

An empirical analysis of Sacha inchi (Plantae: Plukenetia volubilis L) seed proteins and their applications in food and biopharmaceutical industries

Mohammad Harun Ur Rashid (✉ rashid23cu@gmail.com)

Universiti Sains Malaysia <https://orcid.org/0000-0002-0698-6734>

Nor Datiakma Mat Amin

Universiti Sains Malaysia

Erica Kok Jia Yi

Universiti Sains Malaysia

Mohd Nazri Ismail

Universiti Sains Malaysia <https://orcid.org/0000-0002-5878-8190>

Research Article

Keywords: Sacha Inchi, LC-MS/MS, proteomics, nanoparticles, microencapsulation, fusion technology, gene ontology

Posted Date: October 18th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2935638/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Applied Biochemistry and Biotechnology on November 18th, 2023. See the published version at <https://doi.org/10.1007/s12010-023-04783-5>.

Abstract

Sacha Inchi (*Plukenetia volubilis* L) is a plant native to the Amazon rainforest in South America known for its edible seeds, which are rich in lipids, proteins, vitamin E, polyphenols, minerals, and amino acids. Rural communities in developing nations have been using this plant for its health benefits, including as a topical cream for rejuvenating and revitalising skin and as a treatment for muscle pain and rheumatism. Although sacha inchi oil has been applied topically to soften skin, treat skin diseases, and heal wounds, its protein-rich seeds have not yet received proper attention for extensive investigation. Proteins in sacha inchi seeds are generally known to have antioxidant and antifungal activities and are extensively used nowadays in making protein-rich food alternatives worldwide. Notably, large-scale use of seed proteins has begun in nanoparticle and biofusion technologies related to the human health-benefitting sector. To extract and identify their proteins, the current study examined sacha inchi seeds collected from the Malaysian state of Kedah. Our analysis revealed a protein concentration of 73.8 ± 0.002 mg/g of freeze-dried seed flour. Employing liquid chromatography-tandem mass spectrometry (LC-MS/MS) and PEAKS studio analysis, we identified 217 proteins in the seed extract, including 152 with known proteins and 65 unknown proteins. This study marks a significant step towards comprehensively investigating the protein composition of sacha inchi seeds and elucidating their potential applications in the food and biopharmaceutical sectors. Our discoveries not only enhance our knowledge of Sacha inchi's nutritional characteristics but also pave the way for prospective research and innovative advancements in the realms of functional food and health-related domains.

Introduction

Sacha Inchi (*P volubilis* L.), also known as mountain peanut or Incha nut, is an oleaginous plant that belongs to the Euphorbiaceae family (Follegatti-Romero et al. 2009). Sacha inchi is a monoecious plant with triangular leaves to ovate with a truncate to cordate at the base, palmate venation, and bilaminar glands. The fruit capsule of sacha inchi is star-shaped and has a size of 3–5 cm. The colour of the fruit changes from green to blackish brown as it matures. Dark brown oval seeds in the capsules are edible and usually 1.5–2 cm in size (Fu et al. 2014). *Plukenetia volubilis* can be widely found in the tropical rainforest of South America's Amazon area, which includes portions of Peru and northeastern Brazil.

Each chemical constituent varies in different parts of the sacha inchi plant. For example, the seeds consist of lipids (35–60%) (including ω -3, 6, and 9 fatty acids), proteins (25–30%), vitamin E, polyphenols, minerals, and others. The proteins contained in the seeds include essential amino acids such as tryptophan, threonine, cysteine, and tyrosine (Sunan Wang et al., 2018). The raw seeds of sacha inchi contain around 22–30% protein, whereas defatted seeds or press cake remaining after oil extraction contain approximately 53–59% protein. The amount of protein depends on the extraction method and the protein assay used (Hamaker BR et al. 1992; Gutiérrez et al. 2011; Ruiz et al. 2013). Albumins, glutelins, globulins, and prolamins are the major proteins that are primarily soluble (Sathe et al., 2012). Furthermore, the amino acid content of the seeds closely matches dietary guidelines, suggesting that they

might be a beneficial supply of amino acids, especially for malnourished populations (Hamaker BR et al. 1992).

Rural people in developing nations have always utilised plant species as traditional medicines or nutraceuticals due to their ability to provide a suitable replacement for healthcare services. According to a study by Oyeboode et al., around 80% of developing nations rely on conventional medicines for primary health care (Oyeboode et al. 2016). According to a study conducted in San Martín, Peru, it was reported that more than half (67%) of the answers stated that sacha inchi were used for health purposes (Del-Castillo et al. 2019). The mixture of sacha inchi ground seeds and seed oil was used as a topical cream to rejuvenate and revitalise the skin. Besides that, to relieve muscle pain and rheumatism, sacha inchi seed oil has been used to rub on the skin. The oil and roasted sacha inchi seeds have also been consumed to control cholesterol levels and improve cardiovascular and gastrointestinal health (Del-Castillo et al., 2019).

Patents have been issued for cosmetic and medicinal formulations comprising sacha inchi proteins and oils (Moser 2016). When applied to the skin, the extracts provide anti-inflammatory, skin tightening, and anti-ageing effects. Hanssen and Schmitz-Huebsch suggested that using sacha inchi oil, which contains protein from its seed, can help treat diabetes, arthritis, coronary heart disease, and inflammatory skin disorders. However, clinical trials are required to evaluate the efficacy of these health claims (Hanssen & Schmitz-Hübsch, 2011). Moreover, sacha inchi leaves and seeds have antioxidant and anti-proliferative activities against specific cancer cells but are non-toxic to normal cells (Herrera-Calderon et al. 2019; Nascimento et al. 2013). The usage of the beneficial components of sacha inchi oils has also been made easier by recent technologies such as nanoparticles and microencapsulation (Da Silva Soares et al. 2019).

Current research on Sacha inchi seeds mainly concentrates on seed oil's antibacterial, anti-inflammatory, skin tightening, and anti-ageing effects since it is rich in lipids with an unusually high unsaturated fatty acid ratio (accounting for about 85.9–97.5%). Despite the increasing demand for seed oil, large volumes of defatted sacha inchi seed residues containing nutritional protein up to 59.1% dry weight) are produced as a by-product and are frequently discarded (Sunan Wang et al., 2018). To advance the growth of the protein-based sacha inchi industry, it is crucial to discover new uses and values for these defatted sacha inchi seed wastes. Due to its high tryptophan (44 mg/g) and phenylalanine (3 mg/g) content, the World Health Organization (WHO) and Food and Agriculture Organization (FAO) of the United Nations recommend sacha inchi seed protein as a high-quality protein with health benefits for human consumption (Hanssen and Schmitz-Hübsch 2011).

