Characterization and differentiation of quinoa seed proteomes by label-free mass spectrometry-based shotgun proteomics

Rocío Galindo-Luján [a],[1], Laura Pont [a],[1], Zoran Minic [b], Maxim V. Berezovski [b], Victoria Sanz-Nebot [a], Fernando Benavente [a],[*]

[a] Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA·UB), University of Barcelona, 08028 Barcelona, Spain
[b] John L. Holmes Mass Spectrometry Facility, Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

ARTICLE INFO

ABSTRACT

Quinoa seed proteins are of prime importance in human nutrition and in plant breeding for cultivar identification and improvement. In this study, proteins from seeds of black, red, white quinoa from Peru and white quinoa from Bolivia (also known as royal) were extracted, digested and analyzed by nano-liquid chromatography coupled to Orbitrap tandem mass spectrometry (LC-MS/MS). The raw mass spectra data were processed for identification and label-free quantification (LFQ) using MaxQuant/Andromeda against a specific quinoa database from The National Center for Biotechnology Information (NCBI). In total, 1,211 quinoa proteins (85 were uncharacterized) were identified. Inspection and visualization using Venn diagrams, heat maps and Gene Ontology (GO) graphs revealed proteome similarities and differences between the four varieties. The presented data provides the most comprehensive experimental quinoa seed proteome map existing to date in the literature, as a starting point for more specific characterization and nutritional studies of quinoa and quinoa-containing foodstuff.

1. Introduction

Quinoa (*Chenopodium quinoa* Willd.) is an Andean non-model herbaceous flowering plant that belongs to the *Chenopodiaceae* family (*Caryophillales* order). Quinoa produces seeds that can be milled into flour and used as a cereal crop (Burrieza, Rizzo, & Pérez, 2020; Dakhili, Abdolalizadeh, Hosseini, Shojaee-Aliabadi, & Mirmoghtadaie, 2019; Pereira et al., 2019; Vega-Gálvez et al., 2010). The interest in quinoa consumption has grown progressively due to its attractive nutritional composition, mainly based on the presence of lipids, carbohydrates, polyphenols, fiber and especially proteins (Piñuel et al., 2019; Vilcacundo & Hernández-Ledesma, 2017). Quinoa proteins, which are rich in essential amino acids and account for around 12–16% of the seed mass content, have been described to present antidiabetic, antihypertensive, antioxidant and anti-tumoral bioactivities (Capraro et al., 2020; Vilcacundo, Martínez-Villaluenga, & Hernández-Ledesma, 2017; Vilcacundo, Miralles, Carrillo, & Hernández-Ledesma, 2018).

Despite the importance of quinoa proteins in human nutrition, biomedicine, cultivar identification, quality control and authentication, studies focused on the characterization of the quinoa seed proteome are scarce. Until recent years, two-dimensional gel electrophoresis (2D-GE) followed by matrix assisted laser desorption ionization-mass spectrometry (MALDI-MS) has been the most applied technique for visualizing proteome profiles of quinoa seeds (Aloisi et al., 2016). However, 2D-GE possesses several limitations, including a weak ability to detect low-abundance proteins, narrow linear detection dynamic range, long analysis time, low sensitivity and reproducibility. Nowadays, liquid chromatography-tandem mass spectrometry (LC-MS/MS) is the method of choice for shotgun proteomics (Aizat & Hassan, 2018; Gao & Yates, 2019; Schubert, Röst, Collins, Rosenberger, & Aebersold, 2017). This approach provides an indirect measurement of proteins through the bottom-up LC-MS/MS analysis of peptides derived from proteolytic digestion of the original complex protein mixture. So far, only two studies have reported the shotgun proteome analysis of quinoa seeds (Burrieza, Rizzo, Moura, Silveira, & Maldonado, 2019; Capriotti et al., 2015), thereby limiting the current understanding of the quinoa proteome and pointing out that its comprehensive characterization is not so straightforward.

One of the main challenges in shotgun large-scale proteomics when working with non-model plant species such as quinoa is the limited availability of protein sequences in the commonly used databases (Heck & Neely, 2020). In 2015, due to the lack of complete quinoa genomic sequencing and proteomic data, Capriotti et al. (Capriotti et al., 2015) were forced to use a protein sequence database from plants of the *Caryophillales* order, which included a very small number of quinoa proteins. In that study, only 4 quinoa proteins were identified, whereas the rest of proteins (3 4 8) were found by sequence homology to plant species phylogenetically close to quinoa, especially beetroot (*Beta vulgaris*) and spinach (*Spinacia oleracea*). The publication of the quinoa genome in 2017 (Jarvis et al., 2017; Zou et al., 2017) provided an important breakthrough for the experimental characterization of the quinoa seed proteome. Taking advantage of this novel information, in 2019, Burrieza et al. (Burrieza et al., 2019) identified a total of 337 quinoa proteins, including novel lysine-rich seed storage globulins.

Indeed, a critical point when performing shotgun proteomics is the selection of the most suitable database for each particular case (Xu, 2012). UniProt/SwissProt database (https://www.uniprot.org) is the most widely used protein sequence database, especially for studying model organisms (Bateman, 2019). The second most employed database is The National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). The main advantage of the NCBI database is that it works with the Reference Sequence (RefSeq) project (O'Leary et al., 2016), which, contrary to Uniprot, maintains and curates new and updated genome annotations. At the time of writing this article, for the particular case of quinoa, the Uniprot database (unreviewed section) contains 232 *Chenopodium quinoa* protein entries, whereas the RefSeq NCBI database contains a total number of 63,370. This fact suggests the need for large-scale proteomics studies to experimentally confirm and review the data provided for quinoa by the different databases.

In this study, a label-free LC-MS/MS shotgun proteomics approach with a state-of-the-art Orbitrap mass spectrometer was applied to comprehensively characterize the proteome of the most typical quinoa seed varieties, i.e. black (B), red (R), white quinoa from Peru (W) and white quinoa from Bolivia (also referred to as royal, RO). The most appropriate conditions (regarding sample preparation, instrumentation and data analysis) were carefully selected, significantly improving the quinoa proteome coverage from previous studies and revealing differences between varieties. The proposed methodology provides the most comprehensive experimental quinoa seed proteome map existing to date in the literature, as a tool for more specific characterization and nutritional studies.

