



26 **Non-standard abbreviations**

27

28 B: black quinoa

29 BGE: background electrolyte

30 CE-UV-DAD: capillary electrophoresis-ultraviolet absorption diode array detection

31 CE-UV: capillary electrophoresis-ultraviolet absorption detection

32 LV: latent variable

33 MCR-ALS: multivariate curve resolution alternating least squares

34 PC: principal component

35 PCA: principal component analysis

36 PLS-DA: partial least squares discriminant analysis

37 R: red quinoa

38 RO: royal white quinoa

39 UV: ultraviolet absorption detection

40 UV-DAD: ultraviolet absorption diode array detection

41 VIP: variable importance in the projection

42 W: white quinoa

43

44

45

46

47

48

49

50

## Highlights

- A novel CE-UV-DAD method to analyze quinoa soluble protein extracts is described.
- CE-UV-DAD fingerprints from different quinoa grain varieties are obtained.
- Different characteristic components are deconvoluted by MCR-ALS.
- Quinoa varieties are classified by PLS-DA from their differential protein composition.

51 **Abstract**

52

53 Quinoa (*Chenopodium quinoa* Willd.) is an andean grain with exceptional nutritional  
54 properties that has been progressively introduced in western countries as a protein-rich  
55 super food with a broad amino acid spectrum. Quinoa is consumed as whole grain, but it  
56 is also milled to produce high-value flour, which is susceptible to adulteration. Therefore,  
57 there is a growing interest in developing novel analytical methods to get further  
58 information about quinoa at the chemical level. In this study, we developed a rapid and  
59 simple capillary electrophoresis-ultraviolet absorption diode array detection (CE-UV-  
60 DAD) method to obtain characteristic multiwavelength electrophoretic profiles of soluble  
61 protein extracts from different quinoa grain varieties. Then, advanced chemometric  
62 methods (i.e. multivariate curve resolution alternating least squares, MCR-ALS, followed  
63 by principal component analysis, PCA, and partial least squares discriminant analysis,  
64 PLS-DA) were applied to deconvolute the components present in the electropherograms  
65 and classify the quinoa varieties according to their differential protein composition.

66

67

68

69

70

71 **1. Introduction**

72

73 Quinoa (*Chenopodium quinoa* Willd.) is an andean grain with more than 3,000 ecotypes  
74 recognized for its exceptional nutritional properties and its ability to adapt to very diverse  
75 agroecological conditions (Aloisi et al., 2016; Nowak, Du, & Charrondière, 2016; Pereira  
76 et al., 2019; Rojas, Alandía, Irigoyen, Blajos, & Santivañez, 2011; Vega-Gálvez et al.,  
77 2010). Quinoa has been progressively introduced in western countries, where it is sold as  
78 a gluten-free protein-rich super food with a broad amino acid spectrum. Quinoa is  
79 consumed as whole grain, but quinoa flour has been also receiving an increasing attention  
80 as a substitute for wheat flour in the food industry (González-Muñoz, Montero, Enrione,  
81 & Matiacevich, 2016; Laparra & Haros, 2018; Rodríguez, Rolandelli, & Buera, 2019).  
82 This growing interest in quinoa has raised the demand and consequently the prices,  
83 especially if grown organic, being a target for possible adulterations with cheaper cereals  
84 (Rodríguez et al., 2019).

85

86 Nowadays, there is a great interest in developing novel analytical methods for the reliable  
87 characterization of foodstuff as part of quality control, food safety and fraud control  
88 programs (Ojinnaka, 2016). A widespread strategy for this assessment is known as the  
89 fingerprint approach, which is based on obtaining a global profile of certain components  
90 by analytical techniques, such as spectroscopic, spectrometric, chromatographic or  
91 electromigration techniques (Álvarez, Montero, Llorens, Castro-Puyana, & Cifuentes,  
92 2018; Hong et al., 2017; Ropodi, Panagou, & Nychas, 2016). The targeted components  
93 may vary from small bioactive molecules, such as amino acids, organic acids, fatty acids  
94 or polyphenols to large biomolecules, such as proteins (Álvarez et al., 2018). However,

95 fingerprinting of intact proteins in food is specially challenging because of their size,  
96 structural complexity, wide concentration range and heterogeneity of the sample matrices.

97

98 One of the most widely applied fingerprinting techniques to characterize food ingredients,  
99 including proteins, is liquid chromatography with ultraviolet absorption detection (LC-  
100 UV) because of its simplicity, speed and separation performance (Gan et al., 2019;  
101 Jablonski, Moore, & Harnly, 2014; Li Vigni, Baschieri, Marchetti, & Cocchi, 2013;  
102 Rodríguez-Nogales, Cifuentes, García, & Marina, 2007). Protein fingerprinting of  
103 foodstuff by capillary electrophoresis with ultraviolet absorption detection (CE-UV) has  
104 also been demonstrated (Montealegre, García, Del Río, Marina, & García-Ruiz, 2012;  
105 Montealegre, Marina, & García-Ruiz, 2010; Sázelová, Kašička, Leon, Ibáñez, &  
106 Cifuentes, 2012; Vergara-Barberán, Lerma-García, Herrero-Martínez, & Simó-Alfonso,  
107 2014a; Vergara-Barberán, Mompó-Roselló, Lerma-García, Herrero-Martínez, & Simó-  
108 Alfonso, 2017), but to a lesser extent, despite its well-known potential for separation of  
109 complex mixtures of polar and charged compounds, such as peptides and proteins  
110 (Štěpánová & Kašička, 2019). CE-UV provides complementary and, very often, better  
111 separations than LC-UV. Additionally, analyses can be performed using smaller amounts  
112 of sample, operates with lower reagent consumption, no organic solvents are necessary,  
113 separation times are considerably low and it offers good repeatabilities (Heiger, 2010).

114

115 So far, in the typical LC-UV and CE-UV methods that have been described for protein  
116 fingerprinting of foodstuff, peak areas from single-wavelength chromatograms or  
117 electropherograms have been considered for characterization and classification purposes  
118 (Gan et al., 2019; Jablonski et al., 2014; Li Vigni et al., 2013; Montealegre et al., 2012,  
119 2010; Rodríguez-Nogales et al., 2007; Vergara-Barberán, Lerma-García, Herrero-

120 Martínez, & Simó-Alfonso, 2014b; Vergara-Barberán et al., 2017). However, the use of  
121 ultraviolet absorption diode array detection (UV-DAD) in combination with LC and CE  
122 allows the acquisition of three-way datasets (samples, elution/migration times and UV-  
123 spectra), which have proven to be an enhanced tool in profiling of other type of bioactive  
124 components in food and beverages, such as polyphenols in strawberry, olive oil and beer  
125 by LC-UV-DAD or CE-UV-DAD (Godoy-Caballero, Culzoni, Galeano-Díaz, & Acedo-  
126 Valenzuela, 2013; Mas, Fonrodona, Tauler, & Barbosa, 2007; Pérez-Ràfols & Saurina,  
127 2015). There are different data analysis procedures that allow processing of two-, three-  
128 and multi-way data sets (Escandar & Olivieri, 2019; Navarro-Reig, Bedia, Tauler, &  
129 Jaumot, 2018). Among them, multivariate curve resolution alternating least squares  
130 (MCR-ALS) offers several advantages (Jaumot, de Juan, & Tauler, 2015; Jaumot,  
131 Gargallo, de Juan, & Tauler, 2005), as it can resolve overlapped chromatographic or  
132 electrophoretic peaks from the collected data and provide the separation profiles and pure  
133 spectra of the constituents in the analyzed samples. This approach allows overcoming  
134 problems such as elution or migration time shifts, background noise contributions, and  
135 differences in signal-to-noise ratios (S/Ns) among different injections.

136

137 In this study, we describe for the first time a straightforward and simple procedure for  
138 protein fingerprinting of food based on the combination of CE-UV-DAD analysis of  
139 protein extracts and advanced chemometric tools. First, we have developed a CE-UV-  
140 DAD method to obtain characteristic multiwavelength electrophoretic profiles of soluble  
141 protein extracts from different quinoa grain varieties. Then, MCR-ALS has been used to  
142 deconvolute the components present in the CE-UV-DAD fingerprints, and unsupervised  
143 and supervised multivariate data analysis methods (i.e. principal component analysis  
144 (PCA) and partial least squares discriminant analysis (PLS-DA), respectively) have been

145 applied to classify and differentiate the quinoa varieties. The proposed methodology has  
146 allowed classifying the different quinoa varieties and providing a novel insight into their  
147 differential protein composition.