As a result, the sacha inchi seed meal is a viable source of protein for creating tasty, higher-protein snacks. Unfortunately, there is limited research on the protein composition of sacha inchi seeds from various geographical regions. This emphasises the imperative need for proteomic studies on sacha inchi seeds to address the increasing global demand for seed protein, which plays a vital role in the food and

biopharmaceutical industries. In response to this demand, our research aims to extensively investigate the protein composition of sachai inchi seeds and provide detailed bioinformatic analysis.

Materials and Methods

Sample collection

The Sachai inchi seeds utilised in this study were procured from Sungai Petani, Kedah, Malaysia. The collection process was performed by skilled local collectors who sourced the seeds from mature sachai inchi plants. To maintain seed quality, temperature and humidity were kept under controlled conditions during transportation to the laboratory. Upon receipt, the seeds underwent thorough cleaning and sorting to eliminate extraneous materials stored at 4°C for subsequent processing.

Protein extraction

The method used in this study was based on work done by Hooi Xian Lee et al. (Lee et al., 2017) with slight modifications in buffer concentration and centrifugation time. After the sachai inchi seeds were removed from the fruit, the seed kernel was weighed and washed using distilled water to remove impurities. Following overnight freeze-drying in a freeze dryer (Labconco), the seed kernels were pulverised using mortar and pestle in liquid nitrogen, and the powder produced was resuspended in 10 ml of chilled extraction buffer containing 10% (w/v) Trichloroacetic acid (TCA) in acetone with 20 mM dithiothreitol (DTT). Subsequently, the mixture was vortexed for 5 min at 8 °C and stored overnight in the – 20 °C freezer. The next day, the samples were centrifuged at 16,000 g for 30 min at 4°C. Then, the pellet obtained was washed with 20mM DTT in cold acetone and centrifuged again for 5 min. Finally, the supernatant was collected and freeze-dried. The dried solids were then kept for further analysis.

Protein quantification assay

The total protein concentration of sachai inchi seeds was determined by performing a Pierce assay. First, 5 mg of the freeze-dried samples were weighed and solubilised in 1 mL of water. Next, 125 µL of Pierce reagent (Thermo Scientific, USA) was added into the 96-well plate containing 5 µL of the samples, and the solutions were mixed accordingly. Triplicates were prepared, and the solutions were incubated at room temperature for 10 min. To calculate the protein concentration of sachai inchi seeds, a standard curve was plotted using Bovine Serum Albumin (BSA) as standard and water as a buffer (Afiq & Habib, 2021). UV absorbance of the samples at 660 nm was determined using SpectraMax M5 microplate reader (Molecular Devices, USA). The averaged absorbance values obtained were then substituted into the standard curve (y) equation to get the protein concentration (x).

In-solution digestion

A 100-µL aliquot of sachai inchi protein sample was placed in the 6 M urea, 100 mM tris buffer, containing 1 mg of total protein, in a 1.5-mL plastic microcentrifuge tube. Next, 5 µL of the reducing agent (200 mM DTT) was added, and a gentle vortex was used to mix the sample. The protein mixture was then reduced

at room temperature for 1 hr. After reducing protein, 200 mM 20 μ L of alkylating agent iodoacetamide (IAA) was added, and the sample was mixed by gentle vortex, then left at room temperature for 1 hr. Next, 20 μ L of the reducing agent was added to consume unreacted iodoacetamide (Lee et al., 2017). A gentle vortex mixed the sample, allowing the reaction to stand at room temperature for 1 hr. The urea concentration was reduced by diluting the reaction mixture with 775 μ L of water. The solution was mixed with a gentle vortex. The urea concentration was reduced to $\sim$ 0.6 M through the dilution, a concentration at which the trypsin (200 ng/ μ L) maintains its activity. The sample was mixed gently by vortexing, and the digestion was carried out overnight at 37° C. On a subsequent day, the reaction was stopped, and two drops of concentrated acetic acid were added to the solution to adjust the pH to lower than 6. The digest was then concentrated using a concentrator before the analysis.

LC-MS/MS analysis

The sample was diluted with 50 μ L of 0.1% formic acid (FA) in deionised water before LC-MS/MS analysis. Then, the solution was filtered by a 0.2 μ m regenerated cellulose (RC) membrane syringe filter (Sartorius AG, Goettingen, Germany). After that, an LC-MS/MS (Liquid Chromatography-Mass Spectrometry) analysis was carried out by Kwan et al. (Kwan et al., 2015) with minor modifications. The samples were analysed using an LTQ-Orbitrap Velos Pro mass spectrometer coupled with an Easy-nLC II liquid chromatography system (Thermo Scientific, USA) with a running buffer of 0.1% FA in deionised water and 0.1% FA in 100% acetonitrile (ACN) and a flow rate of 0.3 μ L/min. The Easy column C18 (2 cm, 0.1mm i.d., 5 μ m; Thermo Scientific, San Jose, CA, USA) served as a pre-column and Easy column C18 (10 cm, 0.75mm i.d., 3 μ m; Thermo Scientific, San Jose, CA, USA) served as an analytical column. The pre-column was adjusted at a flow rate of 3 μ L/min for 15 μ L whereas the analytical column was equilibrated at a flow rate of 0.3 μ L/min for 4 μ L. When the mass spectrometer reached 220°C (capillary temperature) degrees, the sample eluent was sprayed inside, and a voltage source of 2.1 kV was fixed. To identify proteins and peptides, a full-scan mass analysis from m/z 300 to 2000 was performed at 60,000 resolution, and data-dependent MS/MS analyses were performed on eight ions with the highest abundances for each MS spectrum. High Collision Dissociation (HCD) was the fragmentation method, with a collision energy of 35. Each sample was examined three times, and the results presented were a compilation of the three analyses.

Peptide sequencing and protein identification

Using PEAKS Studio software (PEAKS), version 7.5 (Bioinformatics Solution, Waterloo, Canada), proteins and peptides were identified. Peptide sequences were generated using De Novo sequencing from raw LC-MS/MS data. The post-translational modification (PTM) parameters of oxidation (M) and carboxymethylation (C) are a part of the standard identification process. These peptide sequences were matched with the protein in the UniProt database (UniRef90 taxonomy Viridiplantae and saccha inchi protein database). To get a more accurate result, we filtered the peptide-spectrum matches (PSM) result by setting a protein significant score of $-10 \log P \geq 20$ while the $-10 \log P$ value was ≥ 15 for peptides. The number of the unique peptides was set as at least one or more. In our study, the percentage of coverage

was found as low as 1%, which might be the consequence of the filter for at least 1 or more unique peptide sequences.

