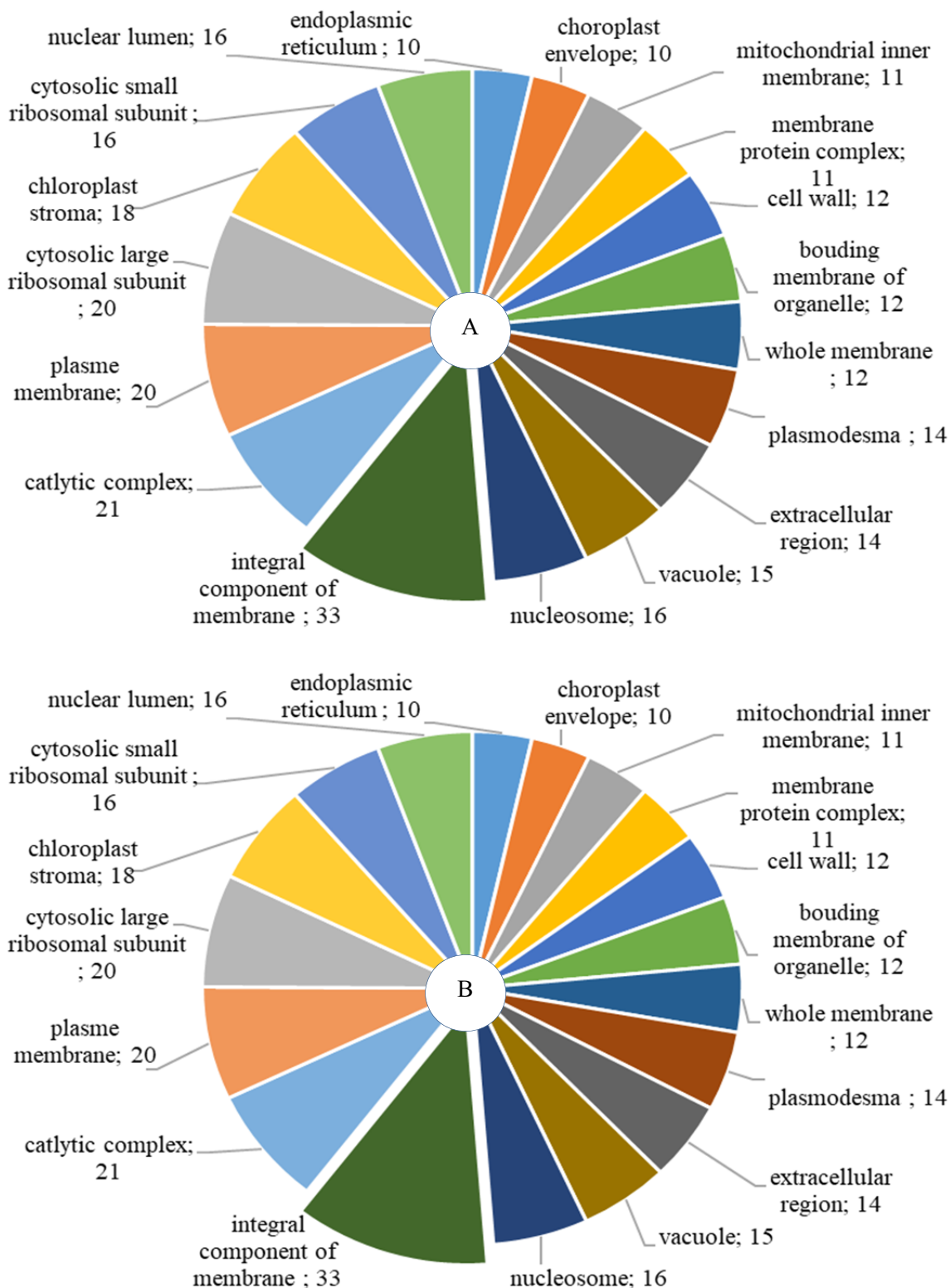


Figure 2. Graphic representation of the distribution of proteins taking into account their location in cellular components (CC) Gene Ontology, in the seeds of the two species studied: (A) *Pinus greggii*, (B) *Pinus patula*. Only those processes involving at least 10 proteins are represented. The numbers indicate the amount of proteins involved in each cell component.