2. Materials and methods

2.1. Chemicals

All the chemicals used in the preparation of buffers and solutions were of analytical reagent grade or better. Sodium hydroxide (NaOH, $\geq$99.0%, pellets), hydrochloric acid (HCl, 37% (v/v)), boric acid (H_3BO_3, $\geq$99.5%), acetonitrile (ACN, LC-MS grade), water (LC-MS grade), bovine serum albumin (BSA, relative molecular mass (M_r) of approximately 66,000), acetic acid (HAc, glacial), formic acid (FA, 99.0%), acetone ($\geq$99.5%), ethanol (96.0%), glycerol ($\geq$99.5%), tris (hydroxymethyl)aminomethane hydrochloride (Tris-HCl, $\geq$99.9%), sodium dodecyl sulfate (SDS, $\geq$99.8%), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, $\geq$99.5%), urea ($\geq$99.0%), Triton™ X-100 (laboratory grade), tris(2-carboxyethyl)phosphine hydrochloride (TCEP, $\geq$98.0%) and iodoacetamide (IAA, $\geq$99.0%) were supplied by Merck (Darmstadt, Germany). β-mercaptoethanol ($\geq$99.0%) was provided by PanReac Applichem (Barcelona, Spain). Trypsin/Lys-C enzyme mix (MS grade) was purchased from Promega (Madison, WI, USA). Bromophenol blue, tetramethylethylenediamine (TEMED, $\geq$99.0%), acrylamide/bis solution (30% (v/v)), ammonium persulfate (APS,

$\geq$98.0%) and Bio-Safe™ Coomassie stain were supplied by Bio-Rad (Hercules, CA, USA). BenchMark™ Protein Ladder was provided by Thermo Fisher Scientific (Waltham, MA, USA). Water with conductivity lower than 0.05 µS/cm was obtained using a Milli-Q water purification system (Millipore, Molsheim, France).

2.2. Apparatus and procedures

2.2.1. Sample preparation

Black (B), red (R), white (W) quinoa from Peru, as well as royal white (RO) quinoa from Bolivia were acquired in local supermarkets from Barcelona. Quinoa grains were dried in an air-current oven at 40 °C for 24 h, ground in a coffee grinder and stored at room temperature (RT) in a desiccator. Quinoa proteins were extracted as in our previous work (Galindo-Luján, Pont, Sanz-Nebot, & Benavente, 2021): 250 mg of the ground sample were mixed with 1 mL of water and 39 µL of 1 M NaOH (final pH was 10.0) using a vortex Genius 3 (Ika®, Staufen, Germany) and then incubated for 1 h at 36 °C with constant shaking in a TS-100 thermoshaker (Biosan, Riga, Latvian Republic). Separation of soluble proteins from the insoluble residue was performed by centrifugation at 15,000 × g for 20 min at 4 °C in a cooled Rotanta 460 centrifuge (Hettich Zentrifugen, Tuttlingen, Germany). For protein purification, the supernatant pH was adjusted with 22 µL of 1 M HCl to obtain a final pH value of 5.0. After centrifugation at 15,000 × g for 20 min at 4 °C, precipitated proteins were resuspended in 1 mL of a solution consisting of 60 mM H_3BO_3 (pH adjusted to 9.0 with NaOH). All pH measurements were made with a Crison 2002 potentiometer and a Crison electrode 52–03 (Crison Instruments, Barcelona, Spain).

2.2.2. CE-UV

A capillary electrophoresis (CE) instrument with a diode-array detector (7100 CE, Agilent Technologies, Waldbronn, Germany) was used to estimate the total amount of protein in the quinoa extracts measuring the absorbance at 280 nm. A calibration curve was established by analyzing BSA standard solutions at concentrations between 50 and 1,000 µg/mL. BSA standards and quinoa protein extracts (three independent replicates for B, R, W and RO quinoa) were injected for 10 s at 50 mbar in a 58 cm total length (L_T) × 50 µm internal diameter (i.d.) × 365 µm outer diameter (o.d.) fused silica capillary (Polymicro Technologies, Phoenix, AZ, USA). Flow injection experiments were performed without voltage, mobilizing the plug of sample by applying 50 mbar of pressure after the injection. Absorbance was measured from the height of the detected protein peaks. The total protein contents estimated by CE-UV in the quinoa extracts (average of the three analyzed replicates, relative standard deviation (%RSD) lower than 5% in all cases) were 2.2% (m/m), 2.3% (m/m), 1.7% (m/m) and 2.6% (m/m) for B, R, W and RO quinoa samples, respectively.

2.2.3. SDS-PAGE

Fifteen microliters of sample solution (protein extracts from RO quinoa) were reduced and denatured with 5 µL of reducing-Laemmli sample buffer (1 M Tris titrated by HCl to pH 6.8, 4% (m/v) SDS, 20% (v/v) glycerol, 10% (v/v) β-mercaptoethanol and 1% (v/v) bromophenol blue). SDS-PAGE was performed on a vertical system Mini-PROTEAN® Tetra Cell with a PowerPac™ HC Power Supply (Bio-Rad, Hercules, CA, USA). Samples were loaded using in-house 10% SDS-polyacrylamide gels (30% (v/v) acrylamide/bis solution, 1.5 M Tris titrated by HCl to pH 8.8, 10% (m/v) SDS, 10% (m/v) APS and TEMED (4 µL/10 mL)). Ten microliters of protein ladder (BenchMark™ Protein Ladder) were also loaded for molecular mass calibration. Gel electrophoresis was performed at 120 V for 2 h at RT. The gel was fixed in 40% (v/v) ethanol and 10% (v/v) HAc for 30 min. Then, it was rinsed with water (3 × 5 min) and incubated in a blue Coomassie staining solution for 24 h. After that, the gel was rinsed with water (3 × 20 min) until a proper degree of staining was achieved.

2.2.4. LC-MS/MS

Appropriate volumes (from 85 to 130 μL, which is to say, 50 μg of total protein estimated by CE-UV) of the quinoa protein extracts (three independent replicates from B, R, W and RO quinoa) were evaporated to dryness using a Savant SPD-111 V SpeedVac concentrator (Thermo-Fisher Scientific) and suspended in 100 μL of ice-cold extraction buffer (25 mM HEPES (pH 8.0), 1.5 M urea, 0.02% (v/v) Triton™ X-100 and 5% (v/v) glycerol). The suspension was vortexed for 2 min and centrifuged for 30 s at $5,000 \times g$. Samples were reduced by addition of 3 mM TCEP for 45 min at RT and then alkylated with 15 mM IAA for 60 min in the dark at RT. Proteolytic digestion was performed by addition of 300 ng of trypsin/Lys-C mix (enzyme:protein ratio 1:167 m/m) and incubated under shaking at $500 \times g$ at RT overnight. The digestion was stopped by addition of FA (1% (v/v) final concentration) and centrifuged at $15,000 \times g$ for 2 min. The supernatant containing the digested proteins was desalted on disposable TopTip C-18 columns (Glygen, Columbia, MD, USA) and was evaporated to dryness.

