INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
18–1080/2019/21–3–679–688
DOI: 10.17957/IJAB/15.0945
http://www.fspublishers.org

Full Length Article

Diversity in Seed Storage Protein Profile of Oilseed Crop *Plukenetia volubilis* from Peruvian Amazon

Petra Hlásná Čepková[1], Michal Jágr[2,4*], Iva Viehmannová[3], Václav Dvořáček[4], Danter Cachique Huansi[5] and Ivan Mikšík[2]

[1]*Gene Bank, Crop Research Institute, Drnovská 507/73, 161 06 Prague 6 – Ruzyně, Czech Republic*
[2]*Department of Analysis of Biologically Important Compounds, Institute of Physiology of the Czech Academy of Sciences, Vídeňská 1083, 142 20 Prague 4, Czech Republic*
[3]*Department of Crop Sciences and Agroforestry, Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Kamýcká 129, 16521 Prague 6, Czech Republic*
[4]*Quality of Plant Products, Crop Research Institute, Drnovská 507/73, 161 06 Prague 6 – Ruzyně, Czech Republic*
[5]*Peruvian Amazon Research Institute, Jr. Belén Torres de Tello 135, Morales, Tarapoto, Peru*
[*]For correspondence: michaljagr@centrum.cz

Abstract

Sacha inchi (*Plukenetia volubilis* L.) is a plant native to the Peruvian Amazon and produces seeds rich in some nutraceutical compounds with a high protein content. The present work aimed to characterize *P. volubilis* seed proteins and compare their proteomic profile between different *P. volubilis* populations. Crude protein content in the seeds was detected at a mean level of 22.56% with the major proportion of albumins and globulins (16.37%), followed by glutelins (5.87%) and a very low content of prolamins (0.33%). Sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis of both total seed flour and Osborne´s protein fractions of different populations showed that proteins were concentrated in the 8–75 kDa range. Differences in abundance of proteins between Sacha inchi populations were detected by SDS-PAGE and two-dimensional (2-DE) gel electrophoresis. Gel protein bands and spots of interest were excised, digested with trypsin and analyzed by nano-liquid chromatography/tandem mass spectrometry (nLC-MS/MS). This is the first direct identification of these proteins in actual Sacha inchi seeds. Three proteins (Oleosin 2, Oleosin 3 and elongation factor 1-alpha) directly from *P. volubilis* and peptides with a sequence identical to 56 proteins from the Euphorbiaceae family were detected. This characterization of seed storage protein, their variability within populations and identification of important proteins provide a basis for further investigations for the food industry and bioengineering. © 2019 Friends Science Publishers

Keywords: Genetic variability; Oilseeds; Proteomics; Storage proteins
Abbreviations: 2-DE, two-dimensional gel electrophoresis; AFLP, amplified fragment length polymorphism; HSD, honest significant difference; IEF, isoelectric focusing; ISSR, inter simple sequence repeat; LC, liquid chromatography; MS, mass spectrometry; MW, molecular weight; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SSP, seed storage proteins

Introduction

Oil seed crops (*i.e.*, worldwide cultivated soya beans with 35% of protein, nuts with 25% protein or more recently cereals and pseudocereals) have been studied as a good source of valuable proteins (Sathe, 1994; Sugano, 2006; Khan *et al.*, 2014; Mota *et al.*, 2015). The importance of plant proteins for human and animal nutrition has been debated intensively. About 80% of proteins consumed by people in developing countries are supplied by plants and this trend has not changed in the past two decades (Day, 2013). Plant proteins are used for food product production, in animal nutrition and plant protection, and in the chemical industry (Bejosano and Corke, 1999; Santoni and Pizzo, 2013). For example, the protein concentrates are produced as by-products after oil extraction in *Jatropha* curcas (Valdez-Flores *et al.*, 2016) or after starch extraction as *Amaranthus* and buckwheat (Bejosano and Corke, 1999).

During the past decade, *Plukenetia volubilis* L. (Euphorbiaceae) has been regarded as a promising crop with great potential in nutrition and for applications in the pharmaceutical industry. It is considered as a superfood mainly due to its high protein content and quality oil (Gutiérrez *et al.*, 2011). This is a wild, climbing, semi-woody, perennial, and oleaginous plant that grows in tropical jungles of South America at altitudes of between 200 and 1500 m above sea level (Gillespie, 1994; Arévalo, 1995). Its fruit capsule usually contains four seeds. These

To cite this paper: Čepková, P.H., M. Jágr, I. Viehmannová, V. Dvořáček, D.C. Huansi and I. Mikšík, 2019. Diversity in seed storage protein profile of oilseed crop *Plukenetia volubilis* from Peruvian amazon. *Intl. J. Agric. Biol.*, 21: 679–688

seeds are of lenticular flattened shape, 1.3–2.1 cm long, and brown in colour with darker margins (Gillespie, 1994). *P. volubilis* (Sacha inchi or "peanut of the Incas") is a highly nutritious crop with seeds containing an unprecedentedly high amount of oil (54%) and a relatively high amount of protein (27%), comparable to soya beans (34%) and used by indigenous people of the Peruvian Amazon for traditional food preparation (Hamaker *et al.*, 1992). According to many recently published studies the great potential of Sacha inchi is in the production of new agro-industrial by-products for the nutraceutical and functional food industries (Chirinos *et al.*, 2016).

All these reasons mentioned above make Sacha inchi's seeds very attractive for the manufacturing industry. However, few data about Sacha inchi's proteins and their solubilization are available to date (Sathe *et al.*, 2012) and with no attempts to evaluate protein polymorphism within different localities and identify the studied proteins by mass spectrometry (MS) technique. Two-dimensional gel electrophoresis in combination with MS is currently the workhouse for proteomics (Magdeldin *et al.*, 2014). In plants, especially for members of Euphorbiaceae family, this technique has been successfully used to characterize proteomes of cassava (*Manihot esculenta*), castor bean (*Ricinus communis*) (Campos *et al.*, 2010; Souza *et al.*, 2015) and *Jatropha curcas* seeds (Shah *et al.*, 2015).

The main objective of this study was to determine seed storage protein content and protein fractions (data about fractionation remain scarcely explored). Another aim of our research has been to determine differences in proteomic composition of seeds of various *P. volubilis* populations, since such protein profiles could to reveal a surprising amount of information about genetic variability of different populations. The final aim was to identify major proteins present in *P. volubilis* seeds. Herein, we report the first characterization of the proteome of *P. volubilis* seeds using 2-DE and tandem mass spectrometry.

Materials and Methods

Plant Material

Collection of *P. volubilis* seeds was done in cooperation with the Peruvian Amazon Research Institute in Tarapoto, Peru, in 2012. Seeds for analysis were harvested in different localities of the Peruvian Amazon, in the region of San Martín. Seeds of ten populations and 171 individual trees were collected, as shown in Table 1. The number of individual samples in each population varied from 2 to 23. A bulked sample was always formed by mixing all individuals of the distinct population. All seeds were stored at 4°C before further processing.

Sample Preparation

Approximately eight seeds (ca 10 g) from all collected individual plants were crushed in a grinding mill (IKA A11 basic, IKA® Werke GMBH & Co.KG, Germany) to create individual samples. Similarly, ten seeds from each genotype (population) were selected randomly, crushed and mixed together to form bulked samples. Samples were further processed according to the procedure shown in Fig. 1.