Representación gráfica de la distribución de las proteínas según su ubicación en los componentes celulares (CC) GO, en las semillas de las dos especies estudiadas. (A) *Pinus greggii*, (B) *Pinus patula*. Solo están representados aquellos procesos que involucran al menos 10 proteínas. Los números indican la cantidad de proteínas involucradas en cada componente celular.

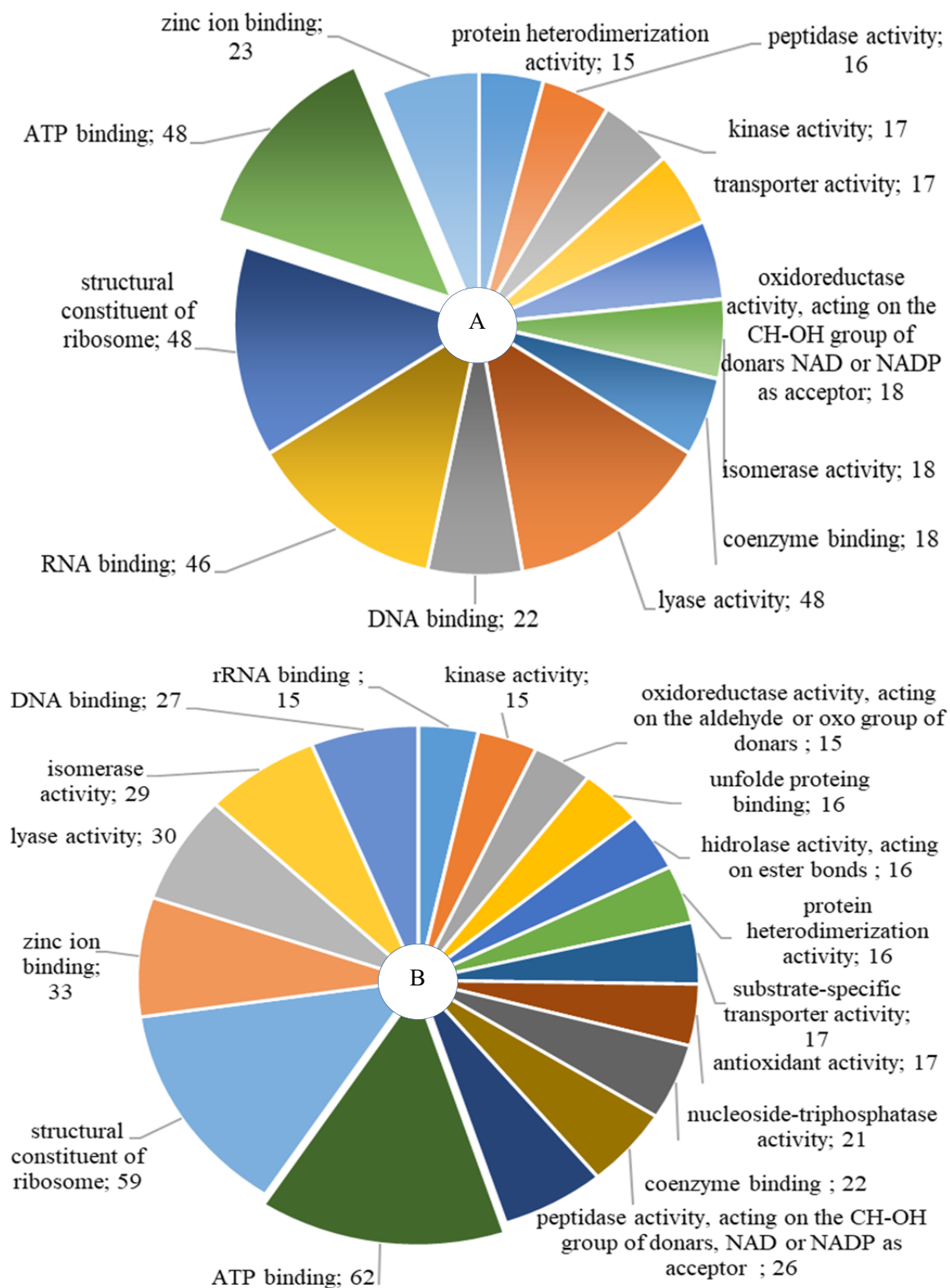


Figure 3. Graphic representation of the distribution of proteins by Molecular Function (FM) Gene Ontology, in the seeds of the two species studied: (A) *Pinus greggii*, (B) *Pinus patula*. Only functions with a minimum of 15 proteins are represented. The numbers indicate the amount of proteins involved in each of the molecular functions. Data for a significance threshold FDR < 0.05.