All experiments were performed on an Orbitrap Fusion™ Lumos™ (Thermo Scientific) coupled to an Ultimate3000 nanoRLSC (Thermo Scientific). Protein digests were reconstituted in 20 μL of water containing 1% of FA (v/v) and separated on a column (15 cm × 75 μm ID × 365 μm OD fused silica capillary, Polymicro Technologies) in-house packed with C18 particles (Luna C18(2), 3 μm, 100 Å, Phenomenex, Torrance, California, USA) using a water/ACN/0.1% (v/v) FA linear gradient at a flow rate of 0.30 μL/min (0–7 min, 2–2% ACN; 7–77 min, 2–38% ACN; 77–86 min, 38–98% ACN; 86–96 min, 98–98% ACN; 96–99 min, 98–2% ACN; 99–109 min, 2–2% ACN). Two μL of sample were injected. The Orbitrap parameters in ESI + were as follows: ion source temperature 250 °C, ion spray voltage 2.1 kV, top speed mode, full-scan MS spectra (*m/z* 350–2,000) acquired at a resolution of 60,000. Precursor ions were filtered according to monoisotopic precursor selection, charge state (+2 to + 7), and dynamic exclusion (30 s with a ± 10 ppm window). The automatic gain control settings were $5*10^5$ for full scan and $1*10^4$ for MS/MS scans. Fragmentation was performed with collision-induced dissociation (CID) in the linear ion trap. Precursors were isolated using a 2 *m/z* isolation window and fragmented with a normalized collision energy of 35%.

2.2.5. Data analysis

MaxQuant (Thermo Scientific, version v1.6.17.0) (Cox & Mann, 2008) with the search engine Andromeda (Cox et al., 2011) was applied for protein and peptide identification for all MS raw files (three independent replicates from B, R, W and RO quinoa). Enzymatic digestion with trypsin was selected, together with a maximum of two missed cleavages, peptide charges from + 2 to + 7, a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.5 Da. Search parameters were set to allow for dynamic modifications of methionine oxidation, acetyl on *N*-terminus, and fixed cysteine carbamidomethylation. The search database consisted of a non-redundant quinoa protein sequence FASTA file containing the 63,370 entries from *Chenopodium quinoa* found in RefSeq NCBI database (FASTA file is provided as Supplementary material). The false discovery rate (FDR) was set to 0.01 for both peptide and protein identifications. Normalized label-free quantification (LFQ) values were obtained by applying the in-built MaxLFQ algorithm (Cox et al., 2014).

MaxQuant normalized LFQ intensities of identified proteins in all quinoa varieties were visualized as a heat map, created using the freely available web server Heatmappper (http://www.heatmapper.ca). The identified proteins were also classified by Gene Ontology (GO) using the PANTHER classification system (http://www.pantherdb.org). However, as *Chenopodium quinoa* is not available in the PANTHER-GO system, which works primarily with UniProt identifiers and modeled organisms (Mi, Muruganujan, Casagrande, & Thomas, 2013), the NCBI accession numbers (IDs) of the identified proteins were blasted (Pertsemlidis & Fondon, 2001) against the Uniprot database of *Arabidopsis thaliana*, a widely recognized model plant organism.

3. Results and discussion

3.1. SDS-PAGE analysis

Before shotgun proteomics experiments, SDS-PAGE was used to investigate the protein banding patterns obtained for quinoa seed extracts with three different methods of protein precipitation (Aloisi et al., 2016; Capriotti et al., 2015; Galindo-Luján et al., 2021). Using a RO quinoa sample as a model, precipitation with acetone, alkaline extraction with NaOH and alkaline extraction with NaOH followed by isoelectric precipitation with HCl were tested. As shown on the gel image in Fig. 1 A-C, the protein profiles were similar with all the extraction procedures and proteins were resolved into distinct bands that spanned a broad range of relative molecular masses (M_r) from 6,000 to 65,000. These results are in consensus with previous studies, such as those reported by Piñuel et al. (Piñuel et al., 2019), which assigned the most abundant bands around 10,000, 15,000–35,000 and 50,000 M_r to 2S albumins, 11S globulins and 7S globulins, respectively. Since there were no significant differences in the number of protein bands observed for the three extraction protocols (only acetone precipitation produced seemingly lower intensities, see Fig. 1-A) and also considering that the alkaline extraction followed by isoelectric precipitation was supposed to provide the most purified protein extracts, this extraction procedure was applied before performing shotgun proteomics.

3.2. Shotgun proteomics

3.2.1. LC-MS/MS analysis

The most commonly used approach for label-free shotgun proteomics is to enzymatically digest proteins into peptides, which are then analyzed by LC-MS/MS. Therefore, efficient protein digestion is critical to the successful identification of proteins. Typically, the solubilization of proteins using buffers that contain salts, denaturing agents and detergents (e.g. urea and Triton™ X-100, among others) and the efficient elimination of these additives from the resulting samples are crucial steps in shotgun proteomics (Tubaon, Haddad, & Quirino, 2017).

Other critical consideration for the successful identification of proteins is the use of a high-performance mass spectrometer. In this study, the Orbitrap Fusion™ Lumos™ mass spectrometer was used for the characterization of the quinoa seed proteome. This mass spectrometer has significantly boosted the sensitivity and sequencing speed, compared to earlier generation Orbitraps (e.g. Linear Trap Quadrupole-Orbitrap), leading to an increase in protein identifications (Zhu et al., 2018). LC-MS/MS conditions normally used for the analysis of peptides in shotgun proteomics were selected, as they are general enough to allow a proper chromatographic separation and detection sensitivity in positive ESI mode for a wide range of complex protein digests using different mass spectrometers. Under these conditions, three independent protein extracts from B, R, W and RO quinoa were analyzed by LC-MS/MS and the raw data files were subjected to data analysis for peptide and protein identification.

3.2.2. Data analysis

Data analysis in shotgun proteomics is much more challenging than for other high-throughput technologies and remains a principal bottleneck in proteomics (Lereim, Oveland, Berven, Vaudel, & Barsnes, 2016; Nesvizhskii, Vitek, & Aebersold, 2007; Sinitcyn, Rudolph, & Cox, 2018). Due to the high complexity of the samples resulting from the enzymatic digestion of proteins, computational proteomics has become a key research area for protein identification and quantification. MaxQuant is a computational proteomics workflow that addresses the above tasks with a focus on high accuracy and quantitative data (Cox et al., 2011, 2014; Cox & Mann, 2008).

