PARAFAC HPLC-DAD metabolomic fingerprint investigation of reference and crossed coffees

Filipe Corrêa Guizellini[a], Gustavo Galo Marcheafave[a], Miroslava Rakocevic[b,c], Roy Edward Bruns[d], Ieda Spacino Scarminio[a,*], Patricia Kaori Soares[e]

[a] Laboratório de Quimiometria em Ciências Naturais, Departamento de Química, Universidade Estadual de Londrina, CP 10.011, 86057-970 Londrina, PR, Brazil
[b] Embrapa Agricultural Informatics, Av. André Tosello 209, P.O. Box 6041, 13083-886 Campinas, SP, Brazil
[c] Embrapa Environment, Rodovia SP 340, km 127.5, 13820-000 Jaguariúna, SP, Brazil
[d] Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970, Campinas, SP, Brazil
[e] Escola de Ciências e Tecnologia, Universidade Federal Do Rio Grande Do Norte, Campus Lagoa Nova, 59078-970, Natal, RN, Brazil

ARTICLE INFO

ABSTRACT

In this study two cultivars of *Coffea arabica* L., Bourbon (reference) and IPR101 (crossing) were analyzed. The extracts were prepared according to a simplex centroid design with four components, ethanol, ethyl acetate, dichloromethane, and hexane. Multiway data were obtained by HPLC-DAD analysis of the fifteen different mixtures for each cultivar. The PARAFAC methodology was used to investigate the chromatographic fingerprint. For both cultivars, Factor 1 was able to discriminate mixtures containing ethyl acetate as solvent. Factor 2 indicated that mixtures in pure ethanol and binary mixtures containing ethanol were the most efficient in extracting chlorogenic acids and factor 3 identified methylxanthines through spectrophotometric profile in all mixtures. Higher concentrations were obtained by the ethanol, dichloromethane and hexane ternary mixture for the Bourbon cultivar and by the quaternary mixture of these solvents with ethyl acetate for the IPR101 cultivar. Trigonelline and cafestol were extracted in both cultivars. The reference coffee showed higher relative abundances of cafestol ester, chlorogenic acids and trigonelline whereas the crossed coffee showed higher levels of caffeine. To confirm these results, UPLC-MS was used as a complementary method to confirm the presence of the metabolites in these extracts. The three way PARAFAC strategy determines correlations of HPLC-DAD chromatographic and spectral data simultaneously with samples permitting a more unambiguous assignment of metabolic groups than can be obtained treating chromatographic and spectral data separately by two way methods. This can provide higher quality chromatographic fingerprints for food chemistry.

1. Introduction

Coffea arabica is a species of greatest economic importance, characterized by difficult gene manipulation with flowering influenced by various environmental factors such as water stress, solar radiation, temperature, wind and relative humidity, as well as other factors such as the interaction between cultures and pruning time (Nascimento, Spehar, & Sandri, 2014). These factors directly affect the concentration of principal metabolites and consequently the production of fruits and exert great influence on the incidence of pests and diseases, like coffee leaf rust, in all phenological phases resulting in many plant breeding studies aimed at solving these issues.

C. arabica green beans contain a wide variety of chemical compounds, primary and secondary metabolites besides intermediate products (Hatumura et al., 2017; Worku, de Meulenaer, Duchateau, & Boeckx, 2018). The latter are used as taxonomic markers, and contribute to the specific smells, tastes, and colors of coffees (Bennett & Wallsgrove, 1994; Crozier, Clifford, & Ashihara, 2006; Martín, Pablos, & González, 1998; Terrile et al., 2016). Many of these compounds, such as alkaloids (Delaroza, Rakocevic, Malta, Bruns, & Scarminio, 2014; Pompelli, Pompelli, de Oliveira, & Antunes, 2013) and chlorogenic acids, interfere with the flavor of the drink (Scholz, da Silva, de Figueiredo, & Kitzberger, 2013) even after the roasting process (Belay & Gholap, 2009). Among these alkaloids are the purine bases, including methylxanthines that primarily act on the central nervous system (Crozier et al., 2006).

Metabolomics, also known as metabolic phenotyping, focuses on the complex interactions of the components of a biological system and

highlights the whole system rather than just partitioned elements, resulting in novel information about cellular equilibrium (Basanta et al., 2010; Liu et al., 2010; Zhang, Sun, Wang, Han, & Wang, 2012). Furthermore, modern techniques are used to analyze samples in order to search for an analytical description, quantifying, and characterizing all low molecular weight components of a biological sample (Kosmides, Kamisoglu, Calvano, Corbett, & Androulakis, 2013; Nicholson & Lindon, 2008; Tang et al., 2016).

To examine a biological system, one must know how cellular regulation processes work, so metabolomics aims at determining the so-called final products of these processes, the metabolites, which are considered the ultimate response of a system to some stimuli, or to genetic and environmental changes (Fiehn et al., 2000; Gomez-Casati, Zanor, & Busi, 2013). The great difficulty in working with metabolites stems from their high dynamisms in time and space, making possible the existence of a very wide number of shapes and structures (Gomez-Casati et al., 2013; Stitt & Fernie, 2003). Considering that each plant material is composed of different metabolite groups, statistical mixture designs provide an excellent strategy for determining the best extraction solvent for obtaining the most relevant chemical information. These designs permit optimization of mixture proportions of solvents with different polarities and strengths. (Bruns, Scarminio, & Neto, 2006; Cornell, 2002; Hatumura et al., 2017).

Metabolomic fingerprinting is a powerful tool for quality control of herbal medicines (Custers, Van Praag, Courselle, Apers, & Deconinck, 2017; Jin-lan Zhang, Cui, He, Yu, & Guo, 2005), and often involves chromatographic separation techniques. Currently, the spectroscopic techniques coupled to chromatographic ones and combined with statistical tools to treat the large amounts of chemical information generated are leading to increased knowledge about complex systems. HPLC or UPLC are the frequently used chromatographic techniques, because of their high adaptive capacity providing information on almost all metabolites (Harris et al., 2007; Oms-Oliu, Odriozola-Serrano, & Martín-Belloso, 2013). HPLC is considered the separation technique most used for metabolomics. A large range of detectors can be coupled to HPLC for this purpose, the most common being the mass spectrometer (HPLC-MS) (Passari, Scarminio, & Bruns, 2014) and the diode array (HPLC-DAD) (Soares, Bruns, & Scarminio, 2012; Soares & Scarminio, 2008).

Modern food research needs good data analysis due to the complex problems to be solved in food chemistry. The chemical composition depends on the raw materials, preparation procedures, storage conditions, and time, among others. In all these cases, chemometric methods have become a helpful tool to explore the obtained data. The progress of chemometrics allowed that the metabolomic analysis from spectroscopic and chromatographic data could be considered, because these tools made possible to extract in an efficient and precise way all the complex information hidden in these spectrums and chromatograms, revealing their true meaning by associating these parameters to properties of interest or chemical substances (Cubero-Leon, Peñalver, & Maquet, 2014). It has been successfully applied in the analysis of yogurts (Cruz et al., 2013), beers (Granato, Branco, Faria, & Cruz, 2011), meat products (Matera et al., 2014; Zhao et al., 2018), milk (Souza et al., 2011), fruits (Filho et al., 2017), tea (Chengying et al., 2018), oils (Bajoub et al., 2018), coffee (Terrile et al., 2016), functional foods (Passari, Scarminio, Marcheafave, & Bruns, 2017) and others.