Representación gráfica de la distribución de proteínas por Función Molecular (FM) GO, en las semillas de las dos especies estudiadas. (A) *Pinus greggii*, (B) *Pinus patula*. Solo están representadas las funciones con un mínimo de 15 proteínas. Los números indican la cantidad de proteínas involucradas en cada una de las funciones moleculares. Datos para un umbral de significación FDR < 0,05.

Table 1. List of proteins involved in the metabolic pathway glycolysis/gluconeogenesis expressed in the seeds of *Pinus greggii* and *Pinus patula*.

Proteínas identificadas implicadas en la ruta metabólica de la glicólisis/gluconeogénesis expresadas en las semillas de *Pinus greggii* y *Pinus patula*.

Acc. Num. ^a	Description ^b	MW [kDa] calc ^c	pI ^d	Coverage (%)	# Unique Peptides	Commons	<i>Pinus greggii</i>	<i>Pinus patula</i>
1032	ATP-dependent 6-phosphofructokinase 2	52	7	5	1	-	X	-
10495	phosphoglycerate chloroplastic	53	9	4	1	X	-	-
10634	phosphoenolpyruvate carboxykinase [ATP]-like	73	7	4	2	-	-	X
1069	dihydrolipoyllysine-residue acetyltransferase component 2 of pyruvate dehydrogenase mitochondrial-like	62	9	3	1	-	X	-
11278	methylmalonate-semialdehyde dehydrogenase [acylating] mitochondrial	60	7	2	1	-	-	X
126671	Enolase	48	6	29	7	X	-	-
126855	glyceraldehyde-3-phosphate dehydrogenase cytosolic	39	8	62	15	X	-	-
127205	Alcohol	51	8	15	1	-	-	-
12998	pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha	67	7	3	1	-	-	X
13226	betaine aldehyde dehydrogenase chloroplastic	27	5	6	1	-	-	X
16504	pyruvate cytosolic isozyme	57	8	5	2	X	-	-
16934	fructose-bisphosphate cytoplasmic isozyme 1	42	9	24	5	X	-	-
17023	fructose-bisphosphate aldolase chloroplastic	49	8	5	2	X	-	-
17064	fructose-bisphosphate aldolase cytosolic	44	8	52	15	X	-	-
17080	pyruvate dehydrogenase E1 component subunit beta-mitochondrial	41	5	7	1	-	-	X
17339	Alcohol	50	8	22	2	-	-	X
18066	phosphoenolpyruvate carboxykinase [ATP]-like	49	9	4	1	-	-	X
18227	glucose-6-phosphate isomerase chloroplastic	71	6	12	5	X	-	-
18273	glucose-6-phosphate cytosolic	68	7	7	3	X	-	-
1839	Alcohol	41	6	11	1	-	-	X
20329	aldehyde dehydrogenase family 2 member mitochondrial	63	8	20	6	X	-	-
2037	glyceraldehyde-3-phosphate cytosolic-like	40	7	51	1	-	X	-
20710	cytosolic triosephosphate isomerase	27	5	30	7	-	X	-
208477	alcohol dehydrogenase class-3	51	8	23	5	-	-	X
21225	glyceraldehyde-3-phosphate dehydrogenase cytosolic	40	7	36	1	-	X	-
22555	Enolase	48	6	45	1	-	X	-
2308	betaine aldehyde dehydrogenase mitochondrial	59	6	2	1	X	-	-
23105	phosphoglycerate kinase	52	7	20	1	-	X	-
2317	pyruvate dehydrogenase E1 component subunit alpha-mitochondrial	44	8	3	1	X	-	-
3137	Dihydrolipoyl dehydrogenase mitochondrial	54	8	16	6	X	-	-
3418	aldose reductase	39	7	44	16	X	-	-
37838	phosphoglycerate kinase	44	8	16	7	X	-	-
4123	dihydrolipoyl dehydrogenase chloroplastic	61	7	2	1	X	-	-
41574	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	54	7	7	2	X	-	-
44645	aldehyde dehydrogenase family 2 member mitochondrial-like	59	7	8	3	X	-	-
72	pyruvate decarboxylase 2	66	6	5	1	-	-	X
782	Alcohol	48	8	25	3	X	-	-
893	aldehyde dehydrogenase	47	8	17	4	X	-	-
9616	pyruvate decarboxylase 2	67	6	7	1	X	-	-