148

## 149 **2. Materials and methods**

### 150 **2.1. Chemicals and samples**

151

152 All the chemicals used in the preparation of buffers and solutions were of analytical  
153 reagent grade or better. Sodium hydroxide ( $\geq 99.0\%$ , pellets), hydrochloric acid (37%  
154 (v/v)), boric acid ( $\geq 99.5\%$ ) and sodium dodecyl sulfate (SDS,  $\geq 99.8\%$ ) were supplied by  
155 Merck (Darmstadt, Germany). Black (B, 6 samples), red (R, 6 samples) and white (W, 6  
156 samples) quinoa from Peru, as well as royal white (RO, 4 samples) from Bolivia were  
157 acquired in local supermarkets from Barcelona. Water with conductivity lower than 0.05  
158  $\mu\text{S}/\text{cm}$  was obtained using a Milli-Q water purification system (Millipore, Molsheim,  
159 France).

160

### 161 **2.2. Background electrolyte solution**

162

163 The background electrolyte (BGE) was prepared from a 60 mM  $\text{H}_3\text{BO}_3$  solution. The pH  
164 of this solution was adjusted to 9.0 with NaOH. Before the analyses, the BGE was  
165 degassed by sonication and filtered through a 0.20  $\mu\text{m}$  nylon filter (Macherey-Nagel,  
166 Düren, Germany).

167

### 168 **2.3. Apparatus and procedures**

169

170 pH measurements were made with a Crison 2002 potentiometer and a Crison electrode  
171 52-03 (Crison Instruments, Barcelona, Spain). Centrifugal filtration at a controlled  
172 temperature (4°C or 25°C) was carried out in a cooled Rotanta 460 centrifuge (Hettich  
173 Zentrifugen, Tuttlingen, Germany). Agitation was performed with a Vortex Genius 3  
174 (Ika<sup>®</sup>, Staufen, Germany). Incubations were carried out in a TS-100 thermoshaker  
175 (Biosan, Riga, Latvian Republic)

176

### 177 **2.3.1. Sample preparation**

178

179 Quinoa grains were dried in an air-current oven at 40°C for 24 h. The dried grains were  
180 ground in a coffee grinder and stored at room temperature in a desiccator. Before protein  
181 extraction, the total crude protein content in the quinoa samples was determined by the  
182 Kjeldahl method following the AOAC official method 2001.11 (conversion factor of N x  
183 6.25) (Nancy J Thiex, Harold Manson, Shirley Anderson, 2002). Protein extraction from  
184 quinoa grain was carried out following a procedure described elsewhere with some  
185 modifications (Aloisi et al., 2016; Giménez, Escudero, Mucciarelli, Luco, & de Arellano,  
186 2004). Briefly, 250 mg of the ground sample were mixed with 1 mL of water and 39 µL  
187 of 1 M NaOH (final pH was 10.0) and then incubated for 1 h at 36°C with constant shaking  
188 in a vortex. Separation of soluble proteins from the insoluble residue was performed by  
189 centrifugation at 15,000 x g for 20 min at 4°C. For protein purification, the supernatant  
190 pH was adjusted with 22 µL of 1 M HCl to obtain a final pH value of 5.0. After  
191 centrifugation at 15,000 x g for 20 min at 4°C, precipitated proteins were resuspended in  
192 1 mL of sodium borate BGE. The supernatant containing the extract of quinoa proteins  
193 was filtered through a 0.20 µm nylon filter before the analysis.

194

### 195 **2.3.2. CE-UV-DAD**

196

197 All CE-UV-DAD experiments were performed in a 7100 CE instrument (Agilent  
198 Technologies, Waldbronn, Germany). Separations were performed at 25°C in 58 cm total  
199 length ( $L_T$ ) (49.5 cm effective length,  $L_D$ )  $\times$  50  $\mu\text{m}$  internal diameter (i.d.)  $\times$  365  $\mu\text{m}$  outer  
200 diameter (o.d.) fused silica capillaries (Polymicro Technologies, Phoenix, AZ, USA). All  
201 capillary rinses were performed at high pressure (930 mbar). New fused silica capillaries  
202 were flushed with 1 M HCl (20 min), water (20 min), 1 M NaOH (20 min), water (20  
203 min) and BGE (20 min). The capillary was finally equilibrated by applying +25 kV  
204 (normal polarity, cathode in the outlet) for 30 min. Protein extracts were injected at 50  
205 mbar for 10 s. Between runs, capillaries were conditioned by rinsing with 0.5% SDS (m/v)  
206 (2 min), water (3 min), 1 M NaOH (3 min), water (3 min) and BGE (3 min). The UV-  
207 spectra were recorded scanning from 190 to 300 nm. Data acquisition was performed with  
208 ChemStation C.01.06 software (Agilent Technologies).

209

### 210 **2.4. Data analysis**

211

212 Experimental data were analyzed by a combination of advanced chemometric tools to  
213 deconvolute the CE-UV-DAD fingerprints, perform multivariate analysis and classify the  
214 different quinoa varieties. Figure 1 shows a summary of the data analysis workflow,  
215 which is explained in detail in this section. Data processing and graphical representation  
216 were performed under MATLAB R2016a (The Mathworks Inc., Natick, MA, USA).  
217 MCR-ALS GUI 2.0 (Jaumot et al., 2015) was run under MATLAB environment, and PLS  
218 toolbox (Version 8.1, Eigenvector Research Inc., Wenatchee, WA, USA) was used for  
219 PCA and, PLS-DA.

220

#### 221 **2.4.1. MCR-ALS**

222

223 First, CE-UV-DAD raw data were converted to comma-separated value (csv) format  
224 using a macro available with the ChemStation software and, then, imported into the  
225 MATLAB environment. The absorbance scale of the imported matrices was normalized  
226 against the maximum absorbance observed between 4 and 7 min at 214 nm, where the  
227 most intense peak corresponding to proteins was observed in all cases (Figure 2). Then,  
228 the normalized matrices were splitted into two submatrices corresponding to the time  
229 regions between 3-11 min and 11-21 min, which presented a differential and characteristic  
230 electrophoretic profile (Figure 2). No other data pre-processing was necessary before  
231 separately applying MCR-ALS to the set of submatrices from both time regions (Figure  
232 1-a and -b).

233

234 MCR-ALS is a decomposition method developed for the deconvolution of overlapping  
235 profiles into the individual contribution of the constituents (Jaumot et al., 2015, 2005). In  
236 case of CE-UV-DAD analysis, the MCR decomposition of a single DAD  
237 electropherogram can be mathematically expressed as follows:

238

$$239 \quad \mathbf{D} = \mathbf{CS}^T + \mathbf{E} \quad (\text{Eq. 1})$$

240

241 where  $\mathbf{D}$  is the data matrix representing the electrophoretic data (with as many rows as  
242 the number of sampled migration times and as many columns as the measured  
243 wavelengths), while  $\mathbf{C}$  and  $\mathbf{S}^T$  are the matrices collecting the resolved electrophoretic  
244 profiles, and the pure UV-spectra, respectively, of the components identified in the

245 mixture. The matrix  $\mathbf{E}$  contains the residuals, i.e., the fraction of the measured signal not  
246 accounted for by the MCR bilinear model.

247

248 The different samples can be simultaneously analyzed and compared by MCR-ALS using  
249 a column-wise augmented data matrix configuration (see matrix  $\mathbf{D}_{\text{aug}}$  in Eq. 2 and Figure  
250 1-c):

251

$$252 \quad \mathbf{D}_{\text{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_{15} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_{15} \end{bmatrix} \mathbf{S}^T + \begin{bmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_{15} \end{bmatrix} = \mathbf{C}_{\text{aug}} \mathbf{S}^T + \mathbf{E}_{\text{aug}} \quad (\text{Eq. 2})$$

253

254 This approach allows obtaining a common matrix of the pure UV-spectra of the resolved  
255 components ( $\mathbf{S}^T$ ) for all the samples, and a set of matrices describing the resolved  
256 electrophoretic profiles ( $\mathbf{C}_{\text{aug}}$ ) in every sample. These electrophoretic peaks resolved in  
257 matrix  $\mathbf{C}_{\text{aug}}$  are allowed to vary in position (shifts) and shape among samples because the  
258 only requirement for a proper resolution is that the resolved UV-spectra are the same for  
259 the common constituents in the different samples (Jaumot et al., 2015, 2005). This aspect  
260 is especially useful in the case of CE data where migration shifts among samples occur  
261 and, hence, the alignment of electrophoretic peaks before analysis is not needed. In this  
262 study, an independent  $\mathbf{D}_{\text{aug}}$  data matrix was built for each of the two selected time  
263 windows (Figure 1-c). MCR-ALS analysis was carried out following standard procedures  
264 for the determination of the number of components (singular value decomposition, SVD)  
265 and initial estimates (simple-to-use interactive self-modelling mixture analysis,  
266 SIMPLISMA). ALS optimization was performed under non-negativity constraints for  
267 electrophoretic ( $\mathbf{C}_{\text{aug}}$ ) and spectral ( $\mathbf{S}^T$ ) profiles, and spectral normalization (equal height)  
268 (Jaumot et al., 2015, 2005).