Some chemometric tools such as principal component analysis and hierarchical analysis for multi-mode data have helped hyphenated techniques describing the best model for the data provided. However, these tools are commonly used for data in two dimensions, samples and analytical signals from one source, allowing individual analysis of each sample. Therefore, in order to simultaneously analyze three dimensional data the higher order chemometric methods such as Parallel Factor Analysis (PARAFAC) (Mortera, Zuljan, Magni, Bortolato, & Alarcón, 2018) and Tucker methods (Azcarate et al., 2016; Soares et al., 2012), are very attractive for complex system studies.

The main objective here is to obtain and compare the modern fingerprints of two cultivars, *C. arabica* Bourbon and IPR101 by means of a statistical mixture design, using the PARAFAC method for exploiting three-way HPLC-DAD data as an investigation strategy to identify major differences in their metabolomes. This could help lead to direct crosses between new species or improved species against harmful factors thus improving the resistance of the coffee tree increasing not only production but also the quality of the coffee.

2. Experimental

2.1. Coffee samples

For this study the green beans of two different cultivars of *C. Arabica* were kindly provided by the Agronomic Institute of Parana (IAPAR), the red Bourbon and the cultivar named as IPR101 which is developed by IAPAR. The Bourbon cultivar is stated as pure Arabica coffee, with no improvement or crossing with other cultivars and was used as reference, while IPR101 was originated by the crossing of Catuaí and Sarchimor cultivars with the SH2 and SH3 rust resistant genes. These samples were prepared and analyzed as described in Moreira and Scarminio (2013).

These arabica coffee cultivars were grown under the edaphoclimatic conditions on the Experimental Field of IAPAR, in Londrina, Paraná state, Brazil. Cherry fruits were selected from the harvest of May to July 2010, washed and sun-dried on a patio. Samples were standardized to a sieve size of 6.5 mm, had their defective beans removed, and were frozen at −18 °C. The coffee beans were previously immersed in liquid nitrogen to make them friable, facilitating their grinding and avoiding the possible oxidation of the chemical compounds, and then sieved.

2.2. Extraction

Based on a statistical mixture design, the extraction procedure followed a simplex centroid design (Bruns et al., 2006; Cornell, 2002) for four solvents, ethanol, ethyl acetate, dichloromethane and hexane. According to Table 1, fifteen different extracts were prepared for each coffee sample, four with pure solvents, six with binary mixtures of equal proportions, three equal proportioned ternary mixtures and one quaternary mixture.

Each coffee extract was prepared by weighing 10 g coffee samples and adding 150 mL of solvent mixture (Table 1). With the aid of ultrasound, extractions were carried out during 30 min, changing the ultrasonic bath water every 15 min or less to prevent heating. The extract was then filtered and stored. This process was repeated twice using the same coffee material and adding a new 150 mL portion of the

Table 1
Extractor mixture composition in terms of proportion.

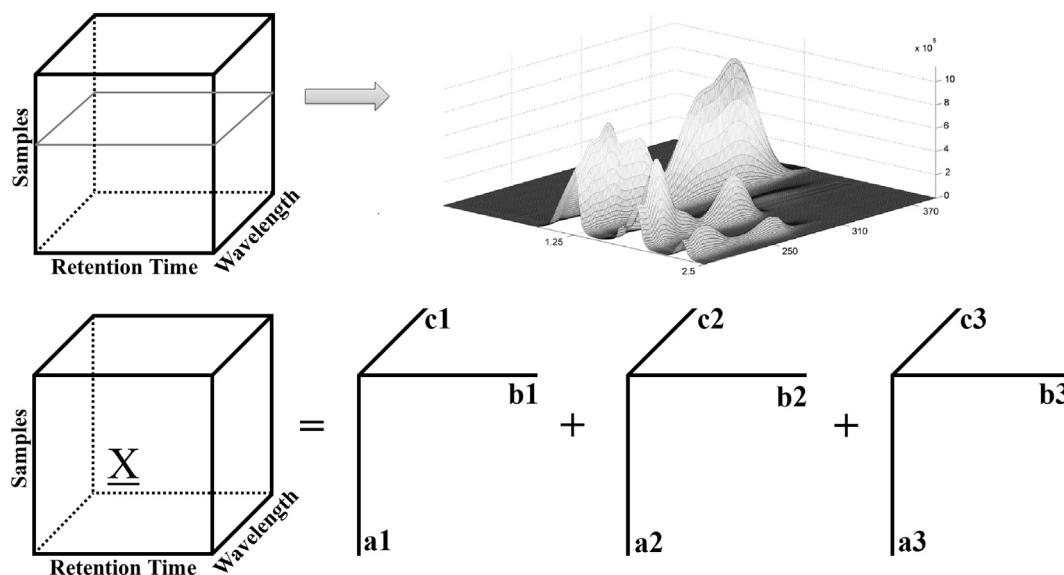| Design Point	Extractor	Solvents			
		Ethanol	Ethyl Acetate	Dichloromethane	Hexane
1	e	1	0	0	0
2	a	0	1	0	0
3	d	0	0	1	0
4	h	0	0	0	1
5	ea	1/2	1/2	0	0
6	ed	1/2	0	1/2	0
7	eh	1/2	0	0	1/2
8	ad	0	1/2	1/2	0
9	ah	0	1/2	0	1/2
10	dh	0	0	1/2	1/2
11	ead	1/3	1/3	1/3	0
12	eah	1/3	1/3	0	1/3
13	edh	1/3	0	1/3	1/3
14	adh	0	1/3	1/3	1/3
15	eadh	1/4	1/4	1/4	1/4

Fig. 1. Data arrangement for PARAFAC analysis.

extractive mixture. At the end of the extraction process, thirty crude extracts were obtained, fifteen of each cultivar.

2.3. HPLC-DAD analysis

After extraction, a 20 μL aliquot of each crude extract was diluted in 800 μL of mobile phase. Diluted samples were filtered through a Millipore Millex 0.22 μm filter and the chromatographic analysis was immediately performed. HPLC analysis was conducted on a Finnigan Surveyor 61,607 liquid chromatograph equipped with a Finnigan Surveyor PDA Plus diode array detector. Elution was performed isocratically using a Phenomenex C18 column, Kinetex 2.6 μm Hilic 100 Å, with dimensions of 150 mm × 4.6 mm, injection volume of 20 μL, mobile phase flow at 1.0 mL min^{-1}, elution monitored at wavelengths at 210, 240 and 254 nm. Satisfactory separation was achieved at 210 nm. ChromQuest Software 4.2 was used for data processing. Nine different mobile phases were tested and a ternary mixture composed of 51% of water, 26% of acetonitrile and 23% of methanol was selected as it presented the highest number of peaks and best resolution. The mobile phase was sonicated for 30 min prior to analysis in order to eliminate dissolved soluble gases that could cause detector instability.