Crude Protein and Protein Fraction Quantification

The dry matter content of seed samples (5 g Sacha inchi flour) was dried in an electric hot-air drier at 105°C for 4 h according to the standard method CSN EN ISO 662 (CSN, 2001). The content of crude protein from each sample was determined using the classical Kjeldahl mineralization method and calculated with conversion factor 6.25 (CSN, 2012; FAO, 2002). Then three protein fractions (first albumins and globulins, second prolamins, and third glutelins) were isolated according to Osborne´s method with a slight modification according to Dvořáček *et al.* (2001).

Statistical Analysis

Final data was analysed by analysis of variance (ANOVA) using Tukey honest significant difference (HSD) test to identify significances between samples from different localities. Statistical analysis was performed by the software Statistica12.0 (StatSoft, Czech Republic) and significance was established at a p-value ≤ 0.05.

Protein and Protein Fractions Solubilization and Extraction

Seed flour from individual plant and/or bulked samples was de-fatted by washing 1 g of flour with 2.5 mL of hexane at 25°C overnight. Then, the hexane fraction was discarded and the flour was lyophilized in aliquots and stored in a cold (4°C), dark place. The de-fatted aliquots were used following Osborne fractionation and/or for SDS-PAGE and 2-DE analysis. Protein concentration in the sample aliquots and pellets was determined with Bradford assay using Bradford Dye Reagent and bovine serum albumin standard (Bio-Rad, Czech Republic).

Sample Preparation for Electrophoretical Osborne Protein Fractionation

The aliquot (0.1 g of sample) was subjected to extraction of protein fractions (Džunková *et al.*, 2011). Tubes containing protein fraction extracts and seed pellets (glutelins) were frozen and lyophilized in a freeze dryer (Christ, Germany) for 24 h at -58°C and 0.018 mBar and stored at 4°C until analysis by SDS-PAGE. Alternatively, ca. 40 mg of every Osborne´s fraction was mixed with trypsin (ratio 1:50, enzyme:protein, w/w) in 50 m*M* ammonium bicarbonate and subjected to cleavage overnight at 37°C. The samples were then processed as described in Eckhardt *et al.* (2014) and the resulting dried tryptic peptide extracts were stored at –80°C before LC-MS/MS analysis.

Table 1: Geographical coordinates of the areas in Peruvian Amazon where the samples of *P. volubilis* were collected

No. of locality	Population	No. of samples	Latitude	Longitude	Altitude (m.a.s.l.)
1	Aguas de Oro (ADO)	20	6°17,570' S	76°39,200' W	385
2	Santa Cruz (SCR)	23	6°36,803' S	76°44,452' W	425
3	Mishquiyacu (MIS)	20	6°21,673' S	76°34,998' W	470
4	Chumbaquihui (CHU)	20	6°21,991' S	76°34,504' W	364
5	Aucaloma (AUC)	18	6°24,816' S	76°26,143' W	740
6	Pacchilla (PAC)	20	6°25,694' S	76°27,729' W	703
7	Ramón Castillo (RAC)	5	6°35,244' S	76°07,884' W	210
8	Santa Lucia (SLU)	2	9°07,622' S	76°01,040' W	562
9	Pucallpa (PUC)	21	6°25,676' S	76°34,689' W	455
10	Dos de Mayo (2DM)	22	6°47,573' S	76°32,108' W	335
Total		171			

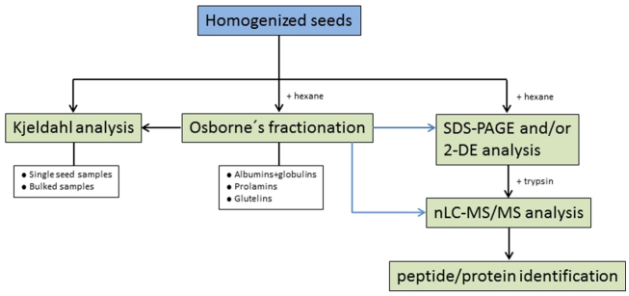

Fig. 1: Flow chart of the sequential *P. volubilis* sample processing and methods used for analysis of the seed samples

Protein Separation by SDS-PAGE

The lyophilized total seed protein and Osborne protein fractions obtained from bulked samples (1 mg) of each population were mixed with 100 μL of sample buffer (0.0625 *M* TrisHCl pH 6.8, 5% (w/v) 2-mercaptoethanol, 2% (w/v) SDS, 10% (w/v) glycerol, 0.01% (w/v) bromophenol blue) and vortexed several times in 1.5 mL tubes. The tubes were heated to 95°C for 5 min. One-dimensional SDS-polyacrylamide gel electrophoresis, was performed using a homogeneous 4% (w/v) stacking gel and 12.5% or 15% resolving gel of 1 mm thickness, as described by Jágr *et al.* (2012). SDS-PAGE analyses were performed in triplicate.

Protein Separation by 2-DE

Proteins (5 mg) in the lyophilized bulked samples from each population were solubilized in 600 μL of lysis buffer (7 *M* urea, 2 *M* thiourea, 2% (w/v) 3-((3-cholamidopropyl) dimethylammonium)-1-propanesulphonate (CHAPS), 0.2% (w/v) Bio-Lyte ampholytes (3–10 buffer), and 1% (w/v) dithiothreitol. Then, the samples were diluted with lysis buffer to protein concentration ca 0.84 mg mL^{-1}, as determined with Bradford assay. Ready Strip IPG Strips (pH 3–10 non-linear, 7 cm; Bio-Rad, Czech Republic) were rehydrated with 105 μg of protein samples in 125 μL lysis buffer by active in-gel rehydration (50 V, 20°C) overnight. Isoelectric focusing was carried out at 15°C with a Protean IEF cell system (Bio-Rad, Czech Republic) at 4,000 V for 20,000 Vh (Eckhardt *et al.*, 2014). Homogeneous 12.5% SDS-polyacrylamide gel was used as second dimension. Finally, the 2-DE gels were stained with Bio-Safe Coomassie Blue G250 stain (Bio-Rad, Czech Republic) and 2-DE gel analyses were repeated at least twice.

Protein Characterization using LC-MS/MS

Protein bands or spots were excised from Coomassie Blue-stained SDS-PAGE or 2-DE gels and then processed (Eckhardt *et al.*, 2014). The resulting dried tryptic peptide extracts were stored at –80°C before analysis.

Analysis of the tryptic digests with nano-liquid chromatography tandem mass spectrometry (maXis, quadrupole time-of-flight as mass spectrometer) was performed as in the previous study of Eckhardt *et al.* (2014). All LC-MS/MS analysis were performed in duplicate.

Data Analysis and Database Searching

Data were processed with ProteinScape software (version 3.0.0.446, Bruker Daltonics, Germany). Proteins were identified by correlating tandem mass spectra to the NCBInr database (downloaded on 28.3.2017), and using the MASCOT search engine (v. 2.3.0). The taxonomy was restricted to *Viridi plantae* (green plants) to remove protein identification redundancy. Trypsin was chosen as the enzyme parameter. One missed cleavage was allowed and initial peptide mass tolerances of ± 10.0 ppm and ±0.05 Da were used for MS and MS/MS analyses, respectively. Cysteines were assumed to be carbamidomethylated, proline and lysine to be hydroxylated, and methionine was allowed to be oxidized. All these modifications were set to be variable. The monoisotopic peptide charge was set to 1^+, 2^+, and 3^+. Only significant hits (MASCOT score ≥80 for proteins) were accepted. Peptide Decoy option was selected during the data-search process to remove false-positive results. Peptide sequences from the significant protein hits were submitted to MS-driven BLAST (MS-BLAST) search against the NCBInr database on a web-accessible server using default settings (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Plukenetieae was chosen as the optional organism parameter.