Gene ontology (GO) study of sacha inchi seed proteins

After identification of the sacha inchi seed proteins by PEAKS studio software, we performed the gene ontology (GO) analysis (Götz et al., 2008). The GO annotation of all identified proteins was done by blast2go, and GO terms (Biological process, Cellular component, and Molecular functions) visualisation was performed by SRplot (<https://www.bioinformatics.com.cn/srplot>) bioinformatics tool (Krasny et al., 2022).

Results

Protein concentration

A calibration curve was plotted, and the total protein yield of seeds extracted from sacha inchi was determined. We found the protein at a concentration of 73.8 ± 0.002 mg/gm of freeze-dried seed flour of sacha inchi.

Protein profiling of sacha inchi seeds extract

The proteins of the seed sample were extracted and profiled subsequently by using LC-MS/MS. We identified a total of 217 proteins based on the Sacha inchi UniProt database. The result showed 152 proteins to be well-known and the remaining 65 are still uncharacterised. Of this number of proteins, 71 proteins were matched with sacha inchi proteins, while 146 proteins identified matched with other plant proteins. The observed disparities in protein quantification could potentially be attributed to the incompleteness of the sacha inchi protein database. Our endeavour encompassed the identification and emphasis of pivotal proteins originating from the protein compendium, specifically focusing on their relevance to the domains of food technology and biopharmaceutical industries. Based on significance and closeness to health-related benefits, we have picked the top 13 proteins to describe them in detail under the Discussion section. To name some of those proteins are Legumin A putative, Legumin B putative, Glutelin type-A 3, Glutelin type-B 5, Oleosin2, Oleosin3, Zeamatin, Glyceraldehyde-3-phosphate dehydrogenase and Heat shock protein 70 kDa, cruciferin, citrin, putative lyase, 11s globulin and others. These proteins are significant in storage, oxidative stress response, ATP-binding, ion binding, glycolysis, and protein-protein interaction. Potential proteins identified through PEAKS studio analysis are listed in Tables 1 and 2 with their protein ID, accession, -10lgP, coverage, unique peptide, average mass, and description.

GO-level distribution of seed proteins of *Plukenetia Volubilis* L.

Gene ontology analyses were performed for the proteins detected in the seed extract of sacha inchi using the Blast2GO software to classify the amino acid sequences. The sequences found after the analyses were visualised by plotting them against the classes of their protein functions using SRplot. (Fig. 1). After

the analyses, all sequences were segregated into three GO categories (biological process, cellular component, and molecular function). The related GO term charts are in the supplementary information (S3, S4, S5).

According to GO annotation as indicated in Fig. 1, the proportion of amino acid sequences associated with the proteins in the cellular metabolic process (GO: 0044237), organic substance metabolic process (GO: 0071704), and nitrogen compound metabolic process (GO: 0006807) sharing a percentage of 71.94%, 55.4% and 46.74% respectively are higher than other proteins under biological process category. The rest of the sequences in this category are related to the proteins of biogenesis, response to stimulus, biosynthetic process, and reproductive process activity. Membrane-bounded organelle (GO: 0043227) followed by Intracellular organelle (GO: 0043229) and chloroplast (GO: 0009507) represent the first three proteins holding a percentage of 73.91%, 75.65%, and 60.87% respectively in the cellular component category of the seed protein. Other proteins dominant here are membrane, envelope, non-membrane bounded organelle and plastid thylakoid.

Again, the proteins involved in the Molecular function category are organic cyclic compound binding (GO:0097159), heterocyclic compound binding (GO:1901363), and ion binding proteins (GO:00043167) showing a percentile share of 47.06%, 47.06% and 33.99% respectively that ranked the top three places in terms of number of sequences. The remaining sequences in this group are linked to the proteins of nutrient reservoir activity, transferase activity, small nucleotide binding, carbohydrate derivative binding and several others.

Discussion

The LC-MS/MS analysis reveals proteins with many beneficial applications in food technology, treatment, and research. Based on the importance and scale of protein usage in the Food and biopharmaceutical industry, we picked the top 13 proteins, and their usage details are described below. As depicted in Fig. 2, we considered 8 proteins to be important in the food industry, while 6 are crucial in the biopharmaceutical industry, sharing a percentage of 57% and 43%, respectively. The details of their application in the respective industry are described below.

Legumin A putative and Legumin B putative were discovered to be seed storage proteins in this experiment, measuring 475 AA and 476 AA long, respectively, and similar to the proteins in *Ricinus communis*. Their sequence similarity belongs to the 11S seed storage protein (globulins) family. Legumin B (fragment) was also detected in our study, which contains many similarities with the *Vernicia fordii*. Legumin A precursor that serves as the primary nitrogen source for the growing plant is a seed storage globulin discovered in *Vicia narbonensis* (Tandang-Silvas et al. 2010). In the legumin A precursor (Faba bean), arginine was found in the highest concentration, followed by leucine, which accounted for 10 and 9.3% of the total amino acid quantity, respectively, according to Liu, Y. et al. (Y. Liu et al., 2017). Cottonseed meal protein isolate (CSMPI) is a promising alternative protein source and is usually considered safe for human consumption. In treated CSMPI, the most abundant seed storage protein was

Legumin A (G3M16), followed by Legumin B (G3M3J0), vicilin A (G3M3J4), uncharacterised protein LOC107929052 (A0A1U8LMX7) (90% sequence similarity with vicilin C72 from *Gossypium australe*), vicilin C72 (A0A1U8LLA0) and 2S sulfur-rich seed storage protein 2-like (A0A1U8PKJ6). However, additional research is needed to determine the effects of long-term use of goods containing CSMPI as well as the frequency and seriousness of CSMPI allergies (Tan et al. 2022). The majority of the expressed sequence tags (ESTs) from the endocarp of neem (*Azadirachta indica*) seeds were related to seed storage proteins like legumin A, legumin B, citrin, and allergen proteins that are accumulated during seed development to serve as carbon and nitrogen sources during the subsequent seed germination phase (Narnoliya et al. 2014).