2.4. UPLC-MS analysis

The separated substances were confirmed using an Acquity UPLC chromatographic system (Waters, Millford, MA) equipped with a binary pump system. The UPLC system was coupled to a Quattro Micro API triple quadrupole (Waters, Manchester, UK) mass spectrometer (MS) using a Z-spray electrospray ionization (ESI) source, with sample conditions similar to those used for HPLC-DAD analyses except that the mobile phase was increased by 0.1% (v/v) formic acid (85% Vetec). The analysis was carried out in both the positive and negative modes to determine the different fragments of compounds in the extracts. Data were acquired in the scan mode covering the 80 to 1000 Da m/z range. The ionization source conditions were: 3 kV capillary voltage, 150 °C source temperature, 80 L h^{-1} conical gas flow, 800 L h^{-1} desolvation gas flow and 350 °C desolvation temperature. Nitrogen nebulizer gas was 99% pure. The data were processed using the v4.1 MassLynx software.

3. Data analysis

3.1. PARAFAC

The multiway chromatographic fingerprint provided by the HPLC-DAD was investigated by the PARAFAC method as it is one of the main decomposition methods for multi-way data. This method basically decomposes the array into sets of scores and loadings that hopefully describe the data in a more condensed form than the original data array (Bro, 1997).

In the PARAFAC method, a three way data array $\underline{\mathbf{X}}$ ($I \times J \times K$) with elements x_{ijk} is decomposed into three loadings matrices, $\mathbf{A}$ ($I \times F$), $\mathbf{B}$ ($J \times F$) and $\mathbf{C}$ ($K \times F$) the elements of which are a_{if}, b_{jf} and c_{kf}, respectively and F is the number of chosen factors (Smilde, Bro, & Geladi, 2004). A trilinear model is found to minimize the sum of squares of the residues, e_{ijk}. The model can be written as follows (Olivieri & Escandar, 2014; Smilde et al., 2004):

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$

The determination of model complexity, i.e. the number of factors, is one of the most important steps in data analysis, and there is no absolute criterion for this purpose. Several tools can be used, such as the explained variance of the model, the chemical knowledge of the system, cross validation (Cruz et al., 2012; Kroonenberg, 2008; Stanimirova, Kita, Malkowski, John, & Walczak, 2009) and the core consistency diagnostic (CORCONDIA) (Bro & Kiers, 2003).

Here, PARAFAC models with different model complexities were calculated and the best one was chosen by means of the CORCONDIA criteria. PARAFAC models were calculated using Matlab version 6.5 equipped with the N-way toolbox for Matlab version 3.00 using the original data cube for the fingerprints (Andersson & Bro, 2000; "Matlab 6.5," 2000). The N-way toolbox can be downloaded at http://www.models.kvl.dk/nwaytoolbox.

3.2. Data arrays and analysis

The HPLC-DAD three-way data arrays (Fig. 1), analyzed for both cultivars have absorbance values as a function of 600 retention times (chromatographic - mode A), corresponding to a 5 min retention time run for the 15 samples (mode C) and absorbance values as a function of 210 wavelengths (spectral-mode B) corresponding to the spectral

Table 2
CORCONDIA value, explained variance and number of iterations for PARAFAC models containing from 1 to 5 components for Bourbon and IPR101 original data cubes.

Sample	Number of components	1	2	3	4	5
Bourbon	CORCONDIA (%)	100	100	98.68	0.08	0.04
	Explained variance (%)	65.11	91.34	96.13	97.70	98.51
	N° of iteration	6	2	10	20	363
IPR101	CORCONDIA (%)	100	100	99.20	17.06	19.68
	Explained variance (%)	68.07	89.17	94.95	97.32	98.17
	N° of iteration	5	2	7	252	122

interval of 190 to 410 nm. As such the data array has dimensions of $15 \times 600 \times 210$.

The three-way data arrays for Bourbon and IPR101 cultivars were treated with the CORCONDIA diagnostic. Each data cube was mean centered in each of the modes, simultaneously in two modes and in all the three modes. Including the original data cube, a total of eight different data arrays were tested with the CORCONDIA algorithm for a model containing from 1 to 5 components. The CORCONDIA values and the explained variances were evaluated. As no differences were noted between the preprocessed and the original data cube for both cultivars, the original data were chosen for analysis. Table 2 presents the COR-CONDIA values, the explained variance and the number of iteration for models containing from 1 to 5 components for Bourbon and IPR101.

Models with one component presented a CORCONDIA of 100%,

although the explained variance does not reach 70%. Models with two components presented high CORCONDIA values and an acceptable explained variance although models with three factors presented CORCONDIA values very close to 100% and explained variances that are 5% higher than models with two components. Models with four and five components presented a very low CORCONDIA value, indicating invalid models. For both cultivars the most adequate models, containing two and three components were evaluated and the best results for both cultivars were achieved with a PARAFAC model with three components applied to the original data cube.

4. Results and discussion

For the Bourbon cultivar, the loadings for the PARAFAC model with three components are shown in Fig. 2. The first column of Fig. 2 represents the loadings of modes A, B and C of component one, the second column the loadings for component two and so on.