269

## 270 **2.4.2. Multivariate data analysis**

271

272 Once MCR-ALS was performed (Figure 1-c), the areas of the deconvoluted components  
273 and the protein content determined by the Kjeldahl method were considered for  
274 unsupervised and supervised multivariate data analysis, i.e. PCA and PLS-DA,  
275 respectively (Figure 1-d). First, PCA was applied to explore the classes present in the data  
276 and the presence of outliers (Joliffe & Morgan, 1992). PLS-DA was performed later to  
277 maximize class separation and rapidly classify the different samples (Barker & Rayens,  
278 2003), as well as to identify which components were the most significant to discriminate  
279 between these classes taking into account the variable importance in the projection (VIP)  
280 scores of the components in the PLS-DA model (Wold, Sjöström, & Eriksson, 2001). A  
281 (leave-one-out) cross validation of the PLS-DA model was performed during model  
282 optimization (Wold et al., 2001).

283

## 284 **3. Results and discussion**

### 285 **3.1 Analysis of quinoa samples by CE-UV-DAD**

286

287 Extraction of the most abundant proteins from quinoa grain was performed adapting a  
288 procedure described by Aloisi et al. (Aloisi et al., 2016). It was basically a protein  
289 solubilization at pH 10.0, followed by isoelectric precipitation at pH 5.0 and redissolution  
290 of the protein precipitate with the sodium borate separation BGE (pH 9.0). Under these  
291 conditions, the protein extract contained albumins and globulins that are the main seed  
292 storage protein fractions in quinoa grain (Aloisi et al., 2016). Specifically, Brinegar et al.  
293 reported that 11S globulin (chenopodin) and 2S albumin polypeptides represent 37 and

294 35% of total proteins, respectively (Brinegar & Goundan, 1993; Brinegar, Sine, &  
295 Nwokocho, 1996). Quinoa grain contains also a small amount of prolamins (Aloisi et al.,  
296 2016), but the concentration in the obtained protein extracts of seed storage proteins from  
297 the alcohol soluble fraction must be extremely low.

298

299 Some preliminary CE experiments were performed using a RO quinoa sample to select  
300 the most appropriate separation conditions to obtain the characteristic multiwavelength  
301 electrophoretic protein extract fingerprints. At first, the protein extract was prepared  
302 redissolving the proteins precipitated at pH 5.0 with a BGE of 60 mM Tris titrated to pH  
303 8.0 with HCl, as suggested by Aloisi et al (Aloisi et al., 2016), but repeatability of the  
304 electrophoretic separation was very low and this BGE was rapidly discarded. The BGE  
305 was changed to sodium borate (pH 9.0) prepared from a 60 mM  $\text{H}_3\text{BO}_3$  solution after  
306 adding NaOH. The good performance in CE-UV of BGEs based on borate buffers at pH  
307 values above the pI of the analyzed proteins is well-known (Heiger, 2010). At pH 9.0,  
308 protein adsorption on the bare fused silica inner capillary wall was minimized and buffer  
309 capacity was high because pH was very close to the  $\text{H}_3\text{BO}_3$  pK<sub>a</sub>. Further experiments were  
310 performed with sodium borate (pH 9.0) BGEs prepared from 100 and 150 mM  $\text{H}_3\text{BO}_3$   
311 solutions, but total analysis time increased, while the number of electrophoretic peaks  
312 decreased, and peak shape deteriorated. Therefore, the sodium borate (pH 9.0) BGE  
313 prepared from 60 mM  $\text{H}_3\text{BO}_3$  solution was selected as the best compromise between the  
314 quality of the electrophoretic profile and the total analysis time applying a voltage of 25  
315 kV. Under these conditions, it was only necessary to add to the typical capillary washing  
316 sequence with 1 M NaOH, water and BGE between consecutive analyses, an extra  
317 cleaning step with 0.5% (m/v) SDS to ensure appropriate separation repeatability. Figure  
318 2-a shows the electropherogram at 214 nm for the protein extract of a RO quinoa sample.

319 As can be observed, the complex electrophoretic profile contained different overlapped  
320 peaks and total analysis time was approximately 20 minutes. Repeatability was evaluated  
321 from 10 consecutive analyses. The relative standard deviation values (%RSD) for the  
322 three peaks labelled with numbers in the electropherogram of Fig. 2-a (peaks at around  
323 3.5, 7, and 15 min, labelled as 1, 2, and 3, respectively) ranged between 1 and 7% for  
324 migration times, and between 7 and 14% for peak areas.

325

326 All the quinoa samples were analyzed under the selected separation conditions for RO  
327 quinoa. Figure 2 also shows the electropherograms at 214 nm for the protein extracts of  
328 a W, a B and a R quinoa sample. As can be observed, the electrophoretic profiles for the  
329 four quinoa varieties presented similarities and differences. However, protein  
330 fingerprinting from the single-wavelength electropherograms was extremely difficult,  
331 because most of the peaks were overlapped and could not be accurately integrated. As an  
332 alternative, we explored the use of the multiwavelength electropherograms combined  
333 with advanced chemometrics methods for data deconvolution followed by multivariate  
334 data analysis for classification and differentiation of the quinoa varieties.

335

### 336 **3.2. MCR-ALS**

337

338 Before deconvolution with MCR-ALS, the raw multiwavelength electropherograms of  
339 the different quinoa samples were simply pre-processed by normalizing the absorbance  
340 scale, and no peak alignment or baseline correction were necessary. In order to minimize  
341 the processing time, while ensuring the good performance of the deconvolution algorithm  
342 and later classification, the normalized matrices were only splitted into two submatrices  
343 corresponding to the time regions between 3-11 min and 11-21 min, which comprised all

344 the detected peaks (Figure 2). Then, MCR-ALS was applied to two separate column-wise  
345 augmented data matrices containing simultaneously the information of the protein  
346 extracts from the 22 samples (6 B, 6 R, 6 W and 4 RO quinoa) to resolve the  
347 electropherogram profiles and the corresponding pure UV-spectra of the resolved  
348 components in both time regions. The number of components selected was lower than the  
349 number of electrophoretic peaks in each region, minimizing the possibility that some of  
350 the resolved components could be due to contributions such as solvent background or  
351 instrumental noise. In this case, two components in each time region allowed explaining  
352 almost 100% of variance (> 99.0% in both cases). As can be observed in Figure 3-a for a  
353 RO quinoa sample, C1 and C2 components were resolved in the first time window (from  
354 3 to 11 min), whereas C3 and C4 were resolved in the second time window (from 11 to  
355 21 min). Only C1 component appeared as a single electrophoretic peak, while the rest  
356 presented a profile with different electrophoretic peaks at lower intensities in the  
357 considered time regions. The studied time windows could have been divided in shorter  
358 time ranges to improve peak resolution, but at the cost of increasing the processing time  
359 and complicating the deconvolution procedure, which we conceived to be simple and  
360 straightforward. Figure 3-b shows the UV-spectra of the four resolved components in the  
361 wavelength range comprised between 190 and 300 nm. Proteins generally absorb strongly  
362 between 190 and 210 nm due to the peptide bonds. From this point of the UV-spectrum,  
363 absorbance decreases and shoulders can be observed at 230 nm due to the carboxylic acid  
364 moieties and again to the peptide bonds. If present, local absorbance maxima at 254 nm  
365 and 280 nm are due to the aromatic side chains of phenylalanine, tryptophan and tyrosine  
366 (Aitken & Learmonth, 1996; Liu, Avramova, & Park, 2009). As can be observed in Figure  
367 3-b, the four resolved components presented UV-spectra compatible with proteins, and  
368 are similar to those reported by CE-UV for olive proteins by Montealegre et al

369 (Montealegre et al., 2012, 2010). However, the presence of other UV-absorbing  
370 compounds such as polyphenols and flavonoids in the four components resolved to  
371 characterize the quinoa protein extracts could not be discarded (Aloisi et al., 2016).

372

### 373 **3.3. Multivariate data analysis. PCA and PLS-DA**

374

375 After MCR-ALS, multivariate data analysis was performed considering the areas of the  
376 four deconvoluted components (C1, C2, C3 and C4) in the protein extracts from the 22  
377 quinoa samples. The total protein content determined by the Kjeldahl method (Table 1)  
378 was also included as a variable to improve discrimination between the different quinoa  
379 varieties. First, we explored the data with PCA for the unsupervised identification of  
380 trends and clustering of the data, as well as outliers. Two principal components (PCs)  
381 allowed explaining 92.4% of the variance (Supplementary Figure S-1). As can be  
382 observed in this figure, PC1 (69.9% of the explained variance) clearly separated R quinoa  
383 samples from the other varieties. Meanwhile, PC2 (22.5% of the explained variance)  
384 allowed a slight separation between B quinoa and the group formed by W and RO quinoa.  
385 This last grouping suggested that the protein extract of RO quinoa, which is a W quinoa  
386 variety from Bolivia, presented similar composition to the protein extract of W quinoa.  
387 Additionally, PCA allowed detecting two W quinoa samples as outliers (W5 and W6,  
388 Supplementary Figure S-1), which were discarded before applying PLS-DA.