We identified Glutelin Type-A 3 putative in our sample, which is the seed storage protein. It is 497 AA long and has a molecular mass of 55,970 Dalton, similar to the protein found in *Ricinus communis* (castor bean). Previously, G. Li et al. found Glutelin type-A 3 putative, expressed in rice (*Oryza sativa*) endosperm nuclear fraction and Glutelin type-A 3 precursors in rice endosperm total protein extracts with a molecular weight of 56,014.69 and 50,509.81 Dalton respectively (G. Li et al., 2008). Later, Glutelin type-A 3, having a molecular mass of 30 kDa, was identified in *Oryza sativa* seed by Shu-Jun Liu and others (2014) in China (S. J. Liu et al. 2015). Again, in 2019, Ji and his co-workers detected Glutelin type-A 3 precursor and other seed storage proteins in flaxseed. In the same study, they analysed 23 mature flaxseed storage proteins and concluded that seed storage proteins have the potential to serve as antioxidant peptides (AP) upon enzymatic digestion (Ji et al. 2019). Antioxidant peptides are generally short (2–10 amino acid residues) and capable of acting as a radical scavenger, a proton donor, and a metal-ion chelator. In many nations, the manufacture of healthful foods includes the direct microbial fermentation of antioxidant peptides as an alternative to purified enzymes (Xiong, 2010).

Glutelin type-B 5 consists of 509 amino acids long with a mass of 57.22kDa and is similar to the protein found in *Ananas comosus*. A previous study identified a Glutelin type-B 5-like protein in the protein database (*Nelumbo nucifera* Gaertn.) from NCBI (Yu et al., 2021). Glutelin has traditionally been linked to rice quality, and it is regarded as a storage protein that makes up more than 80% of rice (Amagliani et al. 2017). Zhao *et al.* (2020) reported increased Glutelin type B proteins in stored rice. They also concluded that Glutelin type- B 5 and Glutelin type- B 5-like proteins were significantly upregulated in *japonica rice* and *indica rice*, respectively (Zhao et al. 2020).

Our study revealed that the oleosin 2 protein has a molecular mass of 15.8 kDa, similar to the oleosin 2 first identified in sacha inchi seed flour by Chandrasekaran *et al.* (Chandrasekaran et al. 2013). Oleosin 3 was detected in sacha inchi seed by Čepková and others in their study (Čepková et al., 2019). Depending on the oleaginous plants and their tissues, the seed oil body (OB) contains a variety of proteins. One of the most practical and economical approaches is oleosin fusion technology. Wei Li and his co-workers successfully expressed human insulin-like growth factor 1 (hIGF-1) as an oleosin-hIGF-1 fusion in *Arabidopsis thaliana* seeds and showed that the oleosin-hIGF-1 expressed in the plant was biologically active (W. Li et al. 2011). It has been suggested that ubiquitination of the sesame seed oleosin and caleosin plays a role in the interaction of OBs with glyoxysomes, which is required for lipid breakdown to

liberate free fatty acids for energy (Hsiao and Tzen 2011). In recent years, there has been much interest in the production of antibiotics, enzymes, and emulsifiers for the food sector as well as for medicinal uses employing gene engineering using oleosin promoters in seeds (Chandrasekaran et al. 2013).

Here, we observed the Zeamatin putative as a defence response protein similar to castor bean with a Molecular weight of 27,445 Dalton; its sequence belongs to the thaumatin family. Zeamatin is an antifungal protein found in corn seeds. To defend themselves against fungal attacks, numerous plants manufacture antifungal proteins, which can be found in various plant tissues, including seeds, bulbs, leaves, tubers, fruits, shoots, and roots (Yan et al. 2015). Zeamatin is mainly recognised for its antifungal activity and has been proposed as a drug working on vaginal murine candidosis cells or in transgenic plants, boosting their resistance to diseases and pests (J. J. Liu et al. 2010).

In the current study, the detected protein glyceraldehyde-3-phosphate dehydrogenase was similar to one in Mesangiospermae with a molecular mass of 37.2 kDa. Within the glycolytic pathway, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a highly conserved enzyme. Glyceraldehyde 3-phosphate is converted to glycerate-1, 3-bisphosphate by the action of GAPDH, which also triggers the synthesis of NADH throughout the process (Shunxian Wang et al., 2020). Gani Z. et al. reported that they extracted 250 µg of highly purified GAPDH from 20 g of pea seeds (Gani et al., 2016). Wang, L.L. et al. concluded that the application of immobilised GAPDH on Fe₃O₄ magnetic nanoparticles (MNPs) to grape and black raspberry wines confirmed its effectiveness in the actual food, destroying over 80% of histamine while barely changing the wine's composition, demonstrating that enzymatic technology based on Fe₃O₄-GAPDH MNPs has excellent potential for reducing biogenic amines in fermented beverages (L. L. Wang et al., 2021). The apoptotic cascade, for which GAPDH nuclear translocation appears crucial, is shown to be involved when GAPDH is knocked down using an antisense approach. GAPDH may be a therapeutic target for disease interventions and future drug research and development because medications used to treat Parkinson's disease (PD), Huntington's disease (HD), and Alzheimer's disease (AD) can bind to GAPDH and decrease its overexpression. These actions correspond with their neuroprotective benefits (Chuang et al. 2005).

We have identified Heat shock protein 70 kDa (Hsp70) in Sacha Inchi seed extract with the similarity of one in the microalgae *Raphidocelis subcapitata*, the primary producer of the freshwater ecosystem. The molecular mass of this protein is 109 kDa with 1,011 amino acid sequences. In cancerous cells, Heat shock proteins, mainly Hsp70 or Hsp90, alert the immune cells to antigens. They deliver antigens to the immune system's antigen-presenting cells (APCs) through surface receptors (Multhoff, 2006). Heat shock protein 70 (Hsp70), a cytoplasmic chaperone, plays a crucial part in cellular homeostasis maintenance through its chaperone functions. In humans, insulin resistance may be linked to low Hsp70 levels. Heat shock protein 70 (Hsp70) guards against accumulating damaged proteins and cell ageing-related functional loss. They strengthen healthy cells by defending them from damage and stress (Chichester et al. 2015; Jindal 1996). Hsp70 has been used as a marker for stressed cells. Viral vector-mediated Hsp70 overexpression increased the survival of neurons and astrocytes against ischemic and ischemia-like insults, demonstrating a direct role in neuroprotection. The outcome is worsened in transgenic mice

lacking Hsp70 compared to those that overexpress it (Giffard and Yenari 2004; J. Y. Kim et al. 2016). In addition to its use in nanoparticle technology, measurements of variations in Hsp70 concentration during pathological circumstances have aided the use of this protein as a diagnostic biomarker. As a result, the early and precise detection of Hsp70 (as the major or secondary biomarker) can be crucial in quickly diagnosing several disorders, such as neurological diseases, diabetes, cardiac diseases, cancers, and some viral infections such as COVID-19 (Negahdary and Angnes 2022).