Mode A of the first component (column 1 of Fig. 2) shows that the chromatographic peak at the 3.41 min retention time presents the highest positive loading, followed by two retention times, 2.56 and 4.56 min., with small negative loadings. For mode B (line 2, column 1 of Fig. 2) the spectral profile shows maxima with positive weights at 208 and 228 nm wavelengths and a small negative value at 272 nm. Mode C (line 3, column 1 of Fig. 2) indicates that mixtures with higher loading values are pure ethyl acetate followed by mixtures containing di-chloromethane whereas pure ethanol and hexane and their 1:1 mixture
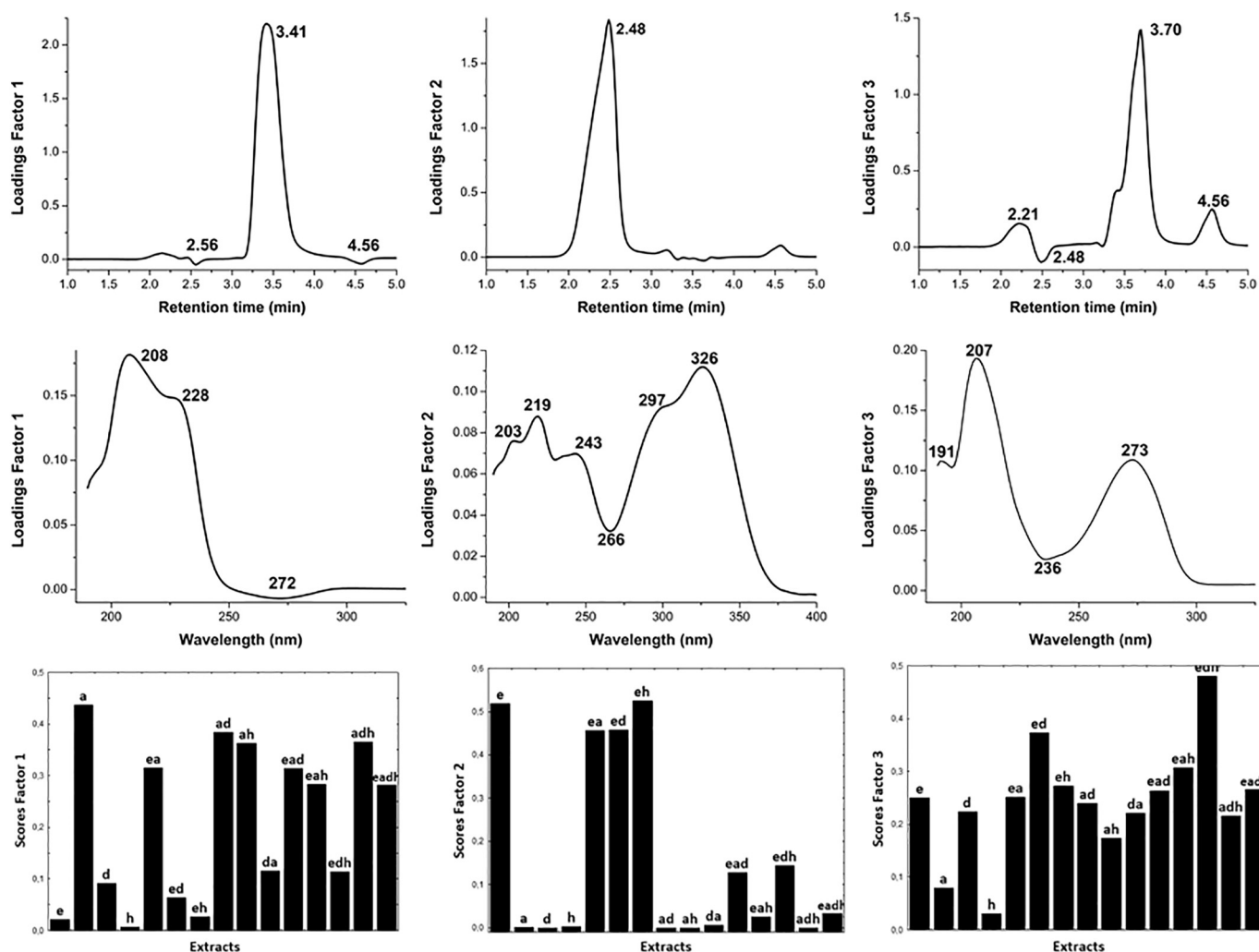

Fig. 2. Loadings and scores of factors 1, 2 and 3 (columns) for modes A, B and C (rows) for the second order chromatographic data of the Bourbon cultivar.
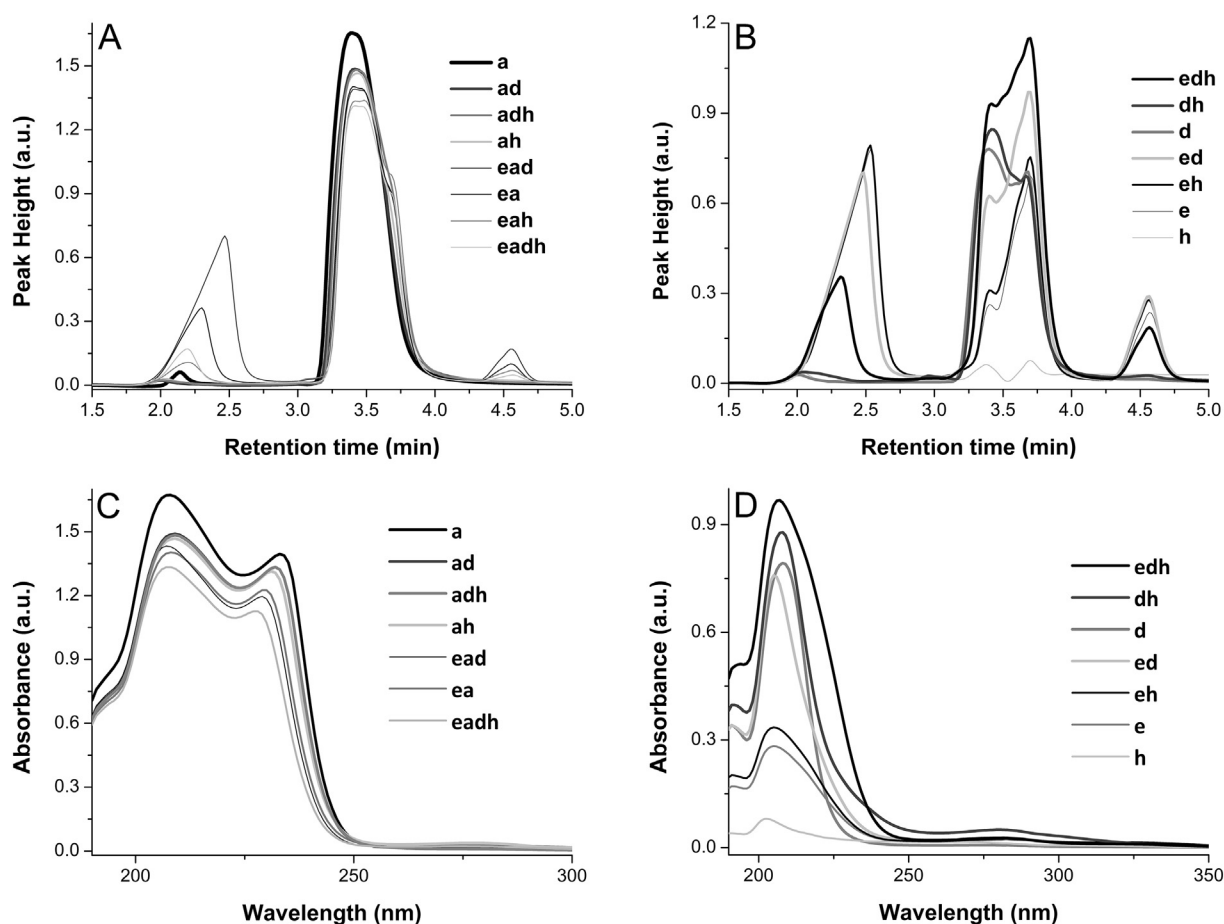
Fig. 3. Chromatograms of the fifteen mixtures of Bourbon cultivar at the 210 nm wavelength separated into two groups (A–B) and spectra of the 3.41 min retention time peak of the chromatographic fingerprint of the Bourbon cultivar (C–D). Chromatographic conditions: mobile phase composed of 51% of water, 26% of acetonitrile and 23% of methanol, isocratic elution at 1.0 mL min^{-1}, and wavelength of 210 nm.
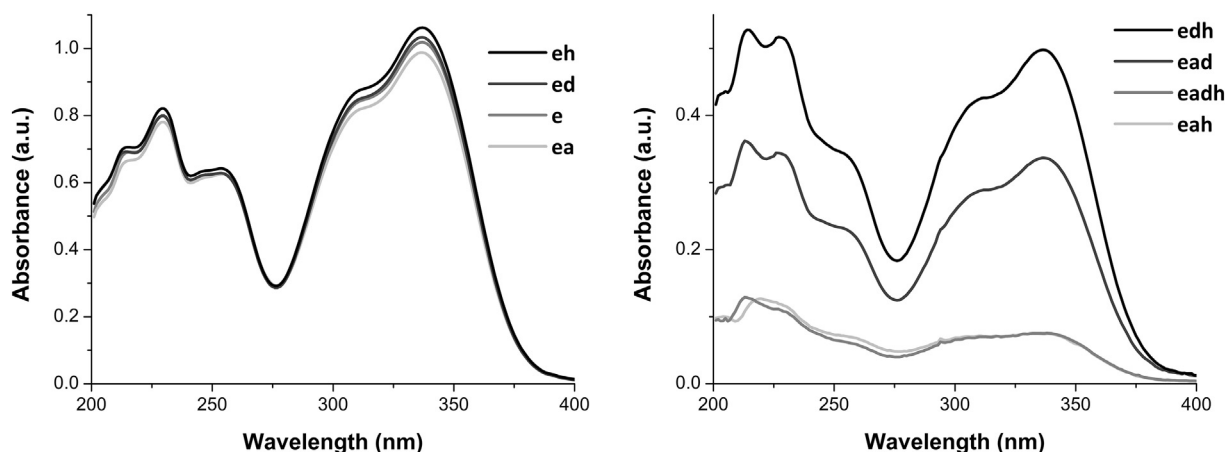