Results

Determination of Protein Content of *P. volubilis* Seeds

Table 2 and 3 summarize the Kjeldahl analysis results of crude protein content and the content of protein fractions in individual and bulked samples, expressed as g.100 g^{-1} in dry matter. Tukey HSD test confirmed the significant statistical difference between tested populations at $p \leq 0.05$. Small differences were confirmed by calculated standard deviations (SD) which corresponded with the variation of

Table 2: Summarization of Sacha inchi protein and protein fraction content of seeds from individual plants

No. of localities	Locality	No. of samples	Total protein *	Albumin-globulin *	Prolamin *	Glutelin *
				($g.100g^{-1}$)		
1	ADO	20	24.55 ± 0.92^{d}	17.42 ± 1.41^{cd}	0.34 ± 0.05^{abc}	6.79 ± 1.71^{a}
2	SRC	23	23.27 ± 1.10^{ab}	16.35 ± 0.97^{ac}	0.36 ± 0.10^{abc}	6.56 ± 1.35^{a}
3	MIS	20	23.67 ± 1.13^{bd}	16.63 ± 1.49^{acd}	0.40 ± 0.08^{bc}	6.65 ± 1.25^{a}
4	CHU	20	22.84 ± 1.42^{ab}	17.46 ± 1.44^{cd}	0.24 ± 0.11^{d}	5.14 ± 1.67^{bc}
5	AUC	18	21.33 ± 1.05^{c}	15.40 ± 0.91^{ab}	0.29 ± 0.13^{ad}	5.64 ± 0.84^{abc}
6	PAC	20	22.74 ± 0.97^{ab}	17.86 ± 0.99^{d}	0.29 ± 0.07^{ad}	4.59 ± 1.54^{c}
7	RAC	5	22.97 ± 1.75^{abcd}	15.04 ± 2.49^{ab}	0.42 ± 0.05^{abc}	7.52 ± 1.30^{a}
8	SLU	2	20.41 ± 1.82^{abcde}	13.93 ± 4.44^{ab}	0.53 ± 0.21^{c}	7.95 ± 2.41^{ab}
9	PUC	21	19.78 ± 1.20^{e}	14.54 ± 1.22^{b}	$0.35 \pm 0.06^{ab}c$	4.88 ± 1.40^{bc}
10	2DM	22	22.20 ± 1.47^{ac}	15.91 ± 1.16^{a}	0.31 ± 0.06^{abd}	5.98 ± 1.33^{ab}
	Total no.	171				
	Mean ± SD		22.56 ± 1.79	16.37 ± 1.68	0.33 ± 0.10	5.87 ± 1.63

*****Different letters in columns denote statistically significant. SD –standard deviation

Table 3: Sumarization of Sacha inchi protein and protein fraction content of seeds from bulked samples

No. of localities	Locality	Total protein	Albumin- globulin	Prolamin	Glutelin
			($g.100g^{-1}$)		
1	ADO	23.56	16.62	0.55	6.39
2	SRC	21.92	15.30	0.39	6.23
3	MIS	23.99	16.38	0.33	7.28
4	CHU	23.8	17.8	0.32	5.67
5	AUC	20.98	14.53	0.19	6.25
6	PAC	21.17	14.71	0.19	6.27
7	RAC	22.23	15.46	0.29	6.47
8	SLU	21.42	14.50	0.51	6.41
9	PUC	19.82	15.62	0.22	3.97
10	2DM	22.66	16.77	0.16	5.73
	Mean	22.082	15.70	0.32	6.7

protein content in each tested population. Mean values of total protein content for all samples were 22.56 g.100 g^{-1} with oscillation between obtained minimum (19.78 g.100 g^{-1}) and maximum (24.55 g.100 g^{-1}) in individual samples. Mean values of bulked samples were only slightly different (22.08 g.100 g^{-1}). ADO population was observed as the population with the highest content of crude protein (mean value 24.55 g.100 g^{-1}). On the other hand, PUC population demonstrated the lowest content of crude protein (mean value 19.78 g.100 g^{-1}). Albumins and globulins presented the major protein fractions; the mean value for all tested samples was 16.37 g.100 g^{-1} with oscillation 1.68 g.100 g^{-1}. Similarly, the content of albumin and globulin fractions in the bulked samples reached ca. 15.70 g.100 g^{-1}. The variability in tested bulked samples was in the range of 14.50-17.80 g.100 g^{-1}. Prolamins were observed as the smallest part of proteins in all tested samples and bulked samples. The mean value of prolamins was 0.33 g.100 g^{-1} in individual Sacha inchi samples and in bulked samples the prolamin level was 0.32 g.100 g^{-1}. Mean value of glutelin fraction for individual samples was calculated as 5.87 g.100 g^{-1} with oscillation 1.63 g.100 g^{-1} in comparison with bulked samples the presence of this fraction was observed slightly higher (6.70 g.100 g^{-1}).

SDS-PAGE and 2-DE Analysis

Total seed proteins isolated from bulked samples of all ten investigated *P. volubilis* populations were also analyzed by SDS-PAGE (Fig. 2). Many protein bands were observed in the wide range of molecular weight 8-75 kDa, most of them were common to the different genotypes, indicating their close relationship with one another. Four major and relatively abundant bands were detected in the following positions: 20.1, 30.8, 32.2 and 36.4 kDa (bands 3, 4, 5 and 8 in Fig. 2) and visible in the SDS-PAGE profile of all *P. volubilis* populations. No differences in intensities between these four most abundant protein bands were detected. However, comparison of other less abundant protein bands revealed some variations between the bulked samples. Some of these less abundant protein bands were unique to the particular populations. 2DM population exhibited the most different profile with one intensive protein band with MW ca. 25.8 kDa (band 6, Fig. 2). This band could also be observed in the profiles of some other populations (SLU, RAC, PUC, MIS and CHU), but as a relatively less abundant band. Moreover, the band was almost invisible in profiles of AUC, PAC, and SCR populations. Similarly, further inter-population variations could be observed in the region of MW ca 40-70 kDa (excision 1 and 2, Fig. 2). All total protein extracts from mature seeds were further analyzed by 2-DE. Intra-population variations within the framework of 2DM population were also studied by SDS-PAGE (Appendix Fig. 1). No significant differences between individual samples of one distinct 2DM population were observed.

According to results of the genetic analyses (Ocelák *et al.*, 2015; Vašek *et al.*, 2017), the most genetically diverse population (2DM) and the population (RAC) were chosen as representatives of bulked samples for further analyses and were fractionated by Osborne´s method. Resulting protein fractions (albumin + globulin, prolamin and glutelin) were electrophoretically characterised by SDS-PAGE (Fig. 3). Albumins and globulins formed a large portion of Sacha inchi proteins. Many abundant bands were detected in the whole MW range, mainly in positions from ca. 22 to 36 kDa and several bands in the region of ca. 8-19 kDa. Glutelin fraction was less abundant with many bands detected at ca. 22-36 kDa and one band at ca. 70 kDa. The samples were very low in prolamin fraction, four protein bands with molecular weight from 22 to 36 kDa were found on gel (Table 2 and 3). Protein diversity between individual samples for each distinct population was also studied, but polymorphism in band abundances and positions was generally found low between all tested individual and bulked samples in 2DM (Supplemental Information Fig. 1) and all other populations (data not shown).