389

390 Once the data were explored and three classes defined (i.e. B, R and white-royal (W-RO)  
391 quinoa) by PCA, PLS-DA was applied to build a refined classification model with  
392 improved class separation and to reveal the importance of the different components for  
393 discrimination between the groups of samples. As can be observed in the scores plot of

394 Figure 4-a, a PLS-DA model with two latent variables (LVs) allowed a perfect  
395 discrimination between the three quinoa classes (92.5% of X-variance and 47.7% of Y-  
396 variance explained). Sensitivity, specificity and classification error in the calibration and  
397 (leave-one-out) cross-validation were excellent. The loadings plot (Figure 4-b) showed  
398 the contribution of the different variables (the four MCR-ALS resolved components and  
399 the total protein content determined by the Kjeldahl method) to the LVs. As can be seen  
400 in this plot, the total protein content was responsible for the separation of B quinoa from  
401 the other quinoa varieties, while the resolved MCR-ALS components allowed the  
402 separation of R quinoa from the other quinoa varieties. In contrast to PCA, the VIP scores  
403 allowed to quantify the influence of the different variables on separation between the  
404 quinoa varieties. The bar plots of Figure 5-a-c show the VIP scores of the different  
405 variables when considering separation of W-RO, B and R quinoa samples from the rest  
406 of classes, respectively. VIP scores estimated the importance of the variables in the  
407 projection and only those with a VIP score over a particular threshold value (usually 1)  
408 were considered important for discrimination (Wold et al., 2001). As can be observed in  
409 the VIP scores plots of Figure 5-a, the total protein content and C3 component (in a minor  
410 extent) were the most important variables for discrimination of W-RO quinoa from B and  
411 R quinoa. The total protein content was also the most important variable to discriminate  
412 B quinoa samples (Figure 5-b). The importance of the total protein content in both cases  
413 could be due to the fact that B and W-RO quinoa samples showed the highest (16.0%  
414 (m/m)) and the lowest (14.8-14.9% (m/m)) total protein content values, respectively  
415 (Table 1). In contrast, the total protein content was not important for discrimination of  
416 red quinoa from the rest of quinoa samples (Figure 5-c). In this case, components C1, C2  
417 and C3 showed to be the most important variables. Therefore, overall it was found that  
418 the component C4, which was a minor component of the protein extracts (see Figures 2

419 and 3-a), was the only variable non-critical for differentiation. The proposed PLS-DA  
420 model allowed rapidly classifying the quinoa varieties, as well as providing information  
421 about the importance of the different protein compositional variables.

422

#### 423 **4. Conclusions**

424

425 We have demonstrated that protein fingerprinting by CE-UV-DAD combined with  
426 advanced chemometric methods is an excellent approach to discriminate between  
427 different quinoa varieties, as well as for getting further insight on protein composition.  
428 After a rapid and simple protein extraction method, CE-UV-DAD was applied to obtain  
429 multiwavelength electrophoretic fingerprints of soluble protein extracts from B, R, W and  
430 RO quinoa samples. Deconvolution with MCR-ALS allowed the resolution of the most  
431 relevant components in the electrophoretic profiles, which showed characteristic UV-  
432 spectra. The areas of the four resolved components and the total protein content  
433 determined by the Kjeldahl method were considered for PCA and PLS-DA. PCA allowed  
434 detecting two white quinoa outlier samples and defining three sample classes (i.e. B, R  
435 and W-RO quinoa). PLS-DA improved sample classification and revealed that  
436 component C4 was not significant for the discrimination. The proposed methodology  
437 demonstrated its potential to rapidly obtain a reliable classification of quinoa varieties  
438 based on protein fingerprinting, and could be used for a simple and enhanced quality  
439 control of quinoa-containing foodstuff. In the future, the approach could be further  
440 validated with larger sample sets of quinoa varieties or ecotypes, which could be also  
441 grown under different conditions (e.g. ecological, salinity, etc). More widely a similar  
442 approach could find application to protein fingerprinting of other foodstuff, presenting  
443 complex electrophoretic profiles with highly overlapped peaks.

444

## 445 **Acknowledgements**

446

447 This study was supported by the Spanish Ministry of Economy and Competitiveness  
448 (RTI2018-097411-B-I00) and the Cathedra UB Rector Francisco Buscarons Úbeda  
449 (Forensic Chemistry and Chemical Engineering). RG thanks the National Agricultural  
450 Innovation Program and the Ministry of Education from Peru for a research stay and a  
451 PhD fellowship, respectively.

452

453 The authors declare no conflicts of interest.

454

## 455 **Figure legends**

456

457 **Figure 1.** Workflow for the analysis and classification of quinoa varieties by CE-UV-  
458 DAD in combination with advanced chemometric tools.

459

460 **Figure 2.** Electropherograms obtained after protein extraction for (a) royal white (RO),  
461 (b) white (W), (c) black (B) and (d) red (R) quinoa samples (at 214 nm). Peaks labelled  
462 as 1, 2 and 3 in (a) were considered for the repeatability studies.

463

464 **Figure 3.** (a) MCR-ALS resolved concentration profiles obtained for the 4 components  
465 of a royal white (RO) quinoa sample and (b) their corresponding pure UV-spectra.

466

467 **Figure 4.** (a) Scores plot and (b) loadings plot of the PLS-DA model applied to the 20  
468 selected quinoa samples using the 4 components resolved by MCR-ALS and the total

469 protein content determined by the Kjeldahl method. (royal white (RO), white (W), black  
470 (B) and red (R) quinoa)

471

472 **Figure 5.** VIP scores of the different variables when considering the separation of the  
473 different quinoa classes (a) white-royal (W-RO) from black (B) and red (R), (b) B from  
474 W-RO and R and (c) R from W-RO and B.

475

476 **References**

477

478 Aitken, A., & Learmonth, M. (1996). Protein determination by UV absorption. *The*  
479 *Protein Protocols Handbook*, 3–6. [https://doi.org/10.1007/978-1-60327-259-9\\_1](https://doi.org/10.1007/978-1-60327-259-9_1)

480 Aloisi, I., Parrotta, L., Ruiz, K. B., Landi, C., Bini, L., Cai, G., ... Del Duca, S. (2016).  
481 New insight into quinoa seed quality under salinity: changes in proteomic and amino  
482 acid profiles, phenolic content, and antioxidant activity of protein extracts. *Frontiers*  
483 *in Plant Science*, 7, 1–21. <https://doi.org/10.3389/fpls.2016.00656>

484 Álvarez, G., Montero, L., Llorens, L., Castro-Puyana, M., & Cifuentes, A. (2018). Recent  
485 advances in the application of capillary electromigration methods for food analysis  
486 and Foodomics. *Electrophoresis*, 39, 136–159.  
487 <https://doi.org/10.1002/elps.201700321>

488 Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of*  
489 *Chemometrics*, 17, 166–173. <https://doi.org/10.1002/cem.785>

490 Brinegar, C., & Goundan, S. (1993). Isolation and characterization of chenopodin , the  
491 11S seed storage protein of quinoa (*Chenopodium quinoa*). *Journal of Agricultural*  
492 *and Food Chemistry*, 41, 182–185. <https://doi.org/10.1021/jf00026a006>

493 Brinegar, C., Sine, B., & Nwokocha, L. (1996). High-cysteine 2S seed storage proteins  
494 from quinoa (*Chenopodium quinoa*). *Journal of Agricultural and Food Chemistry*,  
495 44, 1621–1623. <https://doi.org/10.1021/jf950830+>

496 Escandar, G. M., & Olivieri, A. C. (2019). Multi-way chromatographic calibration—A  
497 review. *Journal of Chromatography A*, 1587, 2–13.  
498 <https://doi.org/10.1016/j.chroma.2019.01.012>

499 Gan, Y., Xiao, Y., Wang, S., Guo, H., Liu, M., Wang, Z., & Wang, Y. (2019). Protein-  
500 based fingerprint analysis for the identification of ranae oviductus using RP-HPLC.