In this study, MaxQuant/Andromeda in combination with a non-redundant quinoa protein sequence FASTA file (containing 63,370 entries) from the RefSeq NCBI database was applied for protein
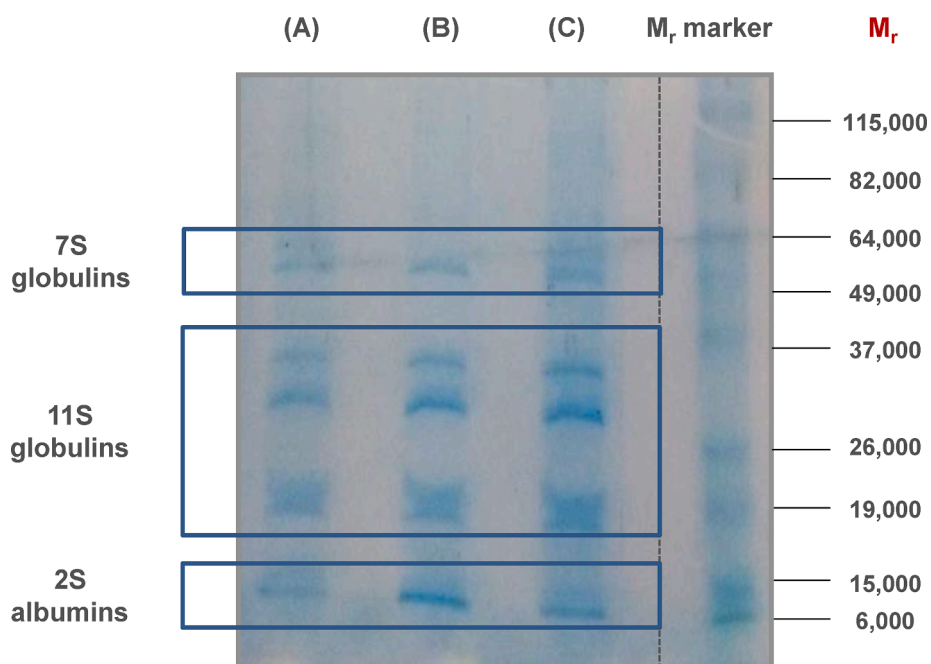
Fig. 1. SDS-PAGE for a RO quinoa sample after protein extraction using (A) precipitation with acetone, (B) alkaline extraction with NaOH and (C) alkaline extraction with NaOH followed by isoelectric precipitation with HCl. Migration region with the bands for the different quinoa proteins are indicated with a box: 10,000 M_r for 2S albumins, 15,000–35,000 M_r for 11S globulins and 50,000 M_r for 7S globulins.
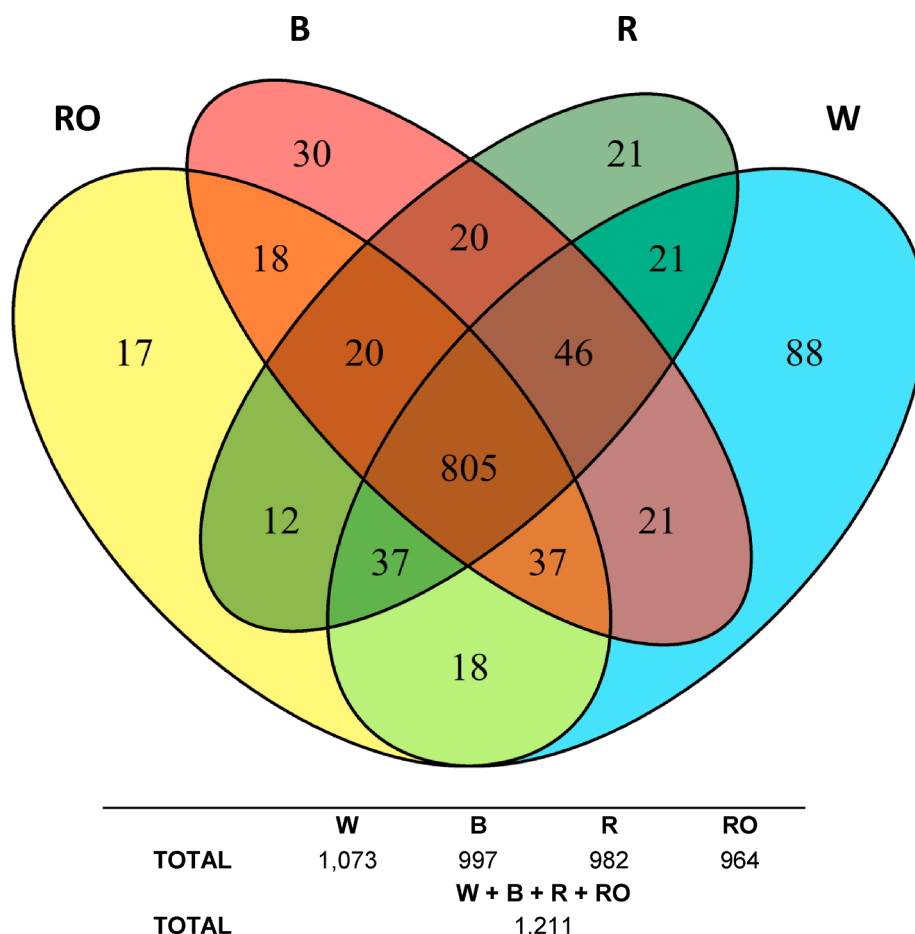

	W	B	R	RO
TOTAL	1,073	997	982	964
		W + B + R + RO		
TOTAL		1,211		

Fig. 2. Venn diagram of the identified proteins in the four analyzed quinoa varieties, W, B, R and RO.

identification and LFQ of all the raw data files. Supplementary Table S-1 shows the protein group level, the accession number (ID), the protein name, the M_r, the Andromeda score, the number of peptides, the sequence coverage and the normalized LFQ intensity for the identified proteins in W, B, R and RO samples (definition of these parameters is provided as a footnote in Supplementary Table S-1). It is worth mentioning that for every quinoa variety, only proteins found in at least two out of the three replicates are reported. Additionally, the number of peptides, the sequence coverage and the normalized LFQ intensity obtained for all the quinoa varieties is presented as an average value for the different protein extract samples (in all cases, relative standard deviation (%RSD) was lower than 10%). As shown in Supplementary Table S-1, a total number of 1,211 proteins (taking into account all the quinoa varieties) could be identified using the described label-free shotgun proteomics approach. This is a significantly larger number of proteins