A 2-DE protein map of each population was performed using a broad range IPG strip to allow simultaneous separation and visualisation of both acidic and basic proteins, rendering a reference map of *P. volubilis* seed proteins. The Coomassie blue stained 2-DE protein map of each population revealed ca 180±20 protein spots through PDQuest image analysis software, and all 2-DE protein maps were dominated by ca. eight highly abundant proteins clustered in two regions of isoelectric point (p*I*) ca 5-6 and 8-9 with molecular masses of 36-22 kDa, which accounted for more than 80% of total protein mass in the seeds, as revealed by 2-DE (Fig. 4). The pI of protein spots in 2-DE were mainly distributed in neutral and basic regions. The 2-DE protein maps of all populations were relatively similar in major spots, but some distinct differences were visible, mainly when comparing 2DM and RAC populations. Based on the comparison of 2-DE protein maps of these 2DM and RAC populations, a minimum of three spots showed population-dependent qualitative changes and more protein spots showed little quantitative changes, respectively. Qualitatively changed spots are marked with a dotted circle (Fig. 4). The remaining 2-DE maps of the relatively less different eight populations are shown in Appendix Fig. 2. These eight 2-DE protein maps did not reveal any further population-specific spots distinct from the 2DM and RAC populations.

Identification of Proteins

The protein bands and spots of interest from SDS-PAGE and 2-DE gels (totally 40) were selected to be excised and subjected to in-gel digestion using trypsin. The digested peptide mixture was then analyzed using nLC-MS/MS. The MS/MS results of tryptic peptides were queried against the NCBInr database via ProteinScape software. Out of 40 protein samples 18 were "positively" identified to contain valid proteins (Table 4), that is the MASCOT score of the identified protein was higher than the significant threshold (≥80). Numbering of the "positive matches" shown in Table 4 is in accordance with numbering bands and spots in Fig. 2 and 4. Six 2DM, AUC, and RAC populations were subjected to Osborne´s fractionation to obtain a total of

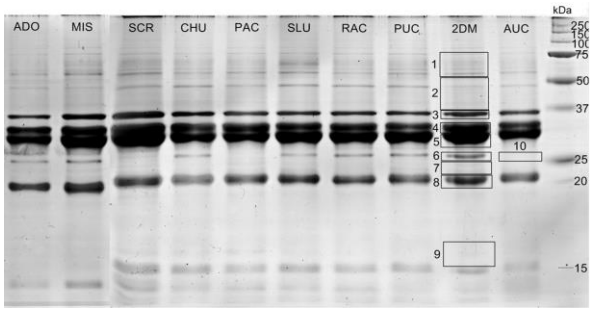

Fig. 2: SDS-PAGE (15% homogeneous gel) of total seed proteins in deffated flour from different *P. volubilis* populations. Protein load, except for the standards, was 40 μg each. Information about the populations corresponding to Table 1

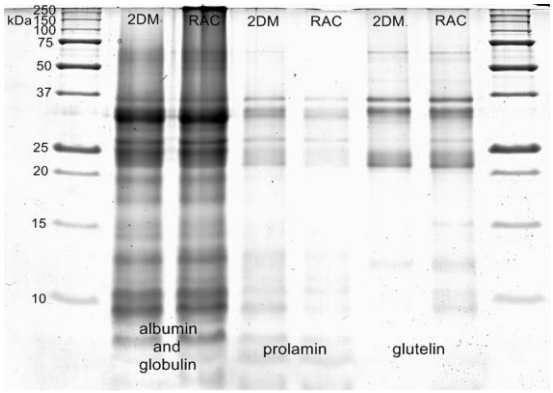

Fig. 3: SDS-PAGE (10% homogeneous gel) of total seed proteins and two protein fractions of *Plukenetia volubilis* population in SDS-PAGE condition. Protein load, except for the standards, was 40 μg each. Information about the populations corresponding to Table 1

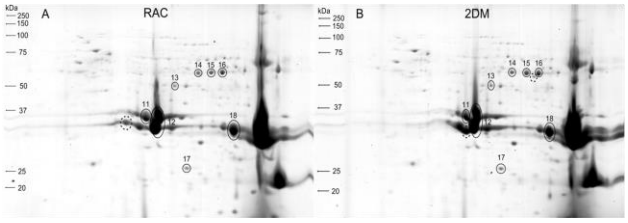

Fig. 4: Two-dimensional gel electrophoresis of storage seed protein of *P. volubilis* deffated flour from different populations. The samples (105 μg proteins) were separated on a 7 cm IPG strip with a non-linear pH gradient (3-10) followed by SDS-PAGE electrophoresis using a 12.5% homogeneous gel. (**A**) Ramón de Castillo (RAC) locality in Peruvian Amazon; (**B**) Dos de Mayo (2DM) locality in Peruvian Amazon. Spots analyzed by nLC-MS/MS are numbered; Difference spots between the population RAC (A) and 2DM (B) are marked with a dotted circle

Table 4: List of identified proteins and peptides detected in Dos de Mayo (2DM) sample by SDS-PAGE and two-dimensional gel electrophoresis and identified by mass spectrometry. Bands and spots in the table column are numbered in accordance with Fig. 3 and 4. *Amino acid modification detected in the peptides: carbamidomethyl (C); oxidation (M); hydroxylation (K, P)