501        *Molecules*, 24, 1–11. <https://doi.org/10.3390/molecules24091687>

502        Giménez, M. S., Escudero, N. L., Mucciarelli, S. I., Luco, J. M., & de Arellano, M. L.  
503        (2004). Comparison of the chemical composition and nutritional value of  
504        amaranthus cruentus flour and its protein concentrate. *Plant Foods for Human*  
505        *Nutrition*, 59(1), 15–21. <https://doi.org/10.1007/s11130-004-0033-3>

506        Godoy-Caballero, M. del P., Culzoni, M. J., Galeano-Díaz, T., & Acedo-Valenzuela, M.  
507        I. (2013). Novel combination of non-aqueous capillary electrophoresis and  
508        multivariate curve resolution-alternating least squares to determine phenolic acids in  
509        virgin olive oil. *Analytica Chimica Acta*, 763, 11–19.  
510        <https://doi.org/10.1016/j.aca.2012.12.014>

511        González-Muñoz, A., Montero, B., Enrione, J., & Matiacevich, S. (2016). Rapid  
512        prediction of moisture content of quinoa (*Chenopodium quinoa* Willd.) flour by  
513        Fourier transform infrared (FTIR) spectroscopy. *Journal of Cereal Science*, 71, 246–  
514        249. <https://doi.org/10.1016/j.jcs.2016.09.006>

515        Heiger, D. (2010). *High performance capillary electrophoresis: A primer, Agilent*  
516        *Technologies*.

517        Hong, E., Lee, S. Y., Jeong, J. Y., Park, J. M., Kim, B. H., Kwon, K., & Chun, H. S.  
518        (2017). Modern analytical methods for the detection of food fraud and adulteration  
519        by food category. *Journal of the Science of Food and Agriculture*, 97, 3877–3896.  
520        <https://doi.org/10.1002/jsfa.8364>

521        Jablonski, J. E., Moore, J. C., & Harnly, J. M. (2014). Nontargeted detection of  
522        adulteration of skim milk powder with foreign proteins using UHPLC-UV. *Journal*  
523        *of Agricultural and Food Chemistry*, 62, 5198–5206.  
524        <https://doi.org/10.1021/jf404924x>

525        Jaumot, J., de Juan, A., & Tauler, R. (2015). MCR-ALS GUI 2.0: New features and

526 applications. *Chemometrics and Intelligent Laboratory Systems*, 140, 1–12.  
527 <https://doi.org/10.1016/j.chemolab.2014.10.003>

528 Jaumot, J., Gargallo, R., de Juan, A., & Tauler, R. (2005). A graphical user-friendly  
529 interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB.  
530 *Chemometrics and Intelligent Laboratory Systems*, 76, 101–110.  
531 <https://doi.org/10.1016/j.chemolab.2004.12.007>

532 Joliffe, I. T., & Morgan, B. J. (1992). Principal Component Analysis and exploratory  
533 factor analysis. *Statistical Methods in Medical Research*, 1, 69–95.  
534 <https://doi.org/10.1177/096228029200100105>

535 Laparra, J. M., & Haros, M. (2018). Inclusion of whole flour from latin-american crops  
536 into bread formulations as substitute of wheat delays glucose release and uptake.  
537 *Plant Foods for Human Nutrition*, 73, 13–17. [https://doi.org/10.1007/s11130-018-](https://doi.org/10.1007/s11130-018-0653-6)  
538 [0653-6](https://doi.org/10.1007/s11130-018-0653-6)

539 Li Vigni, M., Baschieri, C., Marchetti, A., & Cocchi, M. (2013). RP-HPLC and  
540 chemometrics for wheat flour protein characterisation in an industrial bread-making  
541 process monitoring context. *Food Chemistry*, 139, 553–562.  
542 <https://doi.org/10.1016/j.foodchem.2013.01.085>

543 Liu, P. F., Avramova, L. V., & Park, C. (2009). Revisiting absorbance at 230 nm as a  
544 protein unfolding probe. *Analytical Biochemistry*, 389, 165–170.  
545 <https://doi.org/10.1016/j.ab.2009.03.028>

546 Mas, S., Fonrodona, G., Tauler, R., & Barbosa, J. (2007). Determination of phenolic acids  
547 in strawberry samples by means of fast liquid chromatography and multivariate  
548 curve resolution methods. *Talanta*, 71, 1455–1463.  
549 <https://doi.org/10.1016/j.talanta.2006.07.030>

550 Montealegre, C., García, M. C., Del Río, C., Marina, M. L., & García-Ruiz, C. (2012).

551 Separation of olive proteins by capillary gel electrophoresis. *Talanta*, 97, 420–424.  
552 <https://doi.org/10.1016/j.talanta.2012.04.055>

553 Montealegre, C., Marina, M. L., & García-Ruiz, C. (2010). Separation of proteins from  
554 olive oil by CE: An approximation to the differentiation of monovarietal olive oils.  
555 *Electrophoresis*, 31, 2218–2225. <https://doi.org/10.1002/elps.200900675>

556 Nancy J Thiex, Harold Manson, Shirley Anderson, J.-Å. P. (2002). Determination of  
557 Crude Protein in Animal Feed, Forage, Grain, and Oilseeds by Using Block  
558 Digestion with a Copper Catalyst and Steam Distillation into Boric Acid:  
559 Collaborative Study. *Journal of AOAC International*, 85, 309–317. Retrieved from  
560 [doi.org/10.1093/jaoac/85.2.309](https://doi.org/10.1093/jaoac/85.2.309)

561 Navarro-Reig, M., Bedia, C., Tauler, R., & Jaumot, J. (2018). Chemometric strategies for  
562 peak detection and profiling from multidimensional chromatography. *Proteomics*,  
563 18, 1–12. <https://doi.org/10.1002/pmic.201700327>

564 Nowak, V., Du, J., & Charrondière, U. R. (2016). Assessment of the nutritional  
565 composition of quinoa (*Chenopodium quinoa* Willd.). *Food Chemistry*, 193, 47–54.  
566 <https://doi.org/10.1016/j.foodchem.2015.02.111>

567 Ojinnaka, D. (2016). Legislative control of quinoa in the United Kingdom and European  
568 Union. *Madridge Journal of Food Technology*, 1, 53–57.  
569 <https://doi.org/10.18689/mjft.2016-108>

570 Pereira, E., Encina-Zelada, C., Barros, L., Gonzales-Barron, U., Cadavez, V., & Ferreira,  
571 I. C. F. R. (2019). Chemical and nutritional characterization of *Chenopodium quinoa*  
572 Willd (quinoa) grains: A good alternative to nutritious food. *Food Chemistry*, 280,  
573 110–114. <https://doi.org/10.1016/j.foodchem.2018.12.068>

574 Pérez-Ràfols, C., & Saurina, J. (2015). Liquid chromatographic fingerprints and profiles  
575 of polyphenolic compounds applied to the chemometric characterization and

576 classification of beers. *Analytical Methods*, 7, 8733–8739.  
577 <https://doi.org/10.1039/c5ay02113h>

578 Rodríguez-Nogales, J. M., Cifuentes, A., García, M. C., & Marina, M. L. (2007).  
579 Characterization of protein fractions from Bt-transgenic and non-transgenic maize  
580 varieties using perfusion and monolithic RP-HPLC. Maize differentiation by  
581 multivariate analysis. *Journal of Agricultural and Food Chemistry*, 55, 3835–3842.  
582 <https://doi.org/10.1021/jf063579e>

583 Rodríguez, S. D., Rolandelli, G., & Buera, M. P. (2019). Detection of quinoa flour  
584 adulteration by means of FT-MIR spectroscopy combined with chemometric  
585 methods. *Food Chemistry*, 274, 392–401.  
586 <https://doi.org/10.1016/j.foodchem.2018.08.140>

587 Rojas, W., Alandia, G., Irigoyen, J., Blajos, J., & Santivañez, T. (2011). *Quinoa: An*  
588 *ancient crop to contribute to world food security*. Retrieved from  
589 [http://www.fao.org/publications/card/en/c/b1511f39-3d33-5b1a-b722-](http://www.fao.org/publications/card/en/c/b1511f39-3d33-5b1a-b722-2208ac2a71b6)  
590 [2208ac2a71b6](http://www.fao.org/publications/card/en/c/b1511f39-3d33-5b1a-b722-2208ac2a71b6)

591 Ropodi, A. I., Panagou, E. Z., & Nychas, G. J. E. (2016). Data mining derived from food  
592 analyses using non-invasive/non-destructive analytical techniques; determination of  
593 food authenticity, quality & safety in tandem with computer science disciplines.  
594 *Trends in Food Science and Technology*, 50, 11–25.  
595 <https://doi.org/10.1016/j.tifs.2016.01.011>