^a *Pinus_DB Pinus sustainipine v3*. ^b Blast Description Annotator. ^c Molecular weight (kDa) and isoelectric point of protein calculated for each database.

^d Molecular weight (kDa) and isoelectric point of protein calculated by using molecular weight standards.

related to the reserve in seeds are necessary for the mobilization and preparation of the seed protein during germination (Wang *et al.* 2019).

The results of the differential expression analysis showed a high similarity between the proteins profiles of

the two pine species studied, agreeing with Dvorak *et al.* (2008) who found a very little genetic distance between *P. greggii* and *P. patula*. These authors found high levels of variability in the morphology of Mexican *Pinus* spp. accompanied with low levels of interspecific genetic diffe-

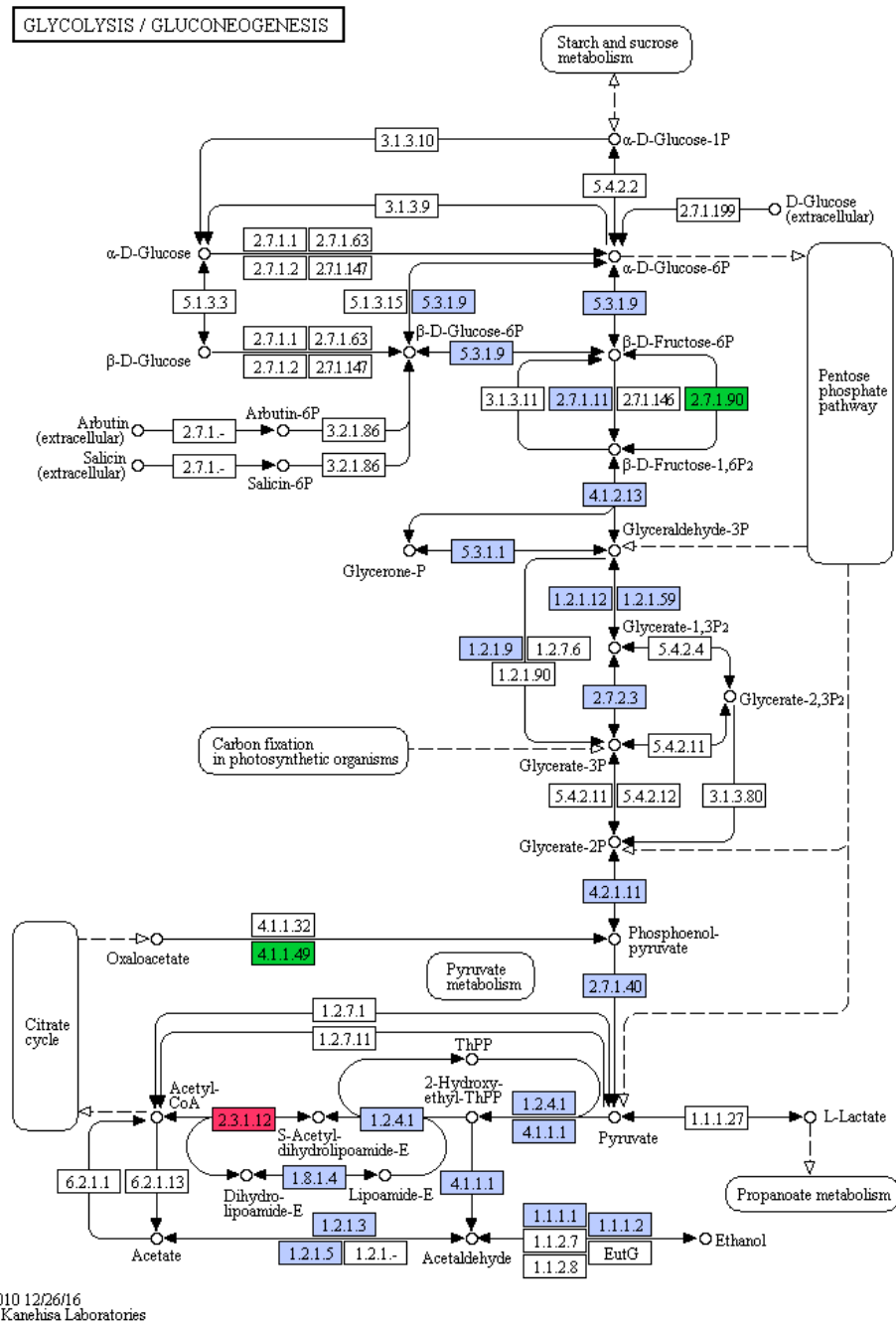


Figure 4. Graphical representation of the metabolic pathway glycolysis/gluconeogenesis KEGGs, in the seeds of the two species. The enzymes that participate in this route are represented with colors. Blue represents the enzymes common in the two species, red represents the enzymes identified only in *P. greggii* and green represents those identified only in the *P. patula* species.

Representación gráfica de la ruta metabólica de la glicólisis/gluconeogénesis KEGGs, en las semillas de las dos especies. Las enzimas que participan en esta ruta se representan con colores. El color azul representa las enzimas comunes en las dos especies, el color rojo representa las enzimas identificadas solo en *Pinus greggii* y el color verde representa las identificadas solo en *Pinus patula*.

rentiation, which may be the result of low rates of evolution among the species of this genus. Moreover, conifers are considered a monophyletic group that descends from the same common ancestor more than 300 million years ago, being their low genetic variability confirmed by several authors (De la Torre *et al.* 2019).