B and/ spot no.	NCBI accession no.	Protein identification	MW (teor.)	pI (teor)	Mascot score	Sequence coverage [%]	Peptides sequence
1	EEF36692.1	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase, putative [*Ricinus communis*]	84.7	6.1	181.0	5.4	R.IPSTEEIADR.I; K.FALESFWDGK.S; K.LNLPILPTTTIGSFPQTIELR.R
2	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	283.7	8.6	R.VTSVNSHNLPILR.Y; R.GLLLPQYVNGPK.L; R.INAVEPSRR.F; R.INAVEPSR.R K.LIYVVQGR.G
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	229.8	6.8	R.VTSVNSHNLPILR.Y; R.GLLLPQYVNGPK.L; R.INAVEPSRR.F; R.INAVEPSR.R
	AFJ04522.1	legumin B precursor, partial [*Vernicia fordii*]	46.5	6.5	102.7	2.4	R.ADVFVPEVGR.L; R.ENIADPSR.A
	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	57.3	9.2	97.2	1.8	R.LSASHVVLR.N; K.ENIADPSR.A
3	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	239.5	8.0	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E; R.INAVEPSR.R; R.INAVEPSRR.F K.LIYVVQGR.G
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	196.2	6.2	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E; R.INAVEPSR.R; R.INAVEPSRR.F
4	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	54.0	8.9	232.0	4.8	R.KLQSENDFR.G; K.LQSENDFR.G; R.NFYLAGNPEDEFQK.Q; R.NFYLAGNP*EDEFQK*.Q
5	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	54.0	8.9	193.7	4.8	R.NFYLAGNPEDEFQK.Q; R.KLQSENDFR.G; K.LQSENDFR.G
6	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	255.6	8.0	R.DFFLAGNPQR.E; R.GLLLPQYVNGPK.L; R.INAVEPSRR.F; R.INAVEPSR.R K.LIYVVQGR.G
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	210.3	6.2	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E; R.INAVEPSR.R; R.INAVEPSRR.F
7	KDP42006.1	hypothetical protein JCGZ_27024 [*Jatropha curcas*]	24.3	6.1	154.8	10.0	K.LSFLYPASTGR.N; K.VTYPIIADPNR.E
	EEF31012.1	Zeamatin precursor, putative [*Ricinus communis*]	27.4	9.5	107.1	8.3	R.YPAGHGPVVAC*K.S; K.GVC*PVVGC*R.A
	EEF43161.1	Peroxiredoxin, putative [*Ricinus communis*]	24.2	5.6	95.8	3.2	K.LSFLYPASTGR.N; R.VVESLQR.A
8	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	133.1	4.5	R.DFFLAGNPQR.E; R.GLLLPQYVNGPK.L
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	133.1	4.4	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E
9	AFR46653.1	oleosin2 [*Plukenetia volubilis*]	15.8	11.1	218.3	20.4	R.LTQQHPFR.E; R.GSLPMQMDQAK.R; R.GSLPM*QM*DQAK.R R.MQETTGQFGQK.A; R.M*QETTGQFGQK.A
10	EEF31012.1	Zeamatin precursor, putative [*Ricinus communis*]	27.4	9.5	122.0	8.3	R.YPAGHGPVVAC*K.S; K.GVC*PVVGC*R.A
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	108.7	1.8	R.INAVEPSRR.F; R.INAVEPSR.R
	AFJ04522.1	legumin B precursor, partial [*Vernicia fordii*]	46.5	6.5	95.2	4.3	R.ENIADPSR.A; R.ADVFVPEVGR.L
11	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	255.7	8.0	R.DFFLAGNPQR.E; R.GLLLPQYVNGPK.L; R.INAVEPSRR.F; R.INAVEPSR.R K.LIYVVQGR.G
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	215.6	6.2	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E; R.INAVEPSR.R; R.INAVEPSRR.F
12	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	308.1	9.2	R.DFFLAGNPQR.E; R.GLLLPQYVNGPK.L; R.INAVEPSRR.F; R.INAVEPSR.R K.LIYVVQGR.G; R.EEEAQR.Q
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	260.6	7.4	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E; R.INAVEPSRR.F; R.INAVEPSR.R; R.EEEAQR.Q
	EEF39577.1	11S globulin subunit beta precursor, putative [*Ricinus communis*]	43.6	9.5	149.5	6.0	R.GIIVSVEHDLEMLAPPR.S;R.GIIVSVEHDLEML APPRSQEEER.E; R.EEEAQR.Q
13	XP_015583443.1	PREDICTED: legumin B [*Ricinus communis*]	55.2	8.9	245.1	8.0	R.DFFLAGNPQR.E; R.GLLLPQYVNGPK.L; R.INAVEPSRR.F; R.INAVEPSR.R K.LIYVVQGR.G
	EEF28918.1	glutelin type-A 3 precursor, putative [*Ricinus communis*]	55.9	8.9	203.7	6.2	R.GLLLPQYVNGPK.L; R.DFFLAGNPQR.E; R.INAVEPSRR.F; R.INAVEPSR.R
14	EEF38210.1	legumin B precursor, putative [*Ricinus communis*]	53.6	9.7	142.1	4.4	R.VSTVNSHNLPILR.W; K.ENIADPSR.S
	AFJ04522.1	legumin B precursor, partial [*Vernicia fordii*]	46.5	6.5	106.8	2.4	R.ADVFVPEVGR.L; R.ENIADPSR.A
	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	54.0	8.9	101.4	1.9	R.LSASHVVLR.N; K.ENIADPSR.A
15	AFJ04522.1	legumin B precursor, partial [*Vernicia fordii*]	46.5	6.5	223.4	4.3	R.ENIADPSRADVFVPEVGR.L; R.ADVFVPEVGR.L; R.ENIADPSR.A
	EEF38210.1	legumin B precursor, putative [*Ricinus communis*]	53.6	9.7	143.1	2.7	R.VSTVNSHNLPILR.W; K.ENIADPSR.S
	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	54.0	8.9	103.6	1.9	R.LSASHVVLR.N; K.ENIADPSR.A
16	AFJ04522.1	legumin B precursor, partial [*Vernicia fordii*]	46.5	6.5	232.2	4.3	R.ENIADPSRADVFVPEVGR.L; R.ADVFVPEVGR.L; R.ENIADPSR.A
	gi|255567250	legumin A precursor, putative [*Ricinus communis*]	54.0	8.9	112.7	1.9	R.LSASHVVLR.N; K.ENIADPSR.A
	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	51.1	6.4	108.3	2.0	R.LDALEPDNR.I; K.ENIADPSR.A
17	KDP42006.1	hypothetical protein JCGZ_27024 [*Jatropha curcas*]	24.3	6.1	155.2	10.0	K.LSFLYPASTGR.N; K.VTYPIIADPNR.E
	EEF43161.1	Peroxiredoxin, putative [*Ricinus communis*]	24.2	5.6	137.1	7.8	K.LSFLYPASTGR.N; K.LLGLSC*DDVLSHVEWIK.D
18	EEF37814.1	legumin A precursor, putative [*Ricinus communis*]	54.0	8.9	177.4	4.8	R.KLQSENDFR.G; K.LQSENDFR.G; R.NFYLAGNP*EDEFQK*.Q
	EEF38210.1	legumin B precursor, putative [*Ricinus communis*]	53.6	9.7	99.3	2.7	R.DQFQC*AGVAVVRR.R;R.DQFQC*AGVAVVRR.T

eighteen albumin + globulin, prolamin, and glutelin fraction samples (each population afforded three fractions). The samples were subjected to direct digestion by trypsin and the resulting mixtures were analyzed by LC-MS/MS. Analysis of these 18 protein samples revealed 62 valid proteins together, six of them were identical with proteins in Table 4 and indicated in bold (Appendix Table 1).

Discussion

During the past decade, the development of nutritionally balanced protein foods to feed a growing population has received an increasing attention from food scientists and nutritionists. Research directions concerning evaluation of plant proteins in human foodstuffs have increased considerably (Day, 2013). Seeds of *P. volubilis* are an important and traditional food in the Amazon region. Sacha inchi is a valuable source not only of high-quality oil but also has a high protein content, first reported by Hamaker *et al.* (1992). In this study crude protein content in *P. volubilis* seeds collected from the Peruvian Amazon was found in a range between 15.95% and 23.83%. In comparison with results of our study (22.56 g.100 g^{-1}), total protein content the value previously published in Hamaker´s was slightly higher (~27%) (Hamaker *et al*., 1992). This discrepancy may be caused using considerably higher number of samples from different localities in our case.

Sacha inchi seeds were characterized by a similar amount of protein as cashew nuts, macadamia, and pistachio (Sathe, 1994; Sathe *et al.*, 2009). Compared with cereals and pseudocereals, the achieved value in Sacha inchi was almost double (Mota *et al.*, 2015). Conversely, the content of crude protein in *P. volubilis* seeds was lower compared with soybean protein content from 38% to 40% (Singh *et al.*, 2007) and it exceeded protein content detected in walnut, hazel nut, brazil nut, and Virginia peanut (Sathe *et al.*, 2009).

Only limited information is available on Osborne fractionation of Sacha inchi seed protein. However, according to our results, albumin and globulin (mean value 16.37 g.100 g^{-1}) were observed as the predominant fractions, followed by glutelin fraction (5.87 g.100 g^{-1}), with the presence of a very small portion of prolamins (0.33 g.100 g^{-1}). Due to the very low content of prolamins, Sacha inchi seeds are interesting from a nutritional point of view as a promising gluten-free nutrient. In comparison with those present in *P. volubilis*, *Jatropha curcas* seed storage protein was similarly characterized with the most abundant fractions of albumins (12%) together with globulins (44%), followed by glutelins (40%) and minor fractions of prolamins (3.4%), which is a somewhat higher content than prolamins in Sacha inchi (0.33 g.100 g^{-1}) (Martinez-Herrera *et al.*, 2004). The flour of *Caryodendron orinocense* (from Euphorbiaceae family) also has albumins and globulins as the major protein fractions, whereas the prolamin fraction is significantly higher (23.10%) compared to Sacha inchi (Padilla *et al.*, 2010).