Cruciferin, another seed storage protein identified in our study, is very promising in the food supplement industry. A study showed that Cruciferin and Napin, two rapeseed storage proteins, are both credited with having well-balanced amino acid compositions. They are ideal for usage as dietary ingredients because they may satisfy both essential and non-essential amino acid needs in humans (Chmielewska et al., 2021).

Citrin insufficiency impacts numerous metabolic pathways and can manifest differently at different periods of life with various symptoms. Neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD) and pre-adult-onset type II citrullinemia (CTLN2) are two manifestations of citrin deficiency (Okamoto et al., 2021). medium-chain triglyceride (MCT) milk and MCT supplementation under a low-carbohydrate diet are recommended for NICCD and CTLN2, respectively (Hayasaka, 2021). 11S globulin protein plays a crucial role in protein emulsion technology. In particular, the use of high internal phase Pickering emulsions (HIPPEs) as a replacement for trans and saturated fats, nutrient encapsulation and delivery, food 3D printing, and complicated porous material design are highlighted as significant research trends. The Quinoa protein, which is made up of 50–60% 11S globulin, can enhance the quality of HIPPEs by converting liquid oils into solid-like fats without partially hydrogenated oils (Ji & Luo, 2023). The putative lyase identified in this study is similar to that of *Rosa chinensis*. Recently, five putative alginate lyase coding genes have been found in the genome of *Microbulbifer* sp. AWL1 and this lyase can be used in the production of active oxygen species and medical treatment of cystic fibrosis (Li et al., 2022)

Conclusion

This research on the protein profile in sacha inchi (*Plukenetia volubilis* L.) raw seeds will pave the way for multipurpose uses in the food and biopharmaceutical industries. Although sacha inchi is rich in nutritional elements, it remains neglected and underutilised. The proteins in sacha inchi seeds have promising potential to revolutionise the ever-growing demand for high-quality food and biomedical technology. The presence of these proteins identified in our study highlights the complexity of the sacha inchi seed protein composition and provides a foundation for future studies to characterise further and understand these proteins' functional significance. More sacha inchi seed proteins are anticipated to be identified after the plant's genome has been fully sequenced and made public. The data found in this study about the seed storage, defence response, and stress response protein and their variability within the population will certainly work as the benchmark to advance the research further in the future.

Declarations

Ethical approval: Not applicable

Consent to participate: Not applicable

Consent to publish: All authors read and approved the final manuscript to publish.

Acknowledgement: The authors convey sincere gratitude to other lab colleagues for their cooperation in experimenting, and *Mohammad Harun Ur Rashid* pays gratitude to the USM authority for supporting his study with gracious Graduate fellowship USM (GFUSM).

Authors' contribution: Protein quantification, preparation of the original draft, and data analysis were done by *Mohammad Harun Ur Rashid*. Sample preparation and participation in LC-MS/MS investigation were performed by *Erica Kok Jia Yi*. Data analysis, providing critical information in protein quantification, and determination of peptide sequences by using PEAKS studio were performed by *Nor Datiakma Mat Amin*. Funding acquisition, sample collection, designing experiments, formal investigation and analysis of LC-MS/MS, and overall supervision of the study were performed by *Mohd Nazri Ismail*.

Funding: This study was funded by the Prototype Research Grant Scheme (PRGS/1/2021/SKK14/USM/02/1) awarded by the Malaysia Ministry of Higher Education.

Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.

Availability of data and materials: All data supporting reported results in this manuscript are available upon request.

References

1. Afiq, M., & Habib, H. (2021). Extraction and identification of biologically important proteins from the medicinal plant God's crown (*Phaleria macrocarpa*). May, 1–12. <https://doi.org/10.1111/jfbc.13817>
2. Amagliani, L., O'Regan, J., Kelly, A. L., O'Mahony, J. A. (2017). The composition, extraction, functionality, and applications of rice proteins: A review. *Trends in Food Science and Technology*, 64, 1–12. <https://doi.org/10.1016/j.tifs.2017.01.008>
3. Čepková, P.H., Jágr, M., Viehmannová, I., Dvořáček, V., Huansi, D.C. & Mikšík, I. (2019). Diversity in seed storage protein profile of oilseed crop *Plukenetia volubilis* from Peruvian Amazon. *International Journal of Agriculture and Biology*, 21(3), 679–688.
4. Chandrasekaran, U., Wang, X., Liu, A. (2013). Characterisation, expression profiling, and heterologous function analysis of two oleosin genes *pvOle1* and *pvOle2* from *sacha inchi* (*Plukenetia volubilis*). *International Journal of Agriculture and Biology*, 15(3), 435–442.