Analysis of the glycolysis/gluconeogenesis pathway. Authors like Chen *et al.* (2016) state that the glycolysis/gluconeogenesis route is one of the most important in seeds, representing an indispensable way to maintain the balance of cellular ATP levels and carbohydrate metabolism. This balance is very important in the seeds of higher plants, being starch the main storage carbohydrate (Duan *et al.* 2016). Glycolysis is a catabolic pathway in which glucose is broken down into two molecules of pyruvic acid under aerobic conditions, or lactate under anaerobic conditions. Complete oxidation of glucose can involve three fundamental biochemical pathways: i) the glycolytic pathway, ii) the Krebs cycle (also called citric acid cycle or tricarboxylic acid cycle) and iii) oxidative phosphorylation. Conversely, gluconeogenesis is the anabolic route which converts non-sugar precursors (lactate, pyruvate, propionate, glycerol and amino acids) into glucose. Therefore, the glycolysis/gluconeogenesis pathway is indispensable to maintain the balance of cellular levels of ATP and carbohydrate metabolism (Chen *et al.* 2016) and starch, which constitutes the main carbohydrate storage in seeds of higher plants (Duan *et al.* 2016). As storage reserves provide the energy necessary for germination, it was expected an overexpression of these proteins in seeds of quick germination (Rosental *et al.* 2014).

Specifically, in our work, 19 enzymes involved in this route were expressed in both species, and 3 were also expressed only in one of them. The acetyltransferase, only identified in *P. greggii*, is an enzyme which catalyzes the oxidative decarboxylation of pyruvate to produce acetyl-CoA. This enzyme is part of the multi-enzymatic pyruvate dehydrogenase (PDC) complex of the mitochondrial matrix in which the acetyl group is passed to coenzyme A, which constitutes a key compound between glycolysis and the Krebs cycle. Some works (Wang *et al.* 2019) that identify this enzyme correlated the expression of PDC with the accumulation of lipids in the seeds, which are relevant for the reorganization of the metabolism of seeds, energy and carbon for the growth of the embryo and the protrusion of radiation (Rosental *et al.* 2014).