than the 337 reported by Burrieza et al. (Burrieza et al., 2019), who were using a different protein extraction protocol, mass spectrometer and data analysis approach with a *Chenopodium quinoa* protein NCBI database of 63,459 entries. In Supplementary Table S-1, proteins are ordered by the Andromeda score that is considered the most important parameter to reflect the reliability of the identification in this environment (Cox et al., 2011; Cox & Mann, 2008). As can be observed, the Andromeda score for the identified proteins ranged between 323 and 2, being more reliable proteins with higher Andromeda scores than those with lower scores. In addition, in certain protein group levels, there are several proteins that cannot be distinguished based on their peptide content, for instance proteins in the group level 3 (with IDs XP_021715439.1 and XP_021720768.1, see Supplementary Table S-1). MaxQuant identification and quantification is reported at the group level to avoid overcounting identification and ambiguous
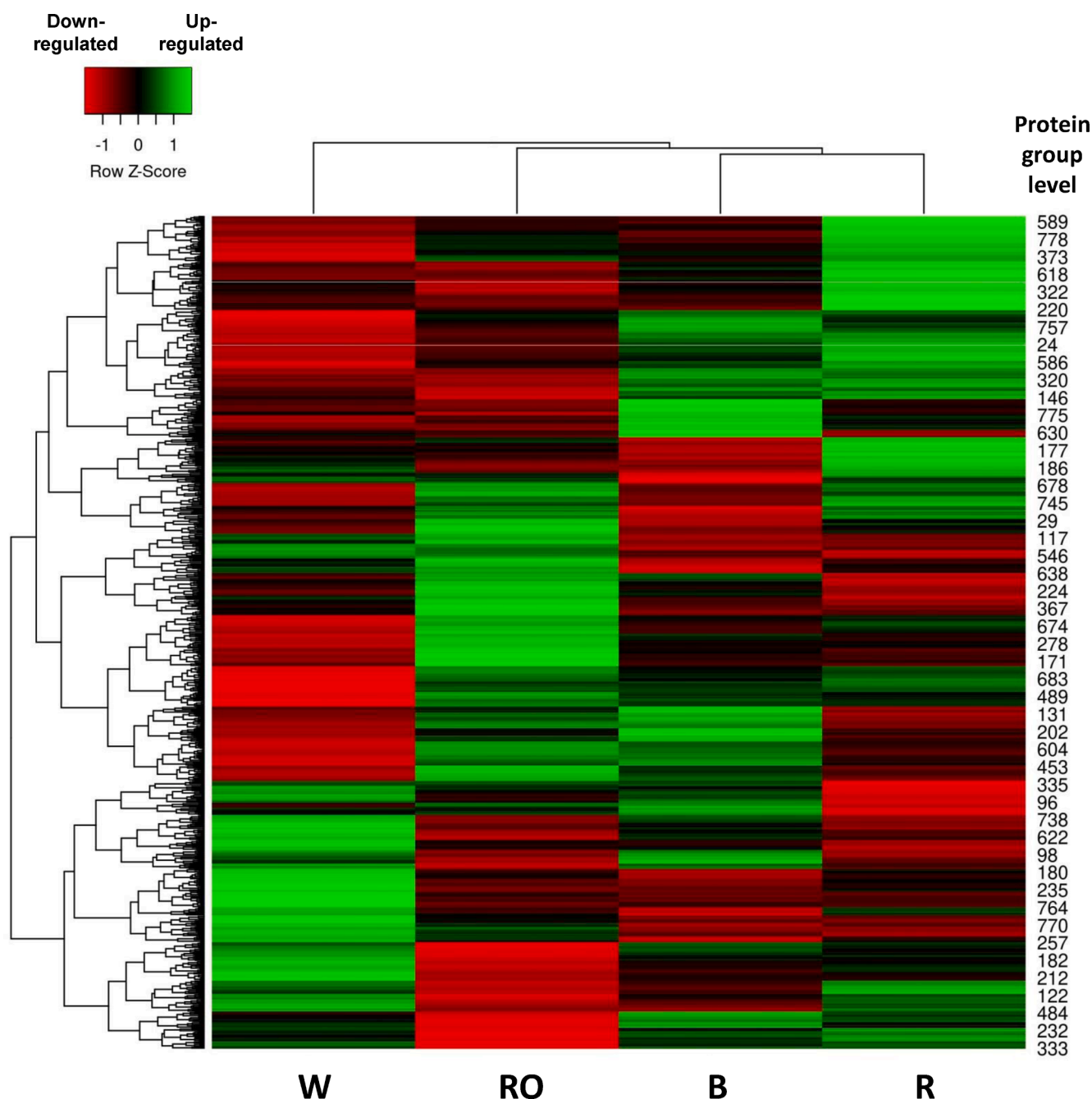
Fig. 3. Heat map obtained using the row z-score normalized LFQ intensities of the identified quinoa proteins in the four analyzed varieties, W, B, R and RO.

quantification. The Venn diagram in Fig. 2 shows the relationships between the identified proteins for W, B, R and RO quinoa. As can be seen in this figure, a similar total number of proteins were identified in W, B, R and RO quinoa varieties (i.e 1,073, 997, 982 and 964, respectively, from a total of 1,211 proteins identified considering the four varieties). Among them, 805 proteins (66% of the total) were identified in all the varieties, while 406 proteins (34% of the total) were only present in some of them. Regarding proteins identified in only one variety, 88 proteins were exclusively identified in W, 30 in B, 21 in R and 17 in RO quinoa (see Fig. 2 and Supplementary Table S-1). These observations suggested that there were differences between the proteomes of the four studied quinoa varieties.

Although the Venn diagram allowed visualizing general relationships in the number of proteins identified in the quinoa varieties, it was necessary to consider differences at the concentration level for a confident discrimination. A Euclidean distance heat map graph (Fig. 3) was constructed from the data matrix of average normalized LFQ intensities of the 1,211 identified proteins (rows) in the four quinoa varieties (columns). Proteins were filtered for complete observations (805 proteins), and z-scores (normalized per protein) were calculated by substracting the mean and dividing by the standard deviation values. In a heat map, the rows and columns are reordered to keep closer those with similar profiles and each row z-score entry in the data matrix is displayed as a color, making it possible to view the relationships and patterns graphically (Benno Haarman et al., 2015; Key, 2012; Krentzman, Robinson, Jester, & Perron, 2011). As can be observed in Fig. 3, each variety presented a characteristic concentration profile, with green, red and black boxes representing up-regulated, down-regulated and unchanged expression proteins, respectively. Most heat maps use an agglomerative hierarchical clustering algorithm to group the data according to the observed characteristic profiles, and display this information using a dendrogram. When two clusters are merged, a line is drawn connecting the two clusters at a height corresponding to how similar the clusters are. As it is shown in Fig. 3, B and R quinoa samples were clustered together, followed by RO and, finally, W quinoa, which, as shown by the clusters, was the least closely related variety based on the quantified protein groups. The differences and similarities clustering the proteomic profiles of the studied quinoa varieties could be attributed to the genetic features of each variety, as well as to the cultivar agroecological conditions. With regard to the observed variety clusters, the results are complementary to our previous work based on protein fingerprinting by capillary electrophoresis with ultraviolet absorption diode array detection (CE-UV-DAD) and advanced chemometrics (Galindo-Luján et al., 2021), where W and RO quinoa showed a high degree of similarity and were discriminated from B and R quinoa. The differences identified between both approaches may be attributed to the information provided by the applied analytical techniques. While results obtained with CE-UV-DAD were more focused on global differences found at the high-abundance protein level, shotgun proteomics provided detailed information about the identity and the abundance of the proteins from the proteomic profiles getting a deeper biological insight.