5. Chichester, L., Wyli, A.T., Craft, S. & Kavanagh, K. (2015). Muscle heat shock protein 70 predicts insulin resistance with aging. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 70(2), 155–162. <https://doi.org/10.1093/gerona/glu015>
6. Chmielewska, A., Kozłowska, M., Rachwał, D., Wnukowski, P., Amarowicz, R., Nebesny, E., & Rosicka-Kaczmarek, J. (2021). Canola/rapeseed protein–nutritional value, functionality, and food application: a review. *Critical Reviews in Food Science and Nutrition*, 61(22), 3836–3856. <https://doi.org/10.1080/10408398.2020.1809342>
7. Chuang, D.M., Hough, C., Senatorov, V.V. (2005). Glyceraldehyde-3-phosphate dehydrogenase, apoptosis, and neurodegenerative diseases. *Annual Review of Pharmacology and Toxicology*, 45,269–290. <https://doi.org/10.1146/annurev.pharmtox.45.120403.095902>
8. Da Silva Soares, B., Siqueira, R.P., De Carvalho, M.G., Vicente, J., Garcia-Rojas, E.E. (2019), Microencapsulation of sachá inchi oil (*Plukenetia volubilis* L.) using complex coacervation: Formation and structural characterisation. *Food Chemistry*, 298, 125045. <https://doi.org/10.1016/j.foodchem.2019.125045>
9. Del-Castillo, Á.M.R., Gonzalez-Aspajo, G., De Fátima, Sánchez-Márquez, M., Kodahl, N. (2019). Ethnobotanical Knowledge in the Peruvian Amazon of the Neglected and Underutilized Crop Sachá Inchi (*Plukenetia volubilis* L.). *Economic Botany*, 73(2), 281–287. <https://doi.org/10.1007/s12231-019-09459-y>
10. Follegatti-Romero, L.A., Piantino, C.R., Grimaldi, R., Cabral, F.A. (2009) Supercritical CO₂ extraction of omega-3 rich oil from Sachá inchi (*Plukenetia volubilis* L.) seeds. *Journal of Supercritical Fluids*, 49(3), 323–329. <https://doi.org/10.1016/j.supflu.2009.03.010>
11. Fu, Q., Niu, L., Zhang, Q., Pan, B.Z., He, H., Xu, Z.F. (2014). Benzyladenine treatment promotes floral feminisation and fruiting in a promising oilseed crop *Plukenetia volubilis*. *Industrial Crops and Products*, 59, 295–298. <https://doi.org/10.1016/j.indcrop.2014.05.028>
12. Gani, Z., Boradia, V.M., Raghu Ram, J., Suryavanshi, P.M., Patil, P., Kumar, S., Singh, R., Raje, M., Raje, C.I. (2016). Purification and characterisation of glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) from pea seeds. *Protein Expression and Purification*, 127, 22–27. <https://doi.org/10.1016/j.pep.2016.06.014>
13. Giffard, R.G., Yenari, M.A. (2004). Many mechanisms for hsp70 protection from cerebral ischemia. *Journal of Neurosurgical Anaesthesiology*, 16(1),53–61. <https://doi.org/10.1097/00008506-200401000-00010>
14. Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36(10), 3420–3435. <https://doi.org/10.1093/nar/gkn176>
15. Gutiérrez, L.F., Rosada, L.M., Jiménez, Á. (2011). Chemical composition of Sachá Inchi (*Plukenetia volubilis* L.) seeds and characteristics of their lipid fraction. *Grasas y Aceites*, 62(1),76–83. <https://doi.org/10.3989/gya044510>

16. Hamaker, B.R., Valles. C., Gilman, R., Hardmeier, R.M., Clark, D., Garcia, H.H., Gonzales, A.E., Kholstad, I., Castro, M., Valdivia, R., Rodregues, D., Lescano, M. (1992). Amino Acid and Fatty Acid Profiles of the Inca Peanut (*Plukenetia volubilis*). *Cereal Chemistry*, 69(4), 461-463
17. Hanssen, H.P., Schmitz-Hübsch, M. (2011), in *Nuts and seeds in health and disease prevention*, sacha inchi (*Plukenetia volubilis* L.) nut oil and its therapeutic and nutritional uses (Preedy, V.R, Watson, R.R., Patel, V.B., First edition), Elsevier, USA, pp. 991-994. <https://doi.org/10.1016/B978-0-12-375688-6.10117-3>
18. Herrera-Calderon, O., Yuli-Posadas, R.Á., Tinco-Jayo, J.A., Enciso-Roca, E., Franco-Quino, C., Chumpitaz-Cerrate, V., Figueroa-Salvador, L. (2019). Neuroprotective effect of sacha inchi oil (*Plukenetia volubilis* L.) in an experimental model of epilepsy. *Pharmacognosy Journal*, 11(6), 1591–1596. <https://doi.org/10.5530/PJ.2019.11.243>
19. Hsiao, E.S.L., Tzen, J.T.C. (2011). Ubiquitination of oleosin-H and caleosin in sesame oil bodies after seed germination. *Plant Physiology and Biochemistry*, 49(1), 77–81. <https://doi.org/10.1016/j.plaphy.2010.10.001>
20. Hayasaka, K. (2021). Metabolic basis and treatment of citrin deficiency. *Journal of Inherited Metabolic Disease*, 44(1), 110–117. <https://doi.org/10.1002/jimd.12294>
21. Ji, C., & Luo, Y. (2023). Plant protein-based high internal phase Pickering emulsions: Functional properties and potential food applications. *Journal of Agriculture and Food Research*, 12(May), 100604. <https://doi.org/10.1016/j.jafr.2023.100604>
22. Ji, D., Udenigwe, C.C., Agyei, D. (2019). Antioxidant peptides encrypted in flaxseed proteome: An in-silico assessment. *Food Science and Human Wellness*, 8(3), 306–314. <https://doi.org/10.1016/j.fshw.2019.08.002>
23. Jindal, S. (1996). Heat shock proteins: Applications in health and disease. *Trends in Biotechnology*, 14(1), 17–20. [https://doi.org/10.1016/0167-7799\(96\)80909-7](https://doi.org/10.1016/0167-7799(96)80909-7)
24. Kim, J.Y., Kim, N., Zheng, Z., Lee, J.E., Yenari, M.A. (2016). 70-kDa Heat Shock Protein Downregulates Dynamin in Experimental Stroke: A New Therapeutic Target? *Stroke*, 47(8), 2103–2111. <https://doi.org/10.1161/STROKEAHA.116.012763>
25. Krasny, L., Wilding, C. P., Perkins, E., Arthur, A., Guljar, N., Jenks, A. D., Fisher, C., Judson, I., Thway, K., Jones, R. L., & Huang, P. H. (2022). Proteomic Profiling Identifies Co-Regulated Expression of Splicing Factors as a Characteristic Feature of Intravenous Leiomyomatosis. *Cancers*, 14(12), 1–14. <https://doi.org/10.3390/cancers14122907>
26. Kwan, S.H., Baie, S., Mohammed, N., Ismail, M.N. (2015). Proteomic Profiling of Freeze- and Spray-Dried Water Extracts of Snakehead Fish (*Channa striatus*): In Search of Biomolecules for Wound Healing Properties. *South Asian Journal of Life Sciences*, 3(1), 22–41. <https://doi.org/10.14737/journal.sajls/2015/3.1.22.41>
27. Lee, H. X., Ahmad, F., Saad, B., & Ismail, M. N. (2017). Evaluation of Extraction Methods for the Identification of Proteins from Date Palm (*Phoenix dactylifera* L .) Seed and Flesh. 6068(September). <https://doi.org/10.1080/10826068.2017.1365250>