Revealing differences in proteomes between the various populations is slightly difficult. The previously mentioned genetic studies of our team based on ISSR (Ocelák *et al.*, 2015) and AFLP markers (Vašek *et al.*, 2017) showed 2DM population as the most genetically diverse population. RAC population was shown as the typical representative among the other populations compared with the 2DM population. However, it did not reveal significant differences between individual samples within 2DM population by SDS-PAGE method (Supplemental Information Fig. 1). Variability was detected only in a few minor band positions. The strategy for cross-pollinated plants (Gardiner and Forde, 1992) using bulked samples for SDS-PAGE was applied. Therefore, bulked samples of ground seed mixture were prepared and were used in the SDS-PAGE (Fig. 2) and 2-DE analysis (Fig. 4 and Supplemental Information Fig. 2). Different loads of protein mixture were tested (data not shown) and the best results were obtained when 40 μg of protein was applied onto each lane. All populations showed a relatively constant quantity of four most abundant proteins (bands 3, 4, 5 and 8). More visible differences between the populations were apparent on less abundant proteins (bands 1, 6 *vs*. 10).

Two selected populations with the highest observed protein diversity (2DM and RAC) were also compared by 2-DE (Fig. 4) thus additional information about the inter-population differences could be obtained. This method allowed us to generate a complete proteome map of *P. volubilis* seeds.

In this case, more differences between the two populations were revealed than in SDS-PAGE gels (spots marked with dotted circle in Fig. 4) because 2-DE allows protein separation not only by molecular weight, but also by pI values. Some spots were present in 2DM population, but not in RAC population and vice versa.

Molecular data from inter simple sequence repeat (ISSR) analysis showed differences between the ten investigated *P. volubilis* populations where 2DM population was among the four most distinguished ones (Ocelák *et al.*, 2015). Moreover, these results support data of AFLP markers (Vašek *et al.*, 2017) performed with the same plant material (10 populations of *P. volubilis*) where population 2DM and RAC were among four populations showing a high level of differentiation. Comparing the results based on molecular data with protein polymorphism, it can probably be assumed that the degree of protein polymorphism will always be different (Aliyu and Awopetu, 2007).

The number of protein spots in 2-DE of our samples was significantly higher than observed by Sathe *et al.* (2012) where especially low abundant proteins were not detected. Good SDS-PAGE and 2-DE separation is important for proper evaluation of differences between the *P. volubilis* populations.

Mass spectrometric identification of bands and spots from gels was carried out by using nLC-MS/MS followed by searching the proteome database by MASCOT and MS-

BLAST homology searches. The identification encountered a lack of *P. volubilis* genome and protein sequences in the public databases as reference sequences to search against. More identified peptides were joined to other plants belonging to the Euphorbiaceae family such, as *J. curcas*, *V. fordii* and *R. communis*. The database of genomic expressed sequence tags can serve as an additional source of data. One recent study investigated gene expression in *P. volubilis* by real-time quantitative PCR (Niu *et al.*, 2015). Under these conditions, cross-species protein identification using mass spectral data has been demonstrated to be an alternative in the absence of annotated genome sequence. This is the most reliable way to search for closely related species and/or highly conserved proteins. Only the successful matches of peptides to proteins of the genetically close plant species (members of Euphorbiaceae family) are shown (Table 4).

Proteins from *P. volubilis* seeds detected in this study cover a broad scale of their roles in the plant. The majority of bands in SDS-PAGE gels and spots in 2-DE gels (bands 2, 3, 6, 8 and spots 11, 12, 13) revealed peptides from proteins, mostly related to nutrient reservoir activity, such as glutelin type-A 3 precursor-putative (*Ricinus communis*) protein and some legumin proteins (legumin A precursor-putative (*R. communis*), predicted legumin B (*R. communis*) and legumin B precursor-partial (*Vernicia fordii*)). Both species (*R. communis* and *V. fordii*) are phylogenetically closely related species to *P. volubilis*, all being members of Euphorbiaceae family. This may explain the presence of peptides in *P. volubilis* seed samples with amino acid sequence identical to *R. communis* and *V. fordii*. In the previous study, *P. volubilis* seed proteins, 2-DE analysis of the alkali glutelin fraction showed the presence of several spots with molecular weight ca. 34 kDa (Sathe *et al.*, 2012). This observation (spots 11 and 12) was confirmed; moreover, a distinct glutelin spot (spot 13) with MW ca. 50.8 kDa and pI ca 7.1 was found. Glutelins are part of a large and extensive group of seed storage proteins most abundant in monocotyledon seeds (Miernyk and Hajduch, 2011). Depending on the pH of the extraction solution, these are usually broken up into two groups: acid glutelins and alkali glutelins. Alkali glutelin fraction in *P. volubilis* seeds was first investigated in Sathe *et al.* (2012). The glutelin fraction of Sacha inchi samples was also analysed (Fig. 2), peptides related to glutelin type-A 3 precursor and putative (*R. communis*) were detected (Table 4).

Several peptides belonging to variants of legumin proteins from *R. communis* and *V. fordii* were detected (bands 2, 4, 5, and spots 14, 15, 16 and 18). Legumin in castor bean seeds (*R. communis*) has recently intensively studied by 2-DE followed by matrix assisted laser desorption ionization-time of flight (MALDI-TOF/TOF) mass spectrometry detection (Campos *et al.*, 2010). The authors identified and investigated several classes of seed reserve proteins such as 2S albumins, legumin-like, legumin B precursor, legumin A2, 11S globulin, 7S globulin precursor, and many other seed storage proteins (Campos *et al.*, 2010). Not surprisingly, legumins from *R. communis* and *J. curcas* were also detected in Sacha inchi seeds in this study, which indicates that their molecular function in *P. volubilis* seeds remains the same or very similar. 11S globulin subunit beta precursor, putative from *R. communis* was detected in one of the largest 2-DE largest spots (spot 12) at MW ca 34.8 kDa and pI ca. 6.5. This is a somewhat different observation in contrast to Sathe *et al*. (2012), where the globulin protein fraction was found in 2-DE gel as a slightly more basic protein than observed in this study.

Zeamatin precursor, putative (*R. communis*), found in bands 7 and 10 with MW ca. 25.9 kDa is another minority protein found here. Our detection of zeamatin peptides in *P. volubilis* seed protein extract with amino acid sequences analogous to zeamatin putative precursor from *R. communis* is interesting, but not surprising. Zeamatin is a corn seed antifungal protein. It is part of a large family of thaumatid-like proteins and had been isolated from numerous plants. Various plants produce antifungal proteins to fight against fungal assault and these proteins can be detected in various plant tissues such as seeds, bulbs, leaves, tubers, fruits, shoots, and roots (Yan *et al.*, 2015).

The peptides belonging to 5-methyltetrahydropteroyl-triglutamate-homocysteine methyltransferase, putative (*R. communis*) in band 1 were also detected. 5-Methyltetrahydropteroyltriglutamate-homocysteine methyltransferase is a plant enzyme and involved in cysteine and methionine formations, a common metabolic pathway present throughout various plant species (Ravanel *et al.*, 2004). Therefore, detection of this enzyme in *P. volubilis* seeds could be expected and its role in the seeds is worth further research.