Fig. 4. Spectra of the 2.48 min retention time substance from mixtures containing ethanol in the composition of extractor solvent for the Bourbon cultivar.

have the lowest weights. As such component one indicates that samples containing ethyl acetate in the extractor solvent present higher quantities of substances represented by the chromatographic peak at the 3.41 min retention time and with a spectral profile maxima at approximately 208 and 228 nm. Samples with smaller weights are associated with substances with lower weights on factor 1 corresponding to retention times of 2.56 and 4.56 min.

Chromatograms of the fifteen samples of Bourbon cultivar obtained at the 210 nm wavelength are presented in Fig. 3A–B and shows that

mixtures containing ethyl acetate in the composition of the extractor solvent presents a peak at 3.41 min, whereas the other mixtures present overlapping peaks at slightly higher retention times. The spectra for the chromatographic peak at 3.41 min retention time are presented in Fig. 3C and shows a very similar profile to mode B of component 1 for mixtures containing ethyl acetate extractor solvent, while the other mixtures, Fig. 3D have spectra profiles with a maximum around 208 nm. These spectral profiles are characteristic of cafestol esters, metabolites containing pentacyclic furan terpenes exclusively found in
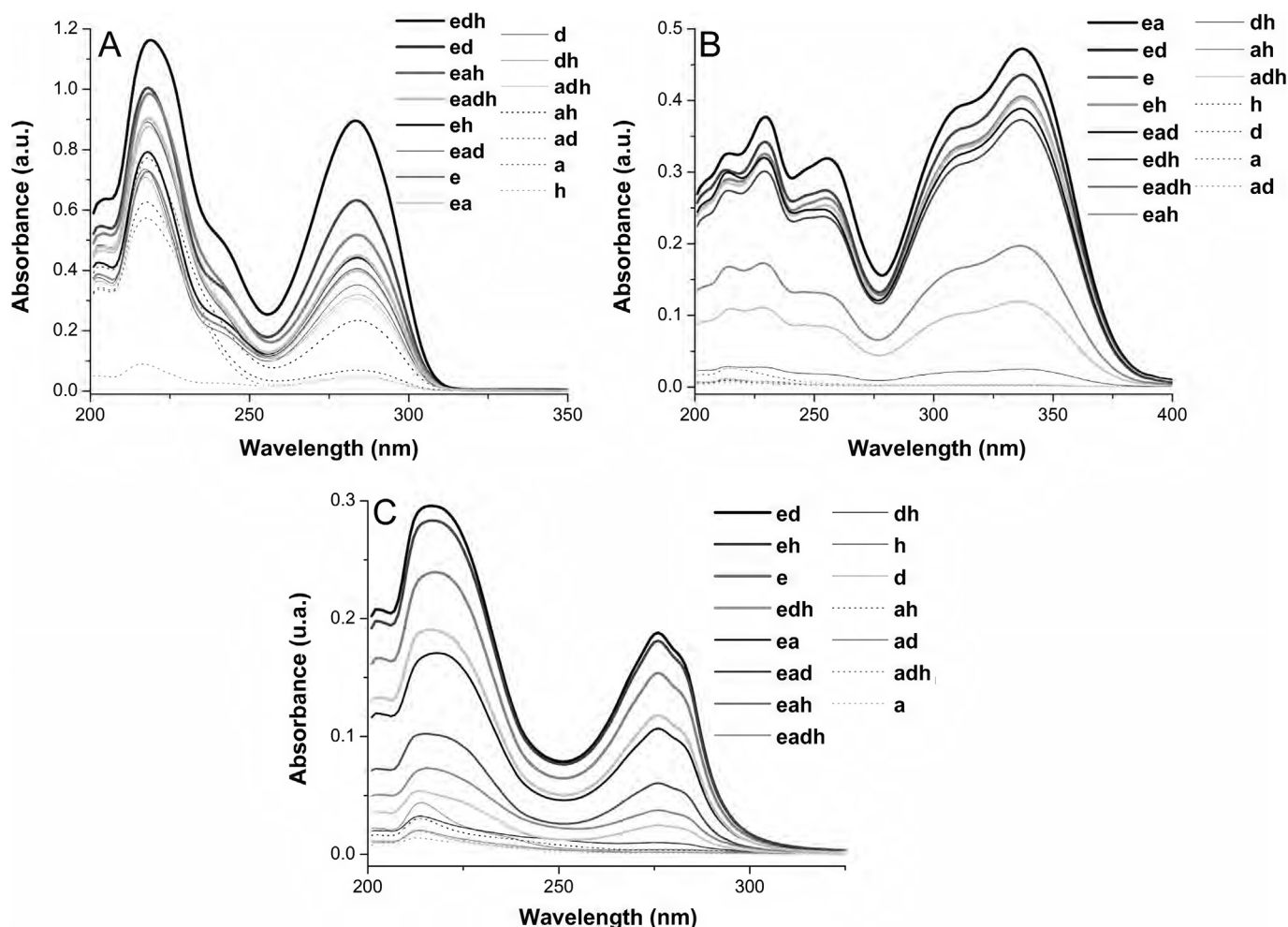
Fig. 5. Spectra of the 3.70 min retention time peak of the chromatographic fingerprint of the Bourbon cultivar (Fig. 5A) and spectra of the 2.21 min (Fig. 5B) and 4.56 min (Fig. 5C) retention time substances from the chromatograms of the Bourbon cultivar. Chromatographic conditions: mobile phase composed of 51% of water, 26% of acetonitrile and 23% of methanol, isocratic elution at $1.0 \, mL \, min^{-1}$, and wavelength of 210 nm.

coffee (Moeenfard, Erny, & Alves, 2016). The spectral profile with the overlapped pair is similar to the one for cafestol linolate while cafestol oleate has one maximum at 208 nm (Ern, Moeenfard, & Alves, 2014; Erny, Moeenfard, & Alves, 2015).