28. Li, G., Nallamilli, B.R.R., Tan, F., Peng, Z. (2008). Removal of high-abundance proteins for nuclear subproteome studies in rice (*Oryza sativa*) endosperm. *Electrophoresis*, 29(3), 604–617. <https://doi.org/10.1002/elps.200700412>
29. Li, W., Li, L., Li, K., Lin, J., Sun, X., Tang, K. (2011). Expression of biologically active human insulin-like growth factor 1 in *Arabidopsis thaliana* seeds via oleosin fusion technology. *Biotechnology and Applied Biochemistry*, 58(3), 139–146. <https://doi.org/10.1002/bab.30>
30. Liu, J.J., Sturrock, R., Ekramoddoullah, A.K.M. (2010). The superfamily of thaumatin-like proteins: Its origin, evolution, and expression towards biological function. *Plant Cell Reports*, 29(5), 419–436. <https://doi.org/10.1007/s00299-010-0826-8>
31. Liu, S.J., Xu, H.H., Wang, W.Q., Li, N., Wang, W.P., Møller, I.M., Song, S.Q. (2015). A proteomic analysis of rice seed germination as affected by high temperature and ABA treatment. *Physiologia Plantarum*, 154(1), 142–161. <https://doi.org/10.1111/ppl.12292>
32. Liu, Y., Wu, X., Hou, W., Li, P., Sha, W., Tian, Y. (2017). Structure and function of seed storage proteins in faba bean (*Vicia faba* L.). *3 Biotech*, 7(1), 74. <https://doi.org/10.1007/s13205-017-0691-z>
33. Li, Z., Du, Z., Li, H., Chen, Y., Zheng, M., Jiang, Z., Du, X., Ni, H., & Zhu, Y. (2022). Characterisation of marine bacterium *Microbulbifer* sp. ALW1 with *Laminaria japonica* degradation capability. *AMB Express*, 12(1). <https://doi.org/10.1186/s13568-022-01482-y>
34. Moser, B.R. (2016). Fuel property enhancement of biodiesel fuels from common and alternative feedstocks via complementary blending. *Renewable Energy*, 85, 819–825. <https://doi.org/10.1016/j.renene.2015.07.040>
35. Multhoff, G. (2005), in *Molecular Chaperon in Health and Disease*, Vol 172: Heat Shock Proteins in Immunity (Matthias, G., First Edition) Springer, Hanover, pp. 279-304.
36. Narnoliya, L.K., Rajakani, R., Sangwan, N.S., Gupta, V., Sangwan, R.S. (2014). Comparative transcripts profiling of fruit mesocarp and endocarp relevant to secondary metabolism by suppression subtractive hybridisation in *Azadirachta indica* (neem). *Molecular Biology Reports*, 41(5), 3147–3162. <https://doi.org/10.1007/s11033-014-3174-x>
37. Nascimento, A.K.L., Melo-Silveira, R.F., Dantas-Santos, N., Fernandes, J.M., Zucolotto, S.M., Rocha, H.A. O., Scortecchi, K.C. (2013). Antioxidant and anti-proliferative activities of leaf extracts from *Plukenetia volubilis* Linneo (Euphorbiaceae). *Evidence-Based Complementary and Alternative Medicine*, 1–10. <https://doi.org/10.1155/2013/950272>
38. Negahdary, M., Angnes, L. (2022). An aptasensing platform for detection of heat shock protein 70 kDa (HSP70) using a modified gold electrode with lady fern-like gold (LFG) nanostructure. *Talanta*, 246, 123511. <https://doi.org/10.1016/j.talanta.2022.123511>
39. Okamoto, M., Okano, Y., Okano, M., Yazaki, M., Inui, A., Ohura, T., Murayama, K., Watanabe, Y., Tokuhara, D., & Takeshima, Y. (2021). Food preferences of patients with citrin deficiency. *Nutrients*, 13(9). <https://doi.org/10.3390/nu13093123>
40. Oyeboode, O., Kandala, N.B., Chilton, P.J., Lilford, R.J. (2016). Use of traditional medicine in middle-income countries: A WHO-SAGE study. *Health Policy and Planning*, 31(8), 984–991.

<https://doi.org/10.1093/heapol/czw022>

41. Ruiz, C., Díaz, C., Anaya, J., Rojas, R. (2013). Approximate analysis, antinutrients, fatty acids, and amino acids profiles of seeds and cakes from 2 species of sacha inchi: *Plukenetia Volubilis* and *Plukenetia huayllabambana*. *Revista de La Sociedad Química Del Perú* ,79(1), 29–36.
42. Sathe, S.K., Kshirsagar, H.H., Sharma, G.M. (2012). Solubilisation, Fractionation, and Electrophoretic Characterization of Inca Peanut (*Plukenetia volubilis* L.) Proteins. *Plant Foods for Human Nutrition*, 67(3), 247–255. <https://doi.org/10.1007/s11130-012-0301-5>
43. Tan, C.F., Kwan, S.H., Lee, C.S., Soh, Y.N.A., Ho, Y.S., Bi, X. (2022). Cottonseed Meal Protein Isolate as a New Source of Alternative Proteins: A Proteomics Perspective. *International Journal of Molecular Sciences*, 23(17), 10105. <https://doi.org/10.3390/ijms231710105>
44. Tandang-Silvas, M.R.G., Fukuda, T., Fukuda, C., Prak, K., Cabanos, C., Kimura, A., Itoh, T., Mikami, B., Utsumi, S., Maruyama, N. (2010). Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1804(7), 1432–1442. <https://doi.org/10.1016/j.bbapap.2010.02.016>
45. Wang, L.L., Fan, M., Xing, X., Liu, Y., Sun, S. (2021). Immobilisation of glyceraldehyde-3-phosphate dehydrogenase on Fe₃O₄ magnetic nanoparticles and its application in histamine removal. *Colloids and Surfaces B: Biointerfaces*, 205, 111917. <https://doi.org/10.1016/j.colsurfb.2021.111917>
46. Wang Shunxian Chen, H., Tang, X., Zhang, H., Hao, G., Chen, W., Chen, Y.Q. (2020). The Role of Glyceraldehyde-3-Phosphate Dehydrogenases in NADPH Supply in the Oleaginous Filamentous Fungus *Mortierella alpina*. *Frontiers in Microbiology*, 11, 1–10. <https://doi.org/10.3389/fmicb.2020.00818>
47. Wang Sunan Zhu, F., Kakuda, Y., (2018). Sacha inchi (*Plukenetia volubilis* L.): Nutritional composition, biological activity, and uses. *Food Chemistry*, 265, 316–328. <https://doi.org/10.1016/j.foodchem.2018.05.055>
48. Xiong, Y.L. (2010), in *Antioxidant Peptides, Bioactive Proteins and Peptides as Functional Foods and Nutraceuticals* (Mine, Y., Li-Chan, E., Jiang, B., First edition), Blackwell, USA, pp. 29-42.
49. Yan, J., Yuan, S.S., Jiang, L.L., Ye, X.J., Ng, T.B., Wu, Z.J. (2015) Plant antifungal proteins and their applications in agriculture. *Applied Microbiology and Biotechnology*, 99(12), 4961–4981. <https://doi.org/10.1007/s00253-015-6654-6>
50. Yu, Y.P., Lai, S., Chang, C.R., Chen, W.C., Wu, S.H., Lu, C.P. (2021). Peptidomic analysis of low molecular weight antioxidative peptides prepared by lotus (*Nelumbo nucifera* Gaertn.) seed protein hydrolysates. *Lwt*, 144, 111138. <https://doi.org/10.1016/j.lwt.2021.111138>
51. Zhao, Q., Xue, Y., Shen, Q. (2020). Changes in the major aroma-active compounds and taste components of Jasmine rice during storage. *Food Research International*, 133, 109160. <https://doi.org/10.1016/j.foodres.2020.109160>