In order to have a global overview of the type of identified proteins, they were also classified in four broad categories by Gene Ontology (GO) (using the PANTHER-GO system): cellular component, molecular function, biological process and protein class (see Fig. 4 A-D, respectively). However, the PANTHER-GO system works primarily with UniProt identifiers and the *Chenopodium quinoa* Uniprot database contains limited amount of data compared to the NCBI database. Therefore, NCBI IDs of the identified proteins were blasted against the Uniprot database of *Arabidopsis thaliana*, a widely recognized model plant organism. After blasting, uncharacterized proteins from quinoa (85, see Supplementary Table S-1) did not match *Arabidopsis thaliana* entries from the Uniprot database and they were discarded for GO annotation. Supplementary Table S-2 shows the 1,126 quinoa proteins considered for GO annotation and its correspondence with 1,085 Uniprot IDs from *Arabidopsis thaliana* (average homology degree was 73%±16% (±s)), which were subjected to PANTHER-GO analysis and classification. In the cellular component category (Fig. 4-A), 489 hits were localized in a cellular anatomical entity (42%), 481 hits in an intracellular part (41%) and 204 hits in a protein-containing complex (17%). In the molecular function category (Fig. 4-B), 331 hits were associated with catalytic activity (46%) and 254 with binding (35%). The rest of hits were associated with structural molecule activity (83 hits), translation regulatory activity (19 hits), transporter activity (19 hits) and molecular function regulation (17 hits) (11%, 3%, 3% and 2%, respectively, Fig. 4-B). In the biological process category, the highest percentage of identified proteins (80%) was involved in cellular and metabolic processes (431 and 349 hits, respectively, Fig. 4-C). The rest of hits were associated with response to stimulus (65 hits), biological regulation (59 hits), localization (51 hits) and other biological processes which represent less than 3%. Finally, in the protein class category, 65% of the identified proteins were classified as metabolite interconversion enzymes and translational proteins (286 and 144 hits, respectively, Fig. 4-D). The rest of hits were classified as protein modifying enzymes (69 hits), chaperones (38 hits), transporters (30 hits), nucleic acid metabolism proteins (21 hits), chromatin/chromatin-binding or regulatory proteins (16 hits), cytoskeletal proteins (12 hits), protein-binding activity modulators (12 hits), membrane traffic proteins (12 hits), and other classes which represent less than 4%. In order to investigate the characteristic features of the small number of proteins exclusively identified in W, B, R and RO quinoa (156 proteins, see the Venn diagram of Fig. 2), GO graphs were represented also for these proteins (Supplementary Figure S-1). However, no significant differences were observed with regard to the GO graphs of the total proteins (Fig. 4), suggesting that the most differential proteins from the four quinoa varieties were related to similar cellular components, molecular functions, biological processes and protein classes.

4. Conclusions

We presented a label-free LC-MS/MS shotgun proteomics approach with a state-of-the-art Orbitrap mass spectrometer for the characterization and differentiation of the most typically commercialized quinoa seed varieties (B, R, W and RO quinoa). A total of 1,073, 997, 982 and 964 quinoa proteins from a non-redundant NCBI quinoa database were identified in the four varieties (1,211 identified proteins in total, 85 of them uncharacterized), significantly improving the quinoa proteome coverage from previous studies. In order to investigate relationships between B, R, W and RO quinoa varieties at the proteome level, Venn diagrams, heat maps and GO classification graphs were represented. As indicated above, a similar number of proteins were identified for each quinoa variety (~1,000), and they were similar with regard to GO annotation. However, the characteristic concentration profiles of the identified proteins were useful to find relationships and discriminate between the four varieties. The study provides the most comprehensive experimental quinoa proteomic map existing to date in the literature that can be used for more specific characterization and nutritional studies on the identified proteins. Proteomic profiling of quinoa seeds may be used for quality control of quinoa and quinoa-containing foodstuff, as well as aid in the enhancement of quinoa seed nutritional value or technological properties. Additionally, it may find applicability in the improvement of industrial processing procedures or cultivar yields under different agroecological conditions.