Diphosphate-fructose-6-phosphate 1-phosphotransferase, only identified in *P. patula*, catalyzes the reversible interconversion between fructose-6-phosphate and fructose-1,6-bisphosphate, which constitutes a limiting step in the regulation of the primary metabolic flow of carbohydrates towards glycolysis or gluconeogenesis. This enzymatic activity is similar to the activity of enzyme 6-phosphofructokinase, which is common to both species (Duan *et al.* 2016), suggesting the presence of both proteins in *P.*

patula and a higher ability to catalyze the flow of carbohydrates to glycolysis. On the other hand, the PEPCK protein, also identified only in *P. patula*, catalyzes the conversion of ATP and oxaloacetate into ADP, phosphoenolpyruvate and carbon dioxide, the first step of the gluconeogenesis route, converting the oxaloacetate into phosphoenolpyruvate. These results suggest that the *P. patula* species could initiate gluconeogenesis using both oxaloacetate and pyruvate as substrate, whereas the *P. greggii* species could hardly use pyruvate as a substrate, which is corroborated by Ahern and Rajagopa (2019) by proposing that in the gluconeogenesis process, it can start from several simple substrates, especially in plants.

The seeds of several *Pinus* spp. presented proteins associated with metabolic functions involved in germination processes (Rey *et al.* 2019). Thus, they are relatively rich in storage proteins synthesized during the last stage of seed development.

Oxidoreductase enzymes overexpression. Results indicate also an overexpression of oxidoreductase enzymes. This coincides with the results of the Gene Ontology biological processes above discussed, which place the redox processes as the most important in terms of the number of proteins involved, followed by the transferase class in *P. greggii* and the hydrolases in *P. patula*. According to the Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) for the general classification of enzymes, it is considered that hydrolases are a special class of transferases in which water serves as an acceptor of the transferred group. Likewise, the work of Behar *et al.* (2018) indicates the existence of flexibility in the specificity of the reactions that catalyze oxidoreductase enzymes.

CONCLUSIONS

The study carried out permitted to characterize the protein profile of the seeds of *P. patula* and *P. greggii*, identifying the main proteins presents in both species. The study of the functional profiles according to terms of the Gene Ontology made the classification of the main biological categories (biological processes, cellular components and molecular function) possible, in which these proteins and their metabolic profiles are involved, comparing the results between both species. The results showed that there is no relevant genetic variability between seeds except for the glycolysis/gluconeogenesis pathway, being expressed in *P. patula* a specific protein which could provide more ability in the mobilization of seed carbohydrates. Neither do they show significant differences in the main functional processes in which these proteins and their metabolic associated profiles are involved. Gene Ontology database search did not show enrichment in any category (between species), since most of these categories had a similar representation in the two profiles analyzed, with the associated proteins predominating in the preparation of the seed during the

germination process. Therefore, both *P. patula* and *P. greggii* presented similar characteristics in the proteomic profile of seeds, being indistinguishable regarding their ability to plant establishment, except for the better ability of *P. patula* to mobilize carbohydrates, which could give this species a slight advantage in the germination process.

ACKNOWLEDGMENTS

This research was funded by Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) through the project “System management of soil fertility, integrating the rational use of fertilizers in agriculture and Bio fertilizers on the basis of rhizobia-Task 147” by the Research Project “Evaluation of the differential response of two *Pinus* species to different climatic and inoculation conditions using molecular techniques” carried out by the University of Córdoba (Spain). The authors are grateful for the technical support of Central Research Support Service (SCAI) of University of Córdoba. Specifically, to the heads of the Proteomics Unit, Consuelo Gómez, Carlos Fuentes and Raquel González, where the identification of proteins was carried out, and to the head of the Bioinformatics Unit, Francisco Amil, where the functional classification of proteins was carried out. Special thanks to the technical staff of the ERSAF research group, Roberto Cabrera Puerto, for their support in the laboratory work. Finally, authors want to thank the heads of the research project “Evaluation of the differential response of two *Pinus* species to different climatic and inoculation conditions using molecular techniques”, Dr. Rafael M^a Navarro (Professor of University of Córdoba) and Dr. Virginia Lacerda (Vice-rector of José Eduardo dos Santos University) for their support in the design of the work.