596 Sázelová, P., Kašička, V., Leon, C., Ibáñez, E., & Cifuentes, A. (2012). Capillary  
597 electrophoretic profiling of tryptic digests of water soluble proteins from *Bacillus*  
598 *thuringiensis*-transgenic and non-transgenic maize species. *Food Chemistry*, 134(3),  
599 1607–1615. <https://doi.org/10.1016/j.foodchem.2012.02.220>

600 Štěpánová, S., & Kašička, V. (2019). Recent developments and applications of capillary

601 and microchip electrophoresis in proteomics and peptidomics (2015–mid 2018).  
602 *Journal of Separation Science*, 42, 398–414. <https://doi.org/10.1002/jssc.201801090>

603 Vega-Gálvez, A., Miranda, M., Vergara, J., Uribe, E., Puente, L., & Martínez, E. A.  
604 (2010). Nutrition facts and functional potential of quinoa (*Chenopodium quinoa*  
605 *willd.*), an ancient Andean grain: A review. *Journal of the Science of Food and*  
606 *Agriculture*, 90, 2541–2547. <https://doi.org/10.1002/jsfa.4158>

607 Vergara-Barberán, M., Lerma-García, M. J., Herrero-Martínez, J. M., & Simó-Alfonso,  
608 E. F. (2014a). Classification of olive leaves and pulps according to their cultivar by  
609 using protein profiles established by capillary gel electrophoresis. *Analytical and*  
610 *Bioanalytical Chemistry*, 406(6), 1731–1738. [https://doi.org/10.1007/s00216-013-](https://doi.org/10.1007/s00216-013-7585-7)  
611 [7585-7](https://doi.org/10.1007/s00216-013-7585-7)

612 Vergara-Barberán, M., Lerma-García, M. J., Herrero-Martínez, J. M., & Simó-Alfonso,  
613 E. F. (2014b). Use of protein profiles established by CZE to predict the cultivar of  
614 olive leaves and pulps. *Electrophoresis*, 35, 1652–1659.  
615 <https://doi.org/10.1002/elps.201300530>

616 Vergara-Barberán, M., Mompó-Roselló, O., Lerma-García, M. J., Herrero-Martínez, J.  
617 M., & Simó-Alfonso, E. F. (2017). Enzyme-assisted extraction of proteins from  
618 Citrus fruits and prediction of their cultivar using protein profiles obtained by  
619 capillary gel electrophoresis. *Food Control*, 72, 14–19.  
620 <https://doi.org/10.1016/j.foodcont.2016.07.025>

621 Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of  
622 chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.  
623 [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

624

625

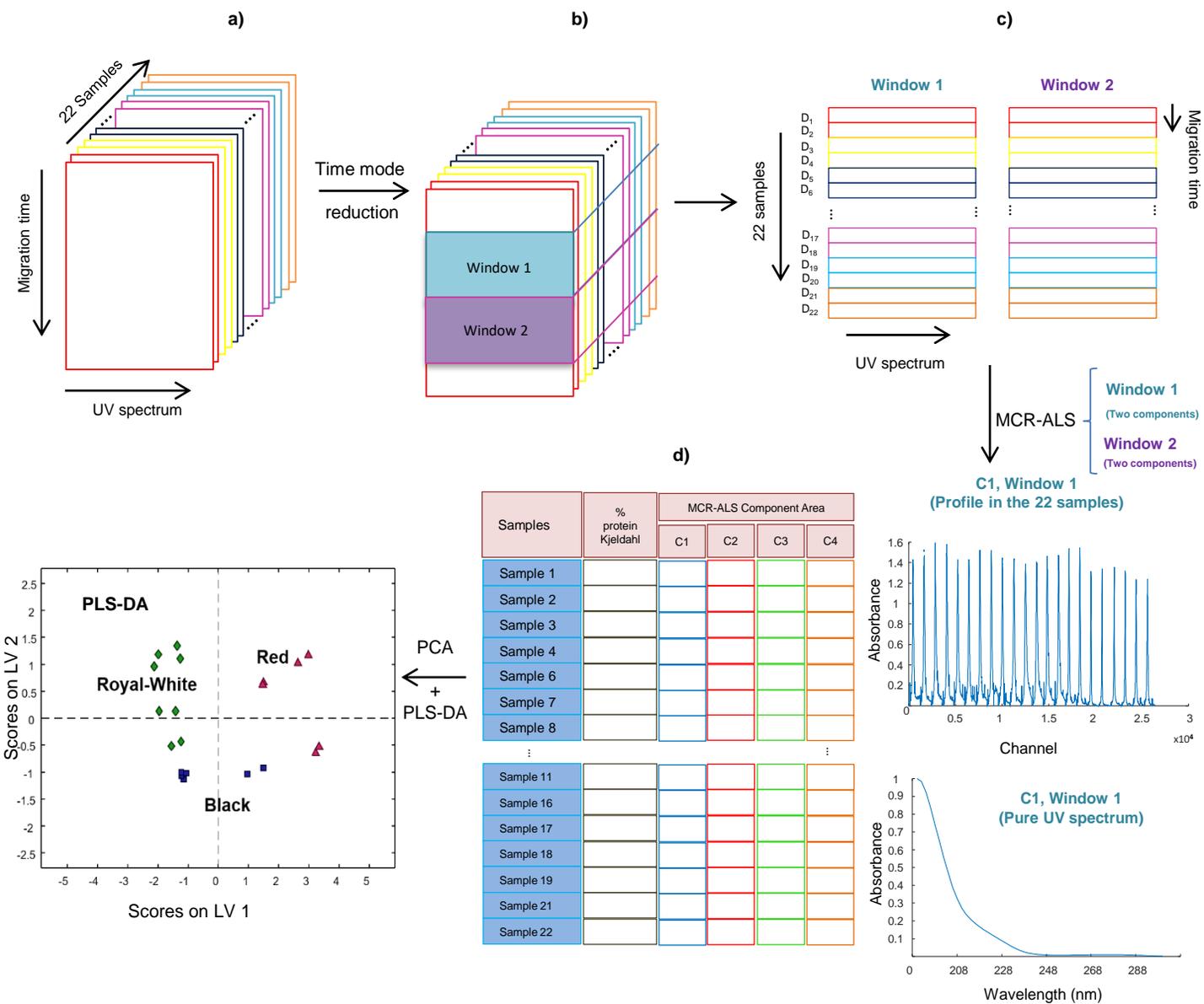
### **CRedit authorship contribution statement**

**Rocío Galindo-Luján:** Methodology, Validation, Investigation, Writing original draft, Writing – review & editing. **Laura Pont:** Investigation, Writing original draft, Writing – review & editing, Supervision. **Victoria Sanz-Nebot:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Fernando Benavente:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

**Table 1.** Total protein content determined by the Kjeldahl method for all the analyzed samples from black (B), red (R), white (W) and royal white (RO) quinoa.

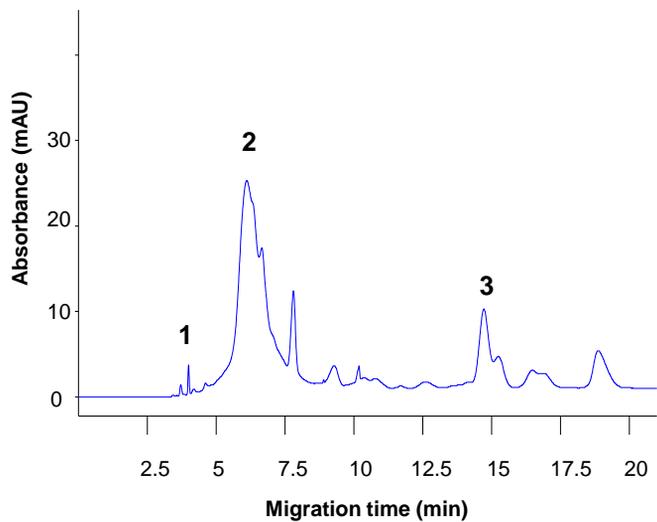
Variety	Code	Total protein content % (m/m)	Average	%RSD
<b>B</b>	B1	15.8	16.0	2
	B2	16.6		
	B3	15.9		
	B4	15.7		
	B5	15.7		
	B6	16.0		
<b>R</b>	R1	15.1	15.6	4
	R2	15.1		
	R3	16.3		
	R4	16.3		
	R5	15.0		
	R6	15.4		
<b>*W</b>	W1	14.4	16.0	11
	W2	14.2		
	W3	15.4		
	W4	15.6		
	W5	18.3		
	W6	18.3		
<b>RO</b>	RO1	15.3	14.8	4
	RO2	14.9		
	RO3	14.9		
	RO4	14.0		

\*Samples W5 and W6, marked in red, were identified as outliers after PCA, see Supplementary Figure 1. Average of total protein content and %RSD values for W quinoa without samples W5 and W6 were 14.9% and 4%, respectively.