Modes A, B and C of the second component are presented in the second column of Fig. 2. Loadings of mode A present one retention time with large weight and positive value at 2.48 min retention time. Loadings of mode B present five relative maxima at 203, 219, 243, 297 and 326 nm. Mode C separates the mixtures by the presence of ethanol in the composition of the extractor solvent, the weight decreasing as the ethanol proportion in the mixture decreases. Mixtures without ethanol present weights near zero. So this component indicates that mixture solvents containing ethanol in any proportion are capable of extracting the substance with retention time around 2.48 min which is confirmed by the chromatograms presented in Fig. 3A-B.

Spectra for chromatographic peak at 2.48 min are plotted in Fig. 4 to confirm this behavior. Spectra of the pure ethanol and binary mixtures containing ethanol present the same spectral profile as the mode B loadings of the second component, ternary and quaternary mixtures presenting a similar profile, although with lower absorbance, possibly indicating an antagonistic effect among these solvents for extracting this substance. This spectral profile matches the one for chlorogenic acids, with maxima around of 326 (max) and 300 nm wavelengths (Belay & Gholap, 2009).

The third column of Fig. 2 presents modes A, B and C for the third PARAFAC component. Loadings of mode A indicates three retention

times with positive values, one at 3.70 min retention time with the highest weight and two others at 2.21 and 4.56 min. One variable with low negative weight occurs at 2.48 min retention time. Loadings of mode B present wavelengths with only with positive values and maxima at 207 and 273 nm. From loadings of mode C it can be seen that mixtures *edh* and *ed* present the highest score values and with intermediate values, between 0.25 and 0.35, for mixtures containing ethanol in the extractor solvent composition, whereas the other mixtures have values under 0.25. The combination of the modes A, B and C for this factor indicates that mixtures *edh* and *ed* present the largest intensities for the peak at 3.70 min. Fig. 3A–B present the chromatograms for the Bourbon cultivar and shows that mixtures containing acetone presented a different peak from the other mixtures at the same retention time.

To evaluate this difference, the original spectra for the chromatographic peak at the 3.70 min retention time were plotted in Fig. 5A for all mixtures. It confirms that mixtures *ed* and *edh* presented a higher absorbance intensity, followed by mixtures containing ethanol and with pure hexane having the lowest absorbance. This profile is very similar to the spectral profile of methylxanthines (caffeine) which presents maxima at 272 nm (López-Martínez, López-De-Alba, García-Campos, & De León-Rodríguez, 2003). To evaluate if the chromatographic peaks indicated by mode A of the third component with retention times of 2.21 and 4.56 min presented the same spectral profile as the mode B loadings of the third component, their spectra were plotted in Fig. 5B–C.

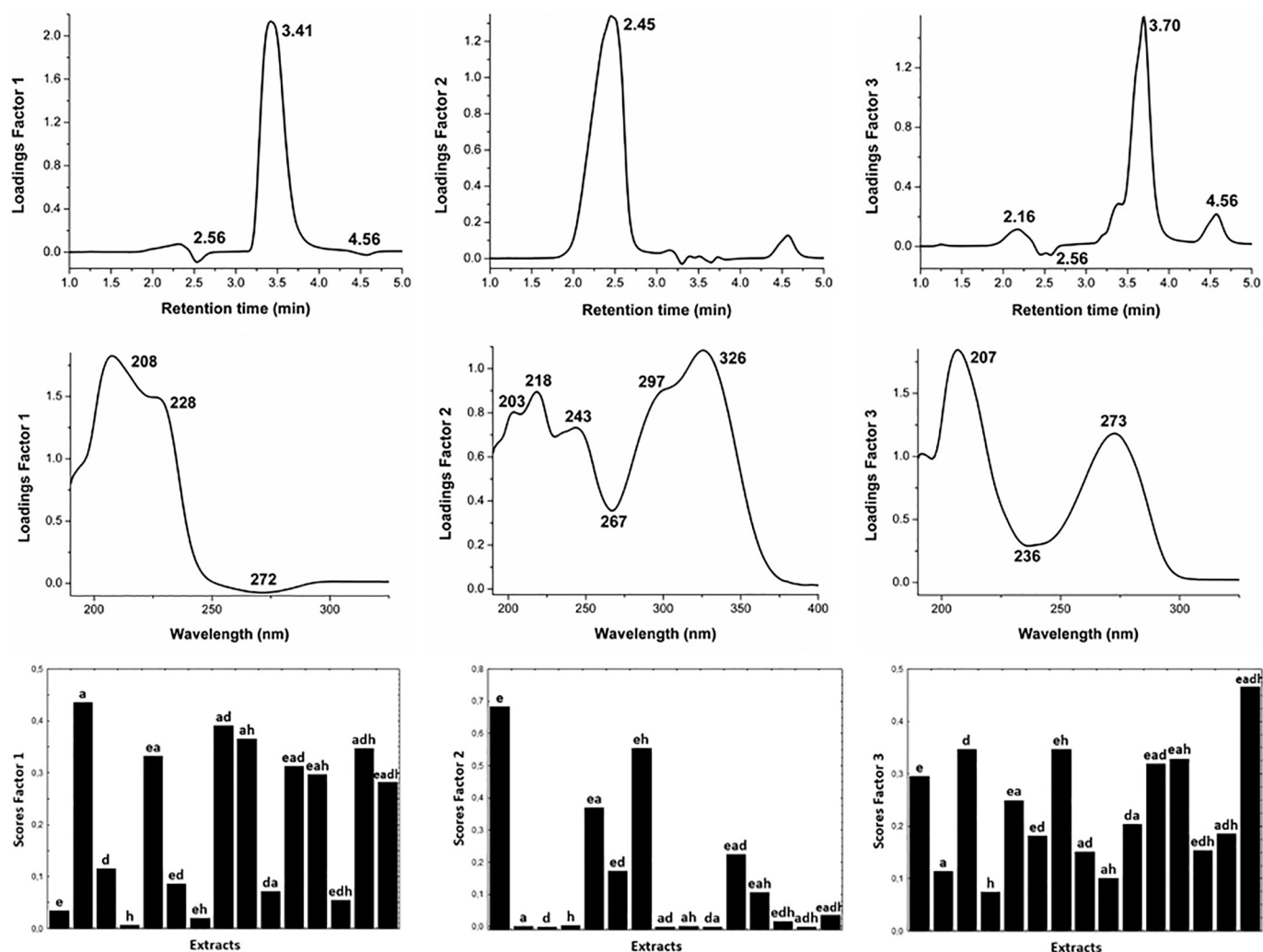Fig. 5B–C indicates that mixtures containing ethanol in the extractor