REFERENCES

- Ahern K, I Rajagopal. 2019. Biochemistry and Biophysics Electronic Book. Oregon, United State. Libretex.
- Behar H, SW Graham, H Brumer. 2018. Comprehensive cross-genome survey and phylogeny of glycoside hydrolase family 16 members reveals the evolutionary origin of EG16 and XTH proteins in plant lineages. *Plant Journal* 95(6): 1114-1128. DOI: <https://doi.org/10.1111/tpj.14004>
- Casique R, R Mendoza, F Galindo, S González, S Sanchez. 2019. Improved parameters of *Pinus greggii* seedling growth and health after inoculation with ectomycorrhizal fungi. *Journal of Southern Forests* 81(1): 23-30. DOI: <https://doi.org/10.2989/20702620.2018.1474415>
- Chen X, K Schreiber, J Appel, A Makowka, B Fähnrich, M Roettger, MR Hajirezaei, FD Sönnichsen, P Schönheit, W Martin, K Gutekunst. 2016. The Entner-Doudoroff pathway is an overlooked glycolytic route in cyanobacteria and plants. *Proceedings of the National Academy of Sciences of the United States of America* 113(19): 5441-5446. DOI: <https://dx.doi.org/10.1073/pnas.1521916113>
- Conesa A, S Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal Plant Genomics* 2008: 619832. DOI: <https://doi.org/10.1155/2008/619832>
- De la Torre AR., A Piot, B Liu, B Wilhite, M Weiss, I Porth. 2019. Functional and morphological evolution in gymnosperms: A portrait of implicated gene families. *Evolutionary Applications* 13(1): 210–227. DOI: <https://doi.org/10.1111/eva.12839>
- Duan E, Y Wang, L Liu, J Zhu, M Zhong, H Zhang, S Li, B Ding, X Zhang. 2016. Pyrophosphate: fructose-6-phosphate 1-phosphotransferase (PF6) regulates carbon metabolism during grain filling in rice. *Plant Cell Reports* 35: 1321-1331. DOI: <https://doi.org/10.1007/s00299-016-1964-4>
- Dvorak WS. 2008. ¿estamos más cerca de entender la ascendencia de las poblaciones de la “variante patula” en la sierra madre del sur, México? *Foresta Veracruzana* 10(1): 59–66. DOI: <https://www.redalyc.org/articulo.oa?id=49711434008>
- Fernández N, J Canales, D Guerrero, DP Villalobos, SM Díaz, R Bautista, A Flores, MA Guevara, P Perdiguero, C Collada, MT Cervera, A Soto, R Ordás, FR Cantón, C Avila, FM Cánovas, MG Claros 2011. EuroPineDB: a high-coverage web database for maritime pine transcriptome. *BMC Genomics* 12: 366. DOI: <https://doi.org/10.1186/1471-2164-12-366>
- Kovačević D, BM Nikolić, S Mladenović-Drinić, SR Bojović, T Dodos, N Rajčević, DP Marin. 2013. Genetic relationships among some *Pinus*, *Picea* and *Abies* species revealed by RAPD markers. *Genetika* 45(2): 493-502. DOI: <http://dx.doi.org/10.2298/GENSR1302493K>
- Loewe V, RM Navarro, R Sánchez, FJ Ruiz, J Jorrín. 2018. Variability studies of allochthonous stone pine (*Pinus pinea* L.) plantations in Chile through nut protein profiling. *Journal of Proteomics* 175: 95-104. DOI: <https://doi.org/10.1016/j.jprot.2018.01.005>
- Maldonado AM, S Echevarría, S Jean, M Hernández, JV Jorrín. 2008. Evaluation of three different protocols of protein extraction for Arabidopsis thaliana leaf proteome analysis by two-dimensional electrophoresis. *Journal of Proteomics* 71(4): 461-472. DOI: <https://doi.org/10.1016/j.jprot.2008.06.012>
- Morel A, JF Trontin, F Corbineau, A Lomenech, M Beaufour, I Raymond, C Le Metté, K Ader, L Harvengt. 2014. Cotyledonary somatic embryos of *Pinus pinaster* Ait. most closely resemble fresh, maturing cotyledonary zygotic embryos: biological, carbohydrate and proteomic analyses. *Planta* 240: 1075-1095. DOI: <https://doi.org/10.1007/s00425-014-2125-z>
- Muniz PH, G Guimarães, GH Peixoto, W Sousa, D de Souza, F Rodrigues, DD Carvalho. 2020. Scientometric Analysis of the major *Pinus* spp. Diseases. *International Journal of Current Microbiology and Applied Sciences* 9(3): 2825-2830. DOI: <https://doi.org/10.1007/s00425-014-2125-z>
- Pullman GS, X Zeng, B Copeland, J Crockett, J Lucrezi, SW May, K Bucalo. 2015. Conifer somatic embryogenesis: improvements by supplementation of medium with oxidation-reduction agents. *Tree physiology* 35(2): 209–224. DOI: <https://doi.org/10.1093/treephys/tpu117>
- Rey MD, MA Castillejo, R Sánchez, VM Guerrero, C López, C Romero, J Valero, B Sghaier, L Simova, S Echevarría, I Jorge, I Gómez, ME Papa, K Carvalho, LE Rodríguez de Francisco, AM Maldonado, L Valledor, JV Jorrín. 2019. Proteomics, Holm Oak (*Quercus ilex* L.) and Other Recalcitrant and Orphan Forest Tree Species: How do They See