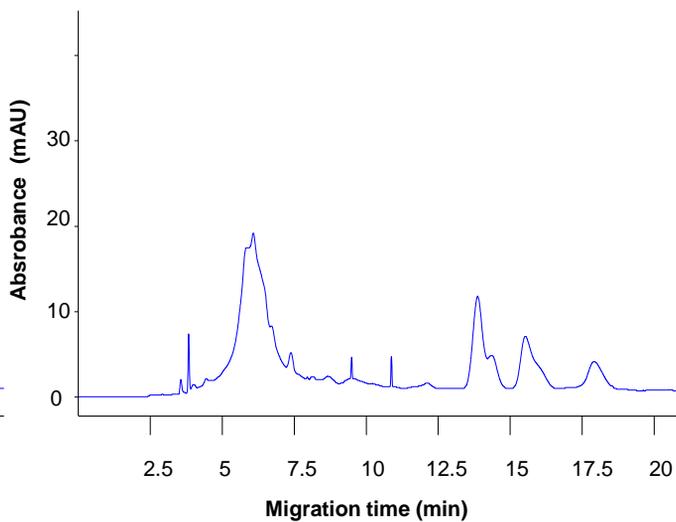


**Figure 1**

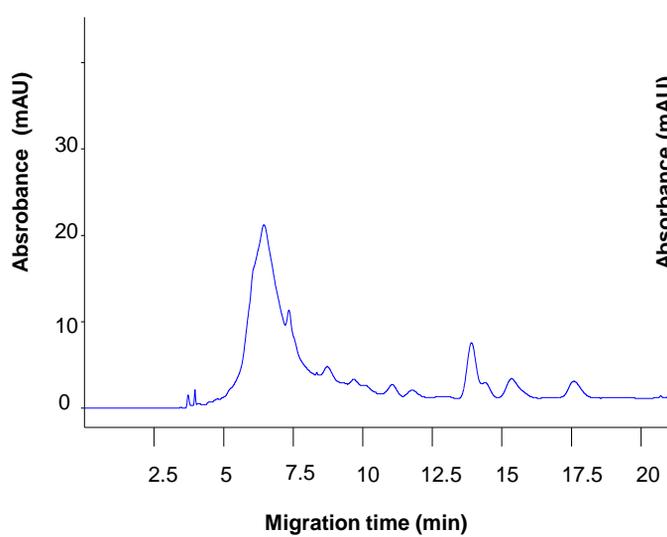
**a) Royal white (RO) Quinoa**



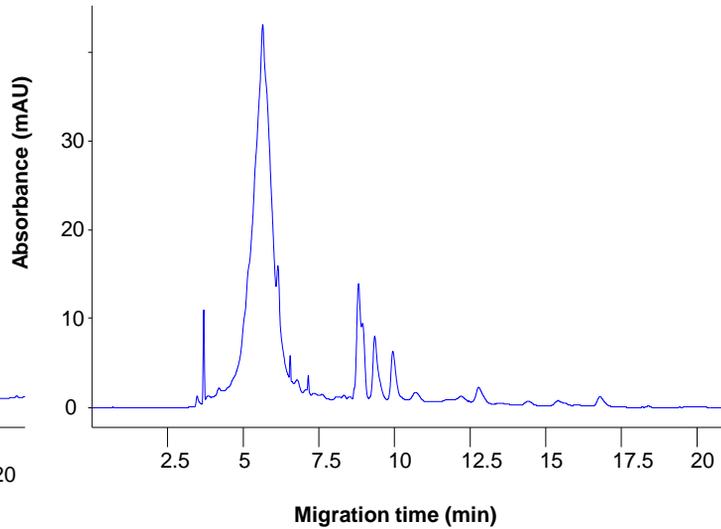
**b) White (W) Quinoa**



**c) Black (B) Quinoa**

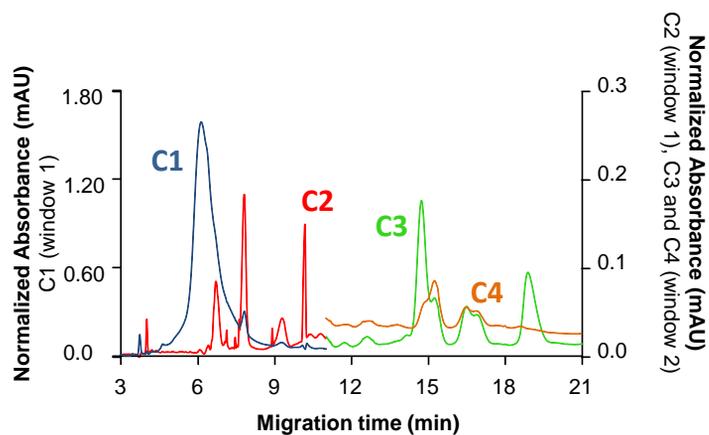


**d) Red (R) Quinoa**



**Figure 2**

### a) MCR-ALS concentration profiles



### b) MCR-ALS UV-spectra

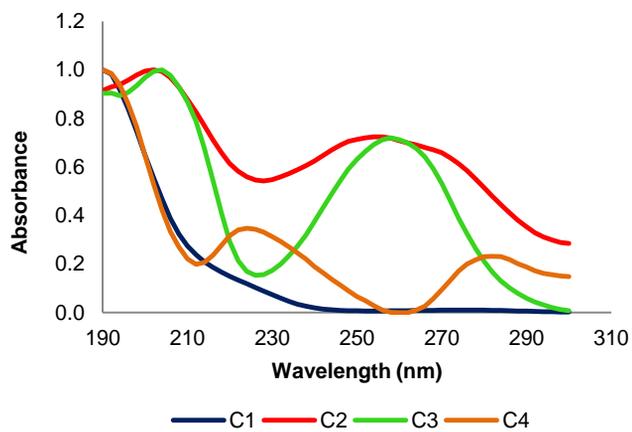
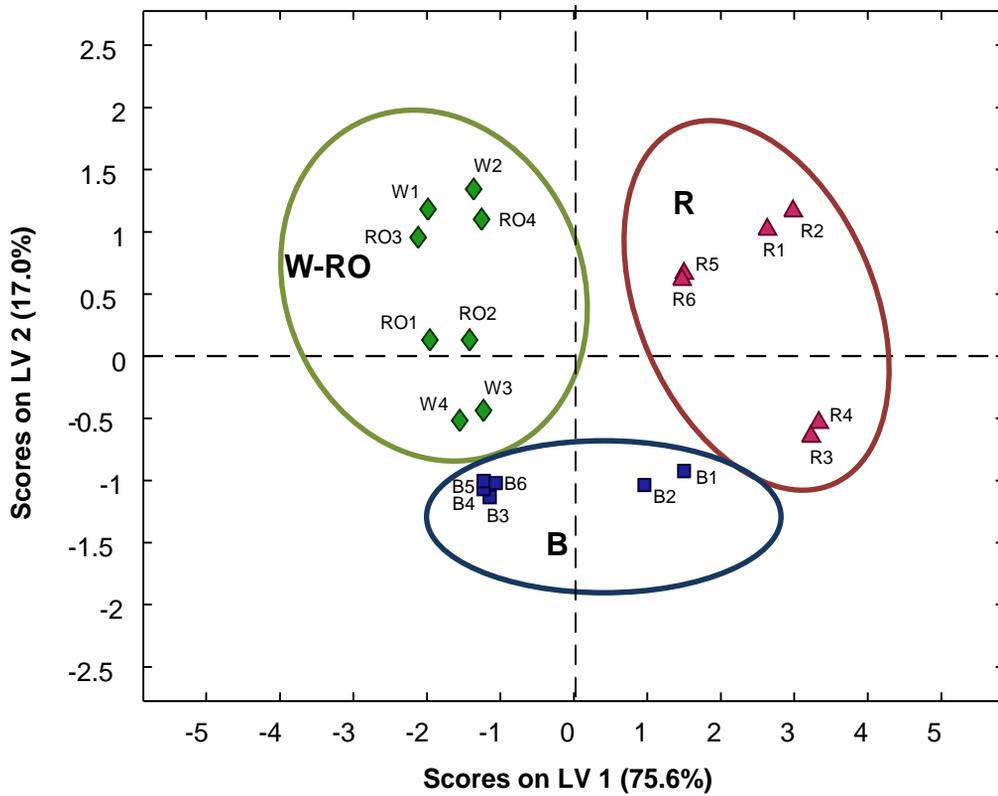
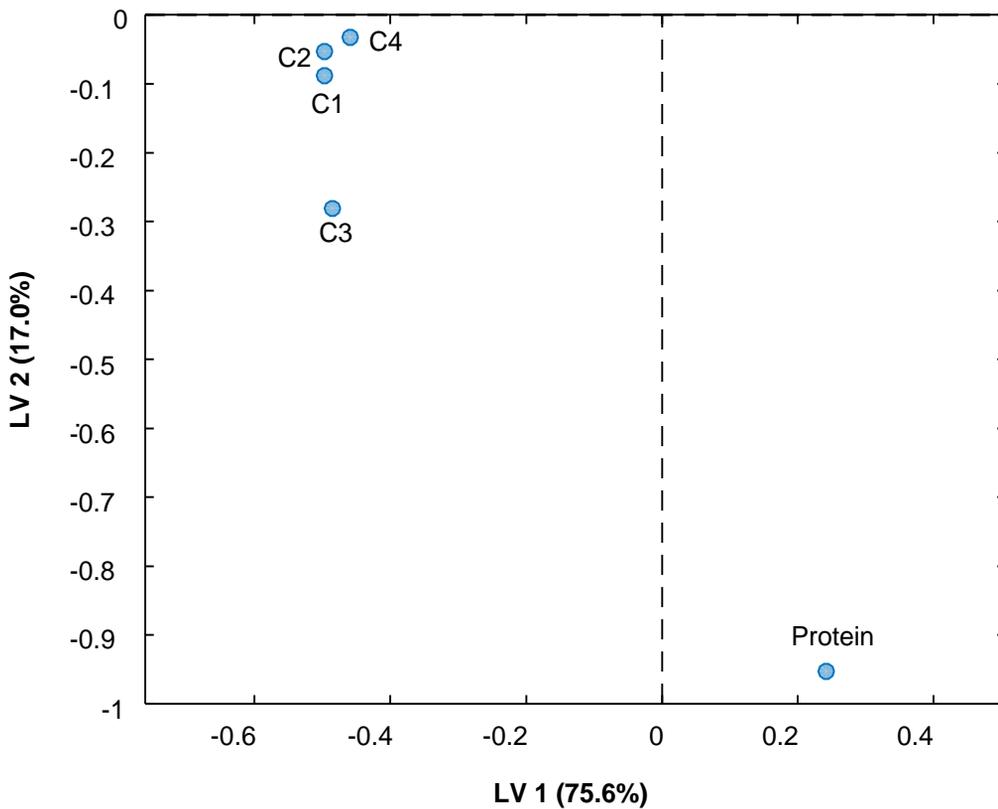