A number of peptides with high analogy to peroxiredoxin, putative from *R. communis*, and hypothetical protein JCGZ_27024 from *Jatropha curcas* were detected in band 7 and also in one distinct 2-DE spot 17. MW and pI values of this protein spot (spot 17 in Fig. 4) from *P. volubilis* are in close accordance with the published values for peroxiredoxin from *R. communis* and/or *J. curcas*, respectively. Peroxiredoxin is a novel antioxidant enzyme reducing phospholipid hydroperoxides and playing an important role in cellular defence mechanisms against oxidant stress (Manevich *et al.*, 2002). 1-Cys peroxiredoxin has peroxidase activity when coupled to the thioredoxin system. In cereal seeds, the thioredoxin system acts in the developing phase, controlling delivery of compounds during seed filling. Zhang *et al.* (2015) observed significant accumulation of 1-Cys peroxiredoxin at desiccation/maturation developmental stages of bread wheat (*Triticum aestivum* L.) during grain development. The function of 1-Cys peroxiredoxin in connection to *P. volubilis* seeds remains unclear.

The only proteins belonging to *P. volubilis* plant found in this study were elongation, factor 1-alpha, oleosin 2 and oleosin 3 (Table 4 and Supplemental Information Table 1). Elongation factor 1-alpha is a protein involved in protein

biosynthesis (Niu *et al.*, 2015). Oleosins are low molecular mass storage proteins (15-26 kDa) in oilseeds and the genes encoding oleosin proteins are usually specifically expressed in seeds (Simkin *et al.*, 2006). This was also been the first study where protein oleosin 2 was detected by 2-DE and MS/MS in real samples of *P. volubilis* seed extract. Oleosin 2 was detected as five peptides belonging to the band 9 in SDS-PAGE gel (Fig. 3). The observed MW of this band is in close accordance with the theoretically published MW (15.8 kDa) for this protein. Unfortunately, Oleosin 2 was not directly detected as a distinct spot in the 2-DE gel because it is a basic protein with pI = 11.1. That pI value is out of the pH range of the used IPG strips used (they have pH range between 3 and 10). Oleosin 2 was also detected as sixteen distinct peptides in glutelin fraction after Osborne´s fractionation of *P. volubilis* seed flour. These peptides cover ca. 36.7% of Oleosin 2 amino acid sequence. Moreover, Oleosin 3 was detected in the glutelin fraction as one distinct peptide with R.TLGAESLDEAR.L sequence. One recent study describes isolation and characterization of two *P. volubilis* oleosin genes. Using yeasts as heterologous expression system, the effective role of two oleosin genes in yeast lipid body accumulation was investigated, which can be helpful for their potential applications in bioengineering in the future (Chandrasekaran *et al.*, 2013) In this study, the presence of oleosin 2 in *P. volubilis* seed flour has been directly described for the first time (Fig. 3 and Table 4). It can be expected that more *P. volubilis* proteins will be discovered in the future, when its genome is fully sequenced and available to the public (Niu *et al.*, 2015).

Detection of the five proteins: Oleosin 2 from *P. volubilis*, glutelin type-A 3 precursor; putative from *R. communis*; Legumin A precursor, putative from *R. communis*; predicted Legumin B precursor, partial from *R. communis*; Legumin B, precursor from *V. fordii* and hypothetical protein JCGZ_27024 from *J. curcas* (Table 4 and Appendix Table 1) in the samples processed by two independent and different ways contributes to our predication of their presence in the *P. volubilis* seed flour.

Conclusion

This research on the protein profile of *P. volubilis* seeds may serve as a starting point to better understand the promising properties of *P. volubilis* seed proteins. The results showed that chemical characteristics and proximate composition of Sacha inchi seeds of different origin available in the Amazonia region are quite variable and mainly dependant on their geographical origin. In contrast, protein characteristics were quite similar among seeds from one locality. In this respect, Sacha inchi seeds should be selected accordingly for derived protein and oil products intended for manufacture by the food and pharmaceutical industries, as well as applications in bioengineering, in order to optimize the production processing and the chemical characteristics of Sacha inchi seed products.

Acknowledgements

This research was supported financially by the projects CIGA (no. 20144207) and IGA (no. 20185015) of the Czech University of Life Sciences Prague and the Ministry of Agriculture of the Czech Republic (no. RO0418). This work was also supported by the Czech Science Foundation (no. 15-01948S), and with support for long-term conceptual development of research organization (RVO:67985823). We also thank to Ing. Martin Ocelák for collection of Sacha inchi seeds and to Ing. Anna Prohasková for help with analysis of protein and protein fraction content.

References

Aliyu, O.M. and J.A. Awopetu, 2007. Assessment of genetic diversity in three populations of cashew (*Anacardium occidentale* L.) using protein-isoenzyme-electrophoretic analysis. *Genet. Resour. Crop Evol*., 54: 1489–1497

Arévalo, G., 1995. *El cultivo del Sacha inchi (Plukenetia volubilis L.) en la Amazonía. Programa Nacional de Investigación en Recursos Genéticos y Biotecnología – PRONARGEB*, p: 21. Estación Experimental El Porvenir – Tarapoto, Perú

Bejosano, F.P. and H. Corke, 1999. Effect of *Amaranthus* and buckwheat proteins on the rheological properties of maize starch. *Food Chem*., 65: 493–501

Campos, F.A., F.C. Nogueira, K.C. Cardoso, G.C. Costa, L.E.D. Bem, G.B. Domont, M.J.D. Silva, R.C. Moreira, A.A. Soares and T.L. Juca, 2010. Proteome analysis of castor bean seeds. *Pure Appl. Chem.*, 82: 259–267

Chandrasekaran, U., X. Wang and A. Liu, 2013. Characterization, expression profiling and heterologous function analysis of two oleosin genes PvOle 1 and PvOle2 from Sacha inchi (*Plukenetia volubilis*). *Intl. J. Agric. Biol.*, 15: 435–442

Chirinos, R., O. Necochea, R. Pedreschi and D. Campos, 2016. Sacha inchi (*Plukenetia volubilis* L.) shell: an alternative source of phenolic compounds and antioxidants. *Intl. J. Food Sci. Technol*., 51: 986–993

Czech State Norm (CSN) EN ISO 662, 2001. Živočišné a rostlinné tuky a oleje – Stanovení vlhkosti a těkavých látek (in Czech)

Czech State Norm (CSN) 5983-1 (467035), 2012. Krmiva - Stanovení obsahu dusíku a výpočet obsahu hrubého proteinu - Část 1: Kjeldahlova metoda (in Czech)

Day, L., 2013. Protein from land plants - Potential resources for human nutrition and food security. *Trends Food Sci. Technol*., 32: 25–42

Dvořáček, V., J. Moudrý and V. Čurn, 2001. Studies of protein fraction in grain of spelt wheat (*Triticum spelta* L.) and common wheat (*Triticum aestivum* L.). *Sci. Agric. Bohem.*, 32: 287–305

Džunková, M., D. Janovská, P.H. Čepková, A. Prohasková and M. Kolář, 2011. Glutelin protein fraction as a tool for clear identification of Amaranth accessions. *J. Cereal Sci*., 53: 198–205

Eckhardt, A., M. Jágr, S. Pataridis and I. Mikšík, 2014. Proteomic analysis of human tooth pulp: Proteomics of human tooth. *J. Endodont*., 40: 1961–1966

FAO, 2002. *Food Energy – methods of Analysis and Conversion factors. Fao, Food and Nutrition Paper 77*. Report of a technical workshop, Rome, 3–6 December. Food and Agriculture Organization of the United Nations. ISSN 0254–4725

Gardiner, S.E. and M.B. Forde, 1992. Identification of cultivars of grasses and forage legumes by SDS-PAGE of seed proteins. *In: Strategies for Cultivar of Cross-fertilized and Self-fertilized Species, in Seed Analysis*, pp: 43–61. Linskens, H.F. and J.F. Jackson (eds.), Springer Verlag: Berlin, Germany