Fig. 6. Factor loadings for components 1, 2 and 3 (columns) and for modes A, B and C (rows) for the second order chromatographic data of the IPR101 cultivar.
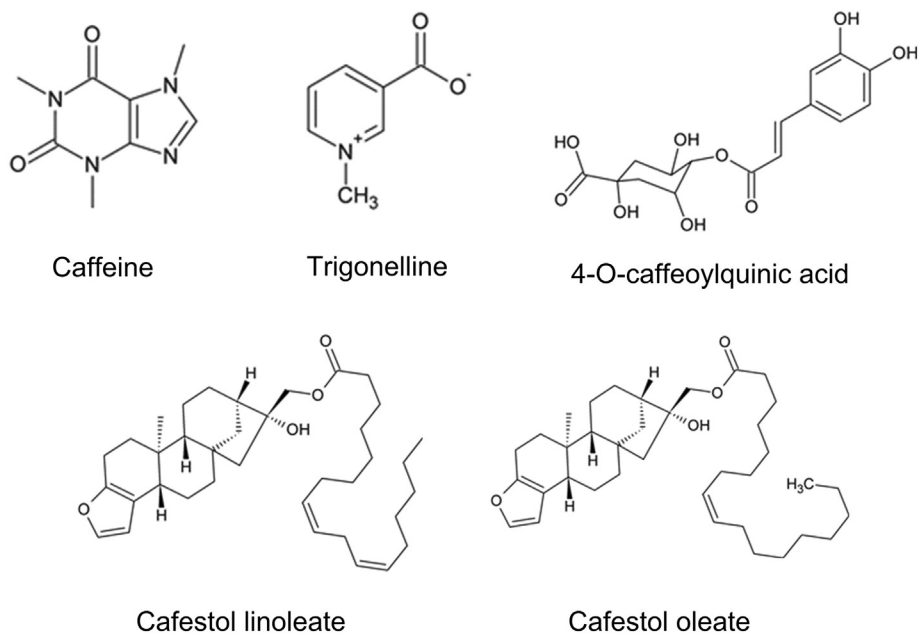

Caffeine

Trigonelline

4-O-caffeoylquinic acid

Cafestol linoleate

Cafestol oleate

Fig. 7. Caffeine, trigonelline, 4-O-caffeoylquinic acid, cafestol linolate and oleate chemical structures.
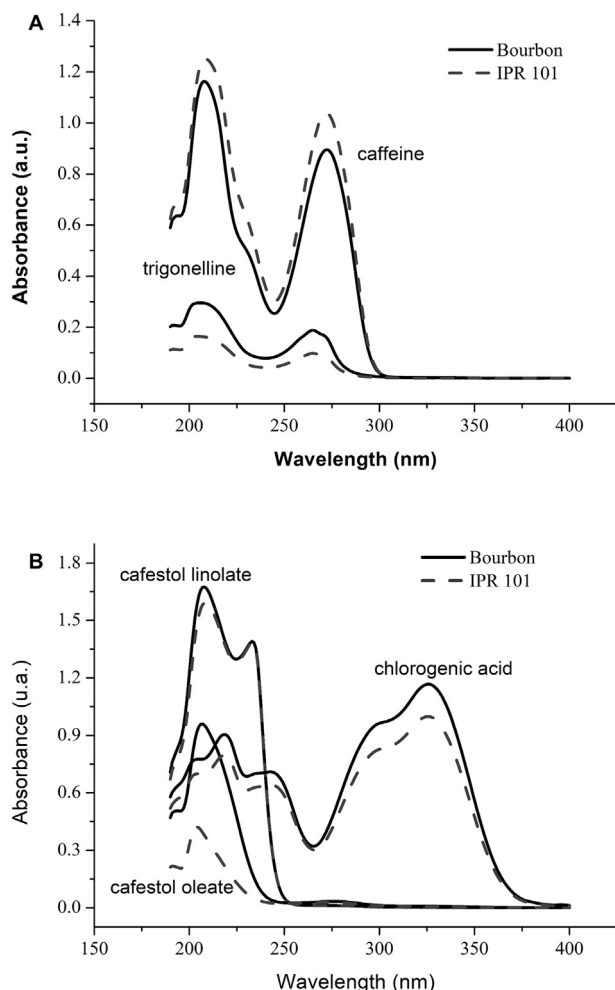
Fig. 8. DAD spectra for A) caffeine and trigonelline B) chlorogenic acids, cafestol oleate and linolate obtained in the best solvent extractor from the PARAFAC model for Bourbon and IPR 101 coffee cultivars.

solvent composition presents the largest absorbance values for the substances represented by the peaks with 2.21 and 4.56 min retention time, although the spectral profile of the 2.21 min peak is more similar to the spectral profile of mode B of the second component, which appears to be associated with the peak with 2.48 min retention time, indicating a possible peak shifting or either substances that belongs to the same class of organic compounds. For the 4.56 min retention time peak, the spectra profile is similar to the loadings of mode B of factor 3. The difference among these spectra and those presented in Fig. 5A and Fig. 5C is that those presented a shoulder at the wavelength of 272 nm. This characteristic indicates that this spectral profile matches the substance trigonelline, whose UV–Vis spectra presents maxima at 265 nm (Berking, 1987).

A PARAFAC model for the IPR101 cultivar with three factors was also determined. Its results are essentially identical to those of the Bourbon cultivar as can be seen in Fig. 6.

To confirm metabolite presence in the Bourbon and IPR101 the ethyl acetate extract was submitted to UPLC-Micromass analysis. In the positive mode $[M + H^+]$ the presence of trigonelline ($m/z = 138.2$), caffeoylquinic acid ($m/z = 355$) and fragment ions ($m/z = 181$ and 193) (Peres, Tonin, Tavares, & Rodriguez-Amaya, 2013), caffeine ($m/z = 195.5$) and cafestol ester ($m/z = 317$) are confirmed. The chemical structures are showed in Fig. 7.

Taking into consideration the almost identical PARAFAC results for the two coffee cultivars and the consequent UPLC-Micromass confirmation of the substances consistent with the chromatograms and

DAD spectra the C mode characterizes the best solvent choice for extracting the various substances indicated by mode B, i.e. ethyl acetate for cafestol, ethanol and its binary mixtures for chlorogenic acids and the ternary edh mixture for caffeine. The relative abundances of these substances can be seen in Fig. 8 A–B. It is evident that the Bourbon cultivar presents higher levels of cafestol ester, chlorogenic acids and trigonelline. On the other hand, the IPR101 cultivar presents higher levels for caffeine. These chemical compounds are correlated with the quality of the coffee beverage. Chlorogenic acids and lipid content are related to the good quality of the beverage, which would be beneficial for the Bourbon cultivar (Novaes, Oigman, De Souza, Rezende, & De Aquino Neto, 2015). Usually its quality is reported to be standard to good (Montagnon, Marraccini, & Bertrand, 2012). Methylxanthines and trigonelline have ambiguous relationships with respect to the quality of the beverage (Novaes et al., 2015).

5. Conclusion

For both cultivars the most suitable PARAFAC model was a three factor model. Factor one was able to discriminate mixtures containing ethyl acetate in the composition of the extractor solvent having a spectrophotometric profile indicating the presence of cafestol ester in the composition. Factor two indicated that mixtures of pure ethanol and binary mixtures containing ethanol were more efficient in extracting the chlorogenic acids. Factor three allowed the identification of methylxanthines by its spectral profile in all mixtures, with larger quantities for the *edh* and *eadh* mixtures for Bourbon and IPR101, respectively. Higher quantities of trigonelline were found for the Bourbon cultivar extracted by *ed* and *eh* mixtures and by the *eh* mixture for the IPR101 cultivar. The crossed IPR101 coffee cultivar when compared to the reference coffee Bourbon cultivar can be highlighted when comparing binary, ternary and quaternary mixtures containing ethanol and ethyl acetate in the composition of the solvent extractor. The reference coffee presents greater abundances of the chemical compounds that are correlated with the higher quality of the beverage when compared to the crossed coffee.