CRediT authorship contribution statement

Rocío Galindo-Luján: Methodology, Investigation, Writing - original draft, Writing - review & editing. **Laura Pont:** Investigation, Writing - original draft, Writing - review & editing, Supervision. **Zoran Minic:** Methodology, Writing - review & editing. **Maxim V. Berezovski:** Conceptualization, Writing - review & editing, Funding acquisition. **Victoria Sanz-Nebot:** Conceptualization, Writing - review & editing, Supervision, Funding acquisition. **Fernando Benavente:**
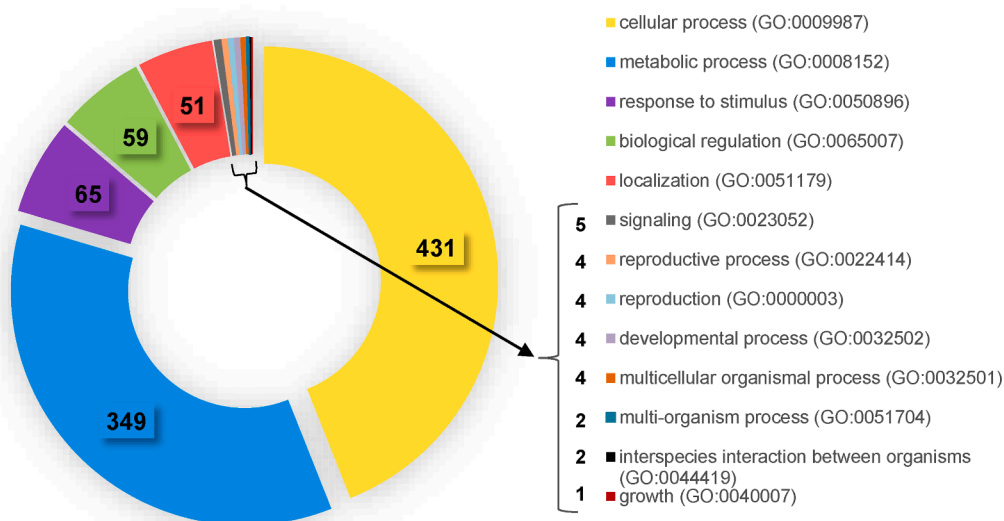
A) Cellular component
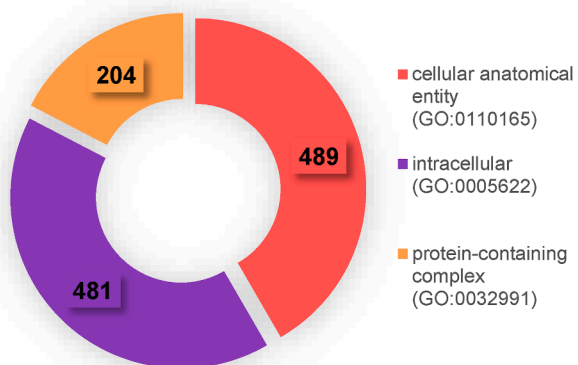
B) Molecular function
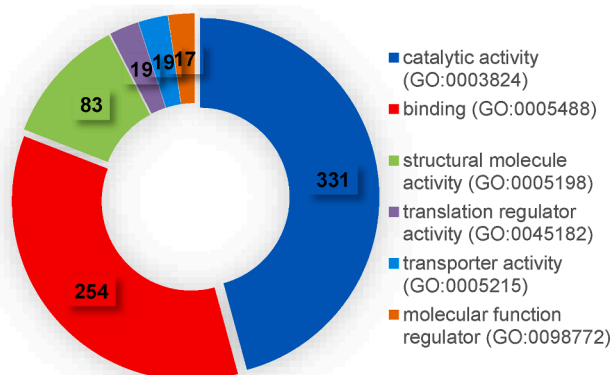
C) Biological process
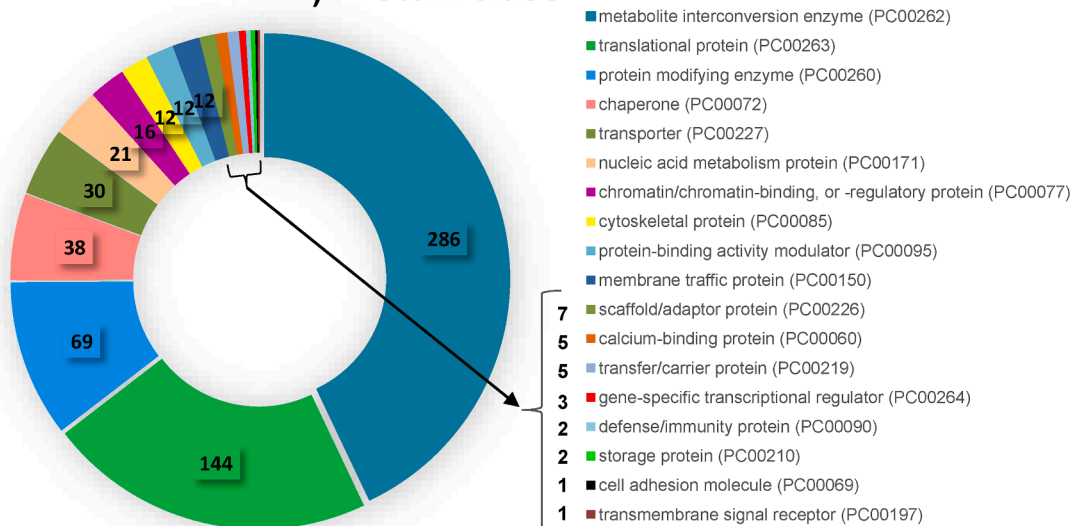
D) Protein class

Fig. 4. Gene Ontologies (GO) graphs classified by (A) cellular component, (B) molecular function, (C) biological process and (D) protein class for the total quinoa proteins blasted against the Uniprot database of *Arabidopsis thaliana*.

Conceptualization, Investigation, Writing - review & editing, Supervision, Funding acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.130250.

References

Aizat, W. M., & Hassan, M. (2018). Proteomics in systems biology. *Advances in Experimental Medicine and Biology, 1102*, 31–49. https://doi.org/10.1007/978-3-319-98758-3_3.

Aloisi, I., Parrotta, L., Ruiz, K. B., Landi, C., Bini, L., Cai, G., et al. (2016). New insight into quinoa seed quality under salinity: Changes in proteomic and amino acid profiles, phenolic content, and antioxidant activity of protein extracts. *Frontiers in Plant Science, 7*. https://doi.org/10.3389/fpls.2016.00656.

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research, 47*, D506–D515. https://doi.org/10.1093/nar/gky1049.

Benno Haarman, B. C. M., Riemersma-Van der Lek, R. F., Nolen, W. A., Mendes, R., Drexhage, H. A., & Burger, H. (2015). Feature-expression heat maps - A new visual method to explore complex associations between two variable sets. *Journal of Biomedical Informatics, 53*, 156–161. https://doi.org/10.1016/j.jbi.2014.10.003.

Burrieza, H. P., Rizzo, A. J., Moura, E., Silveira, V., & Maldonado, S. (2019). Shotgun proteomic analysis of quinoa seeds reveals novel lysine-rich seed storage globulins. *Food Chemistry, 293*, 299–306. https://doi.org/10.1016/j.foodchem.2019.04.098.

Burrieza, H. P., Rizzo, A. J., & Pérez, O. E. (2020). Quinoa does not contain prolamins. Comments on "Quinoa protein: Composition, structure and functional properties", Dakhili et al. (2019). *Food Chemistry, 325*, 126934. https://doi.org/10.1016/j.foodchem.2020.126934.

Capraro, J., De Benedetti, S., Di Dio, M., Bona, E., Abate, A., Corsetto, P. A., et al. (2020). Characterization of chenopodin isoforms from quinoa seeds and assessment of their potential anti-inflammatory activity in caco-2 cells. *Biomolecules, 10*(5), 795. https://doi.org/10.3390/biom10050795.

Capriotti, A. L., Cavaliere, C., Piovesana, S., Stampachiacchiere, S., Ventura, S., Zenezini Chiozzi, R., et al. (2015). Characterization of quinoa seed proteome combining different protein precipitation techniques: Improvement of knowledge of nonmodel plant proteomics. *Journal of Separation Science, 38*(6), 1017–1025. https://doi.org/10.1002/jssc.201401319.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular and Cellular Proteomics, 13*(9), 2513–2526. https://doi.org/10.1074/mcp.M113.031591.