Gillespie, L.J., 1994. Pollen morphology and phylogeny of the tribe Plukenetieae (Euphorbiaceae). *Ann. Missour. Bot. Gard*., 81: 317–347

Gutiérrez, L.F., L.M. Rosada and Á. Jiménez, 2011. Chemical composition of Sacha Inchi (*Plukenetia volubilis* L.) seeds and characteristics of their lipid fraction. *Gras. Aceit.*, 62: 76–83

Hamaker, B.R., C. Valles, R. Gilman, R.M. Hardmeier, D. Clark, H.H. García, A.E. Gonzales, I. Kohlstad and M. Castro, 1992. Amino acid and fatty acid profiles of the Inca peanut (*Plukenetia volubilis* L.). *Cereal Chem.*, 69: 461–463

Jágr, M., A. Eckhardt, S. Pataridis and I. Mikšík, 2012. Comprehensive proteomic analysis of human dentin. *Eur. J. Oral Sci*., 120: 259–268

Khan, A.H., N. Khan, N.M. Minhas, M.A. Rabbani and A. Ghafoor, 2014. Diversity in seed storage proteins in maize genetic resources: I. variation in alcohol soluble zein protein fraction. *Intl. J. Agric. Biol.*, 16: 1015–1018

Magdeldin, S., S. Enay, Y. Yoshida, B. Xu, Y. Zhang, Z. Zureena, I. Lokamani, E. Yaoita and T. Yamamoto, 2014. Basics and recent advances of two dimensional-polyacrylamide gel electrophoresis. *Clin. Proteom.*, 11: 16

Manevich, Y., T. Sweitzer, J.H. Pak, S.I. Feinstein, V. Muzykantov and A.B. Fisher, 2002. 1-Cys peroxiredoxin overexpression protects cells against phospholipid peroxidation-mediated membrane damage. *Proc. Natl. Acad. Sci.*, 99: 11599–11604

Martinez-Herrera, J., L. Chel-Guerrero and A.L. Martinez-Ayala, 2004. The Nutritional Potential of Mexican Pinon (Jatropha curcas), Toxic and antinutritional factors. *In: Recent Advances of Research in Antinutritional Factors in Legume Seeds and Oilseeds*, pp: 185–188. Muzquiz, M., G.D. Hill, C. Cuadrado, M.M. Pedrosa and C. Burbano (eds.). EAAP-European Association for Animal Production Publication. Toledo, Spain

Miernyk, J.A. and M. Hajduch, 2011. Seed proteomics. *J. Proteom.*, 74: 389–400

Mota, C., M. Santos, R. Mauro, N. Samman, A.S. Matos, D. Torres and I. Castanheira, 2015. Protein content and amino acids profile of pseudocereals. *Food Chem*., 193: 55–61

Niu, L., Y.B. Tao, M.S. Chen, Q. Fu, C. Li, Y. Dong, X. Wang, H. He and Z.F. Xu, 2015. Selection of reliable reference genes for gene expression studies of a promising oilseed crop, *Plukenetia volubilis*, by real-time quantitative PCR. *Intl. J. Mol. Sci*., 16: 12513–12530

Ocelák, M., P.H. Čepková, I. Viehmannová, Z. Dvořáková, D.C. Huansi and B. Lojka, 2015. Genetic diversity of *Plukenetia volubilis* L. assessed by ISSR markers. *Sci. Agric. Bohem.*, 46: 145–153

Padilla, F., T. Guédez, M.J. Alfaro, M. Regnault, C. Rincón and A.M.C. Rincón, 2010. Fractionation and characterization of soluble proteins from nuez de Barinas (*Caryodendron orinocense* K.) flour. *Rev. Inst. Nac. Hig. Raf. Range.*, 41: 38–42

Ravanel, S., M.A. Block, P. Rippert, S. Jabrin, G. Curien, F. Rébeillé and R. Douce, 2004. Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import S-adenosylmethionine from the cytosol. *J. Biol. Chem*., 21: 22548–22557

Santoni, I. and B. Pizzo, 2013. Evaluation of alternative vegetable proteins as wood adhesives. *Ind. Crop Prod*., 45: 148–154

Sathe, S.K., 1994. Solubilization and electrophoretic characterization of cashew nut (*Anacardium occidentale*) proteins. *Food Chem*., 51: 319–324

Sathe, S.K., H.H. Kshirsagar and G.M. Sharma, 2012. Solubilization, fractionation, and electrophoretic characterization of inca peanut (*Plukenetia volubilis* L.) proteins. *Plant Food Hum. Nutr*., 67: 247–255

Sathe, S.K., M. Venkatachalam, G.M. Sharma, H.H. Kshirsagar, S.S. Teuber and K.H. Roux, 2009. Solubilization and electrophoretic characterization of select edible nut seed proteins. *J. Agric. Food Chem.*, 57: 7846–7856

Singh, R.J., R.L. Nelson and G. Chung, 2007. Soybean (Glycine max (L.) Merr.). Genetic Resources, Chromosome Engineering, and Crop Improvement. *In: Oilseeded Crops*, p: 304. Singh, R.J. (ed.). CRC Press Taylor & Francis Group, Boca Raton, Florida, USA

Shah, M., E.L. Soares, P.C. Carvalho, A.A. Soares, G.B. Domont, F.C.S. Nogueira and F.A.P. Campos, 2015. Proteomic analysis of the endosperm ontogeny of *Jatropha curcas* L. seeds. *J. Proteom. Res*., 14: 2557–2568

Simkin, A.J., T. Qian, V. Caillet, F. Michoux, M.B. Amor, C. Lin and S. Tanksley, 2006. Oleosin gene family of *Coffea canephora*: quantitative expression analysis of five oleosin genes in developing and germinating coffee grain. *J. Plant Physiol*., 163: 691–708

Souza, C.R.B.D., S.P.D. Reis and L.J.C.B. Carvalho, 2015. An overview of protein identification studies in cassava. *Curr. Protein Pept. Sci*., 16: 219–227

Sugano, M., 2006. *Nutritional Implications of Soy. Soy in Health and Disease Prevention*, p: 313. Sugano, M. (ed.). CRC Press Taylor & Francis Group, Boca Raton, Florida, USA

Valdez-Flores, M., L.J. Germán-Báez, R. Gutiérrez-Dorado, S. Medina-Godoy, C. Norzagaray-Valenzuela, S. Hernández-Verdugo, C. Reyes-Moreno and A. Valdez-Ortiz, 2016. Improving bioactivities of *Jatropha curcas* protein hydrolysates by optimizing with response surface methodology the extrusion cooking process. *Ind. Crop Prod*., 85: 353–360

Vašek, J., P. Hlásná Čepková, I. Viehmannová, M. Ocelák, D. Cachique and P. Vejl, 2017. Dealing with AFLP genotyping errors to reveal genetic structure in *Plukenetia volubilis* (Euphorbiaceae) in the Peruvian Amazon. *PloS One*, 12: e0184259

Yan, J., S.S. Yuan, L.L. Jiang, X.J. Ye, T.B. Ng and Z.J. Wu, 2015. Plant antifungal proteins and their applications in agriculture. *Appl. Microbiol. Biotechnol*., 99: 4961–4981

Zhang, N., F. Chen, W. Huo and D. Cui, 2015. Proteomic analysis of middle and late stages of bread wheat (*Triticum aestivum* L.) grain development. *Front. Plant Sci*., 6: 735

(Received 02 August 2018; Accepted 27 September 2018)