Figures

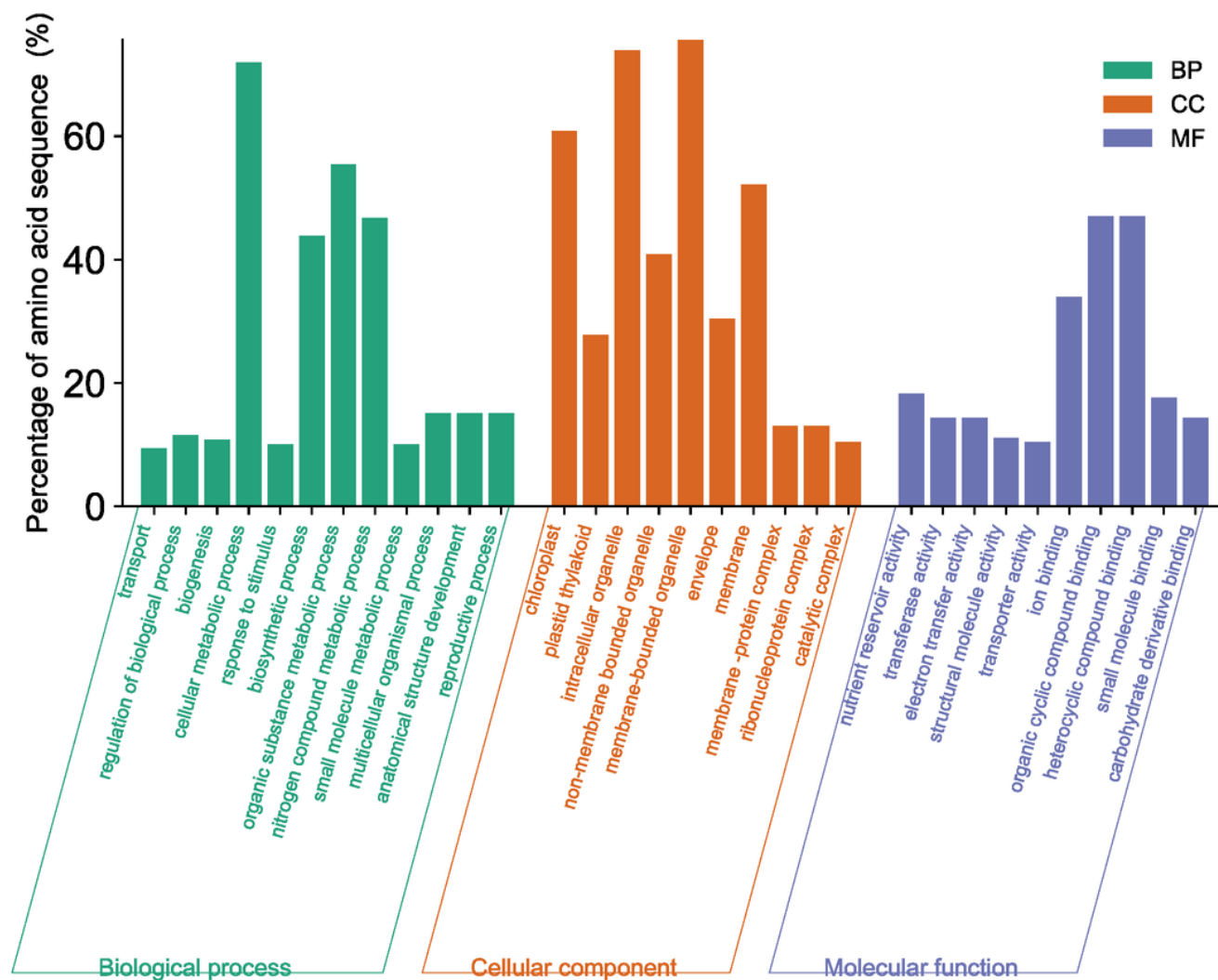


Figure 1

The distribution of amino acid sequences involved in potential proteins identified in the Sacha Inchi seed extract are linked to either biological process (P), Cellular component (C), or Molecular function (F).

Distribution of protein based on Industry

- Legumine A putative : Food Industry
- Legumine B putative : Food Industry
- Glutelin Type-A 3 putative : Food Industry
- Glutelin type-B 5 : Food Industry
- glyceraldehyde-3-phosphate dehydrogenase : Food Industry
- Cruciferin : Food Industry
- Citrin : Food Industry
- 11S globulin : Food Industry
- Oleosin 2 : Biopharmaceutical industry
- Oleosin 3 : Biopharmaceutical industry
- Zeamatin putative : Biopharmaceutical industry
- glyceraldehyde-3-phosphate dehydrogenase : Biopharmaceutical industry
- Heat shock protein 70 kDa : Biopharmaceutical industry
- Putative lyase : Biopharmaceutical industry

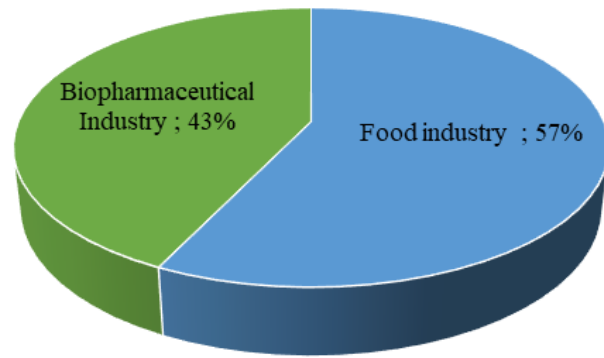


Figure 2

Proportion of proteins found in Sacha inchi seeds based on their application in industries.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)