Figure 3

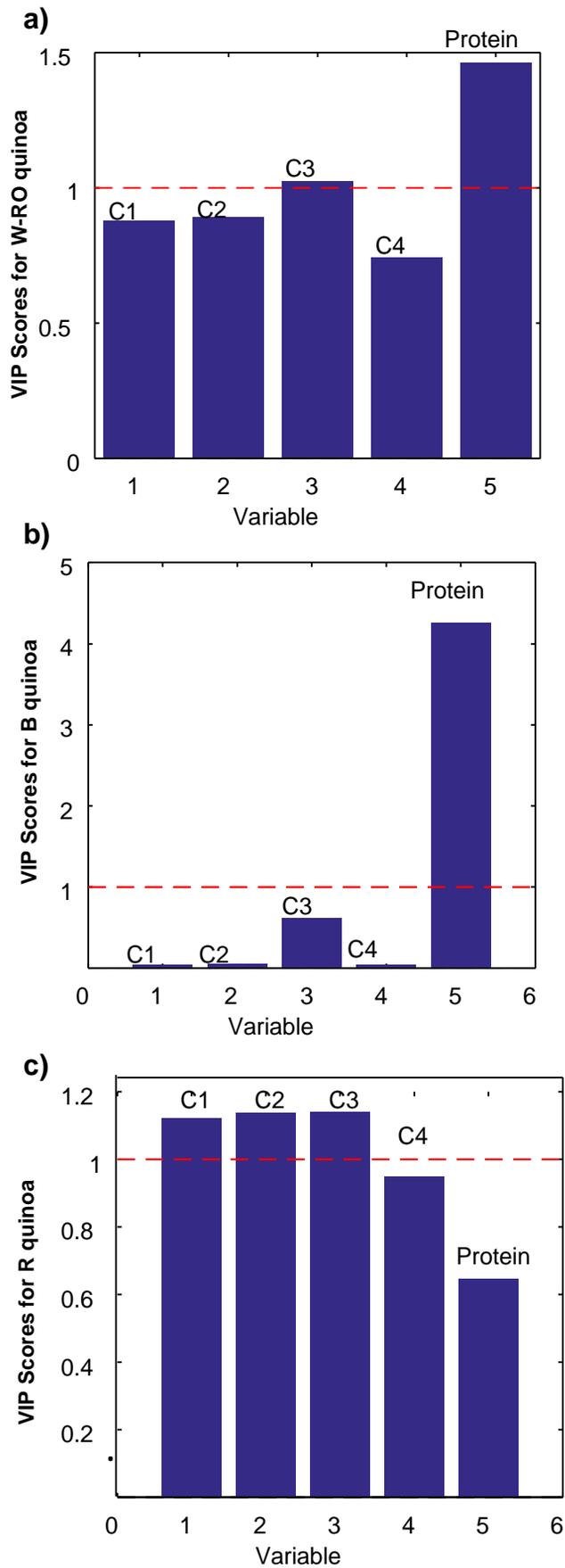
**a) PLS-DA. Scores plot**



**b) PLS-DA. Loadings plot**



**Figure 4**



**Figure 5**

## Supplementary material

### **Classification of quinoa varieties based on protein fingerprinting by capillary electrophoresis with ultraviolet absorption diode array detection and advanced chemometrics**

Rocío Galindo-Luján, Laura Pont, Victoria Sanz-Nebot, Fernando Benavente\*

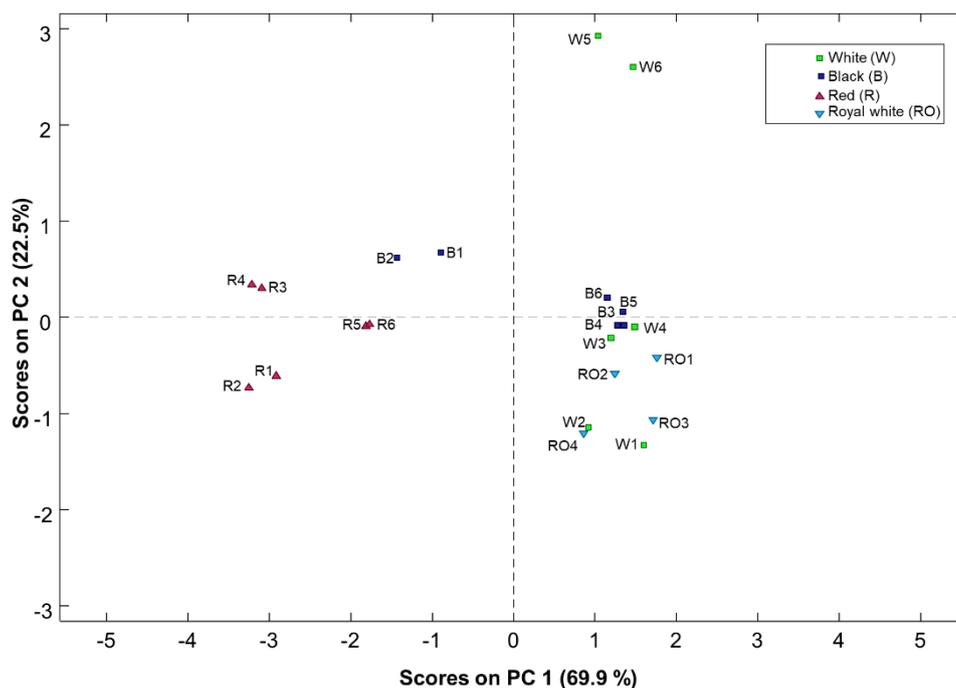
Department of Chemical Engineering and Analytical Chemistry, Institute for Research on Nutrition and Food Safety (INSA·UB), University of Barcelona, 08028, Barcelona, Spain

\*Corresponding author: [fbenavente@ub.edu](mailto:fbenavente@ub.edu) (F. Benavente, PhD)

Tel: (+34) 934039116, Fax: (+34) 934021233

## Table of contents

**Figure S-1.** Scores plot of the PCA model applied to the 22 quinoa S-3 samples using the 4 components resolved by MCR-ALS and the protein content determined by the Kjeldahl method. Two white (W) quinoa samples (W5 and W6) were identified as outliers and were discarded before PLS-DA.



**Figure S-1.** Scores plot of the PCA model applied to the 22 quinoa samples using the 4 components resolved by MCR-ALS and the protein content determined by the Kjeldahl method. Two white (W) quinoa samples (W5 and W6) were identified as outliers and were discarded before PLS-DA.