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology, 26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., & Mann, M. (2011). Andromeda: A peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research, 10*(4), 1794–1805. https://doi.org/10.1021/pr101065j.

Dakhili, S., Abdolalizadeh, L., Hosseini, S. M., Shojaee-Aliabadi, S., & Mirmoghtadaie, L. (2019). Quinoa protein: Composition, structure and functional properties. *Food Chemistry, 299*, 1–10. https://doi.org/10.1016/j.foodchem.2019.125161.

Galindo-Luján, R., Pont, L., Sanz-Nebot, V., & Benavente, F. (2021). Classification of quinoa varieties based on protein fingerprinting by capillary electrophoresis with ultraviolet absorption diode array detection and advanced chemometrics. *Food Chemistry, 341*, 128207. https://doi.org/10.1016/j.foodchem.2020.128207.

Gao, Y., & Yates, J. R. (2019). Protein analysis by shotgun proteomics. *Mass Spectrometry-Based Chemical Proteomics*, 1–38. https://doi.org/10.1002/9781118970195.ch1.

Heck, M., & Neely, B. A. (2020). Proteomics in non-model organisms: A new analytical frontier. *Journal of Proteome Research, 19*(9), 3595–3606. https://doi.org/10.1021/acs.jproteome.0c00448.

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., et al. (2017). *The genome of Chenopodium quinoa, 542*, 307–311. https://doi.org/10.1038/nature21370.

Key, M. (2012). A tutorial in displaying mass spectrometry-based proteomic data using heat maps. *BMC Bioinformatics, 13*, 1–13. https://doi.org/10.1186/1471-2105-13-S16-S10.

Krentzman, A. R., Robinson, E. A. R., Jester, J. M., & Perron, B. E. (2011). Heat maps: A technique for classifying and analyzing drinking behavior. *Substance Use and Misuse, 46*(5), 687–695. https://doi.org/10.3109/10826084.2010.528126.

Lereim, R. R., Oveland, E., Berven, F. S., Vaudel, M., & Barsnes, H. (2016). Visualization, inspection and interpretation of shotgun proteomics identification results. *Advances in Experimental Medicine and Biology, 919*, 227–235. https://doi.org/10.1007/978-3-319-41448-5_11.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the panther classification system. *Nature Protocols, 8*(8), 1551–1566. https://doi.org/10.1038/nprot.2013.092.

Nesvizhskii, A. I., Vitek, O., & Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods, 4*(10), 787–797. https://doi.org/10.1038/nmeth1088.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research, 44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189.

Pereira, E., Encina-Zelada, C., Barros, L., Gonzales-Barron, U., Cadavez, V., & Ferreira, I. C. F. R. (2019). Chemical and nutritional characterization of Chenopodium quinoa Willd (quinoa) grains: A good alternative to nutritious food. *Food Chemistry, 280*, 110–114. https://doi.org/10.1016/j.foodchem.2018.12.068.

Pertsemlidis, A., & Fondon, J. W. (2001). Having a BLAST with bioinformatics (and avoiding BLAST phemy). *Genome Biology, 2*, 1–10. https://doi.org/10.1186/gb-2001-2-10-reviews2002.

Piñuel, L., Boeri, P., Zubillaga, F., Barrio, D. A., Torreta, J., Cruz, A., et al. (2019). Production of white, red and black quinoa (Chenopodium quinoa willd var. real) protein isolates and its hydrolysates in germinated and non-germinated quinoa samples and antioxidant activity evaluation. *Plants, 8*(8), 257. https://doi.org/10.3390/plants8080257.

Schubert, O. T., Röst, H. L., Collins, B. C., Rosenberger, G., & Aebersold, R. (2017). Quantitative proteomics: Challenges and opportunities in basic and applied research. *Nature Protocols, 12*(7), 1289–1294. https://doi.org/10.1038/nprot.2017.040.

Sinitcyn, P., Rudolph, J. D., & Cox, J. (2018). Computational methods for understanding mass spectrometry-based shotgun proteomics data. *Annual Review of Biomedical Data Science, 1*(1), 207–234. https://doi.org/10.1146/annurev-biodatasci-080917-013516.

Tubaon, R. M., Haddad, P. R., & Quirino, J. P. (2017). Sample clean-up strategies for ESI mass spectrometry applications in bottom-up proteomics: Trends from 2012 to 2016. *Proteomics, 17*(20), 1700011. https://doi.org/10.1002/pmic.v17.2010.1002/pmic.201700011.

Vega-Gálvez, A., Miranda, M., Vergara, J., Uribe, E., Puente, L., & Martínez, E. A. (2010). Nutrition facts and functional potential of quinoa (Chenopodium quinoa willd.), an ancient Andean grain: A review. *Journal of the Science of Food and Agriculture, 90*, 2541–2547. https://doi.org/10.1002/jsfa.4158.

Vilcacundo, R., & Hernández-Ledesma, B. (2017). Nutritional and biological value of quinoa (Chenopodium quinoa Willd.). *Current Opinion in Food Science, 14*, 1–6. https://doi.org/10.1016/j.cofs.2016.11.007.

Vilcacundo, R., Martínez-Villaluenga, C., & Hernández-Ledesma, B. (2017). Release of dipeptidyl peptidase IV, α-amylase and α-glucosidase inhibitory peptides from quinoa (Chenopodium quinoa Willd.) during in vitro simulated gastrointestinal digestion. *Journal of Functional Foods, 35*, 531–539. https://doi.org/10.1016/j.jff.2017.06.024.

Vilcacundo, R., Miralles, B., Carrillo, W., & Hernández-Ledesma, B. (2018). In vitro chemopreventive properties of peptides released from quinoa (Chenopodium quinoa Willd.) protein under simulated gastrointestinal digestion. *Food Research International, 105*, 403–411. https://doi.org/10.1016/j.foodres.2017.11.036.

Xu, D. (2012). Protein databases on the internet. *Current Protocols in Protein Science, 1*, 1–15. https://doi.org/10.1002/0471140864.ps0206s70.

Zhu, Y., Zhao, R., Piehowski, P. D., Moore, R. J., Lim, S., Orphan, V. J., et al. (2018). Subnanogram proteomics: Impact of LC column selection, MS instrumentation and data analysis strategy on proteome coverage for trace samples. *International Journal of Mass Spectrometry, 427*, 4–10. https://doi.org/10.1016/j.ijms.2017.08.016.

Zou, C., Chen, A., Xiao, L., Muller, H. M., Ache, P., Haberer, G., et al. (2017). A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Research, 27*(11), 1327–1340. https://doi.org/10.1038/cr.2017.124.