- Each Other? *International Journal of Molecular Sciences* 20(3): 692. DOI: <https://doi.org/10.3390/ijms20030692>
- Rodríguez de Francisco L, MC Romero, RM Navarro, V Miniño, O Perdomo, JV Jorrín. 2016. Characterization of the orthodox *Pinus occidentalis* seed and pollen proteomes by using complementary gel-based and gel-free approaches. *Journal of Proteomics* 143(30): 382-389. DOI: <https://doi.org/10.1016/j.jprot.2016.03.041>
- Romero MC, J Pascual, L Valledor, JV Jorrín. 2014. Improving the quality of protein identification in non-model species. Characterization of *Quercus ilex* seed and *Pinus radiata* needle proteomes by using SEQUEST and custom databases. *Journal of proteomics* 105: 85-91. DOI: <https://doi.org/10.1016/j.jprot.2014.01.027>
- Rosental L, H Nonogaki, A Fait. 2014. Activation and regulation of primary metabolism during seed germination. *Seed Science Research* 24(1): 1-15. DOI: <https://doi.org/10.1017/S0960258513000391>
- Singh L, P Dixit, RS Prakash, S Pandey, PV Chandra. 2019. Morpho-anatomical variation and their phylogenetic implications in native and exotic species of *Pinus* growing in the Indian. *BioRxiv* 754804. DOI: <https://doi.org/10.1101/754804>
- Taïbi K, D del Campo, A Vilagrosa, M Bellés, M López, D Pla. 2017. Drought Tolerance in *Pinus halepensis* seed sources as identified by distinctive physiological and molecular markers. *Frontiers in Plant Science* 8(1202): 1-13. DOI: <https://doi.org/10.3389/fpls.2017.01202>
- Wang W, R Vignani, M Scali, M Cresti. 2006. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis* 27(13): 2782-2786. DOI: <https://doi.org/10.1002/elps.200500722>
- Wang X, Y Shi, R He, B Li, A Huang. 2019. Label-free quantitative proteomic analysis of the biological functions of *Moringa oleifera* seed proteins provides insights regarding the milk-clotting proteases. *International Journal of Biological Macromolecules* 1(144): 325-333. DOI: <https://10.1016/j.ijbiomac.2019.12.070>
- Woo C, R Gupta, K Agrawa, R Kumar, S Tae. 2019. Concepts and strategies of soybean seed proteomics using the shotgun proteomics approach. *Journal Expert Review of Proteomics* 16(9): 795-804. DOI: <https://doi.org/10.1080/14789450.2019.1654860>
- Zhang H, WQ Wang, SJ Liu, IM Møller, SQ Song. 2015. Proteome analysis of poplar seed vigor. *PloS one* 10(7): e0132509. DOI: <https://doi.org/10.1371/journal.pone.0132509>

Recibido: 16/04/20
Aceptado: 08/09/20