As the three-way PARAFAC strategy determines correlations of HPLC-DAD chromatographic and spectral data simultaneously with samples it permits a more unambiguous assignment of metabolic groups than can be obtained treating chromatographic and spectral data separately by two way methods. This can provide higher quality chromatographic fingerprints for food chemistry. This could lead to direct crosses between new species or improved species against harmful factors thus improving the resistance of the coffee tree increasing not only production but also the quality of the coffee.

References

Andersson, C. A., & Bro, R. (2000). The N-way toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems, 52*(1), 1–4.

Azcarate, S. M., Gomes, A. d. A., Vera-Candioti, L., de Araújo, M. C. U., Camiña, J. M., & Goicoechea, H. C. (2016). Second-order capillary electrophoresis diode array detector data modeled with the Tucker3 algorithm: A novel strategy for Argentinean white wine discrimination respect to grape variety. *Electrophoresis, 37*(13), 1902–1908.

Bajoub, A., Medina-Rodríguez, S., Ajal, E. A., Cuadros-Rodríguez, L., Monasterio, R. P., Vercammen, J., ... Carrasco-Pancorbo, A. (2018). A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication.

Food Research International, 106, 233–242.

Basanta, M., Jarvis, R. M., Xu, Y., Blackburn, G., Tal-Singer, R., Woodcock, A., ... Fowler, S. J. (2010). Non-invasive metabolomic analysis of breath using differential mobility spectrometry in patients with chronic obstructive pulmonary disease and healthy smokers. *Analyst, 135*(2), 315–320.

Belay, A., & Gholap, A. V. (2009). Characterization and determination of chlorogenic acids (CGA) in coffee beans by UV-Vis spectroscopy. *African Journal of Pure and Applied Chemistry, 3*(11), 234–240.

Bennett, R. N., & Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytologist, 127*(4), 617–633.

Berking, S. (1987). Homarine (N-methylpicolinic acid) and trigonelline (N-methylnicotinic acid) appear to be involved in pattern control in a marine hydroid. *Development, 99*(2), 211–220.

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems, 38*(2), 149–171.

Bro, R., & Kiers, H. A. L. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics, 17*(5), 274–286.

Bruns, R. E., Scarminio, I. S., & Neto, B.d. B. (2006). *Statistical design — Chemometrics* (1st ed.). Amsterdam: Elsevier.

Chengying, M., Junxing, L., Wei, C., Wenwen, W., Dandan, Q., Shi, P., & Aiqing, M. (2018). Study of the aroma formation and transformation during the manufacturing process of oolong tea by solid-phase micro-extraction and gas chromatography–mass spectrometry combined with chemometrics. *Food Research International, 108*, 413–422. http://dx.doi.org/10.1016/J.FOODRES.2018.03.052.

Cornell, J. A. (2002). *Experiments with mixtures: Designs, models, and the analysis of mixture data* (3rd ed.). New York: Wiley.

Crozier, A., Clifford, M. N., & Ashihara, H. (Eds.). (2006). *Plant secondary metabolites: Occurrence, structure and role in the human diet*(1st ed.). Oxford: Blackwell Publishing.

Cruz, A. G., Cadena, R. S., Alvaro, M. B. V. B., Sant'Ana, A. S., Oliveira, C. A. F., Faria, J. A. F., ... C, M. M. (2013). Assessing the use of different chemometric techniques to discriminate low-fat and full-fat yogurts. *LWT - Food Science and Technology, 50*, 210–214.

Cruz, A. G., Cadena, R. S., Faria, J. A. F., Bolini, H. M. A., Dantas, C., Ferreira, M. M. C., & Deliza, R. (2012). PARAFAC: Adjustment for modeling consumer study covering probiotic and conventional yogurt. *Food Research International, 45*, 211–215.

Cubero-Leon, E., Peñalver, R., & Maquet, A. (2014). Review on metabolomics for food authentication. *Food Research International, 60*, 95–107.

Custers, D., Van Praag, N., Courselle, P., Apers, S., & Deconinck, E. (2017). Chromatographic fingerprinting as a strategy to identify regulated plants in illegal herbal supplements. *Talanta, 164*, 490–502.

Delaroza, F., Rakocevic, M., Malta, G. B., Bruns, R. E., & Scarminio, I. S. (2014). Spectroscopic and chromatographic fingerprint analysis of composition variations in *Coffea arabica* leaves subject to different light conditions and plant phenophases. *Journal of the Brazilian Chemical Society, 25*(11), 1929–1938.

do Nascimento, L. M., Spehar, C. R., & Sandri, D. (2014). Produtividade de cafeeiro orgânico no cerrado após a poda sob diferentes regimes hídricos. *Coffee Science, 9*(3), 354–365.

Ern, G. L., Moeenfard, M., & Alves, A. (2014). Analytical performance of spectral deconvolution approach for the analysis of diterpene esters in coffee using HPLC-DAD. *Book of Extended Abstracts* (pp. 1–3). Porto: FEUP Edições.

Erny, G. L., Moeenfard, M., & Alves, A. (2015). Liquid chromatography with diode array detection combined with spectral deconvolution for the analysis of some diterpene esters in Arabica coffee brew. *Journal of Separation Science, 38*(4), 612–620.

Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R. N., & Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nature Biotechnology, 18*, 1157–1161.

Filho, E.d. G. A., Rodrigues, T. H. S., Fernandes, F. A. N., Pereira, A. L. F., Narain, N., de Brito, E. S., & Rodrigues, S. (2017). Chemometric evaluation of the volatile profile of probiotic melon and probiotic cashew juice. *Food Research International, 99*, 461–468.

Gomez-Casati, D. F., Zanor, M. I., & Busi, M. V. (2013). Metabolomics in plants and humans: Applications in the prevention and diagnosis of diseases. *BioMed Research International*, 1–11.

Granato, D., Branco, G. F., Faria, J.d. A. F., & Cruz, A. G. (2011). Characterization of Brazilian lager and brown ale beers based on color, phenolic compounds, and antioxidant activity using chemometrics. *Journal of the Science of Food and Agriculture, 91*, 563–571.

Harris, C. S., Burt, A. J., Saleem, A., Le, P. M., Martineau, L. C., Haddad, P. S., ... Arnason, J. T. (2007). A single HPLC-PAD-APCI/MS method for the quantitative comparison of phenolic compounds found in leaf, stem, root and fruit extracts of *Vaccinium angustifolium. Phytochemical Analysis, 18*(2), 161–169.

Hatumura, P. H., de Oliveira, G. S., Marcheafave, G. G., Rakocevic, M., Bruns, R. E., Scarminio, I. S., & Terrile, A. E. (2017). Chemometric analysis of ^{1}H NMR fingerprints of *Coffea arabica* green bean extracts cultivated under different planting densities. *Food Analytical Methods*. http://dx.doi.org/10.1007/s12161-017-1104-y.

Kosmides, A. K., Kamisoglu, K., Calvano, S. E., Corbett, S. A., & Androulakis, I. P. (2013). Metabolomic fingerprinting: Challenges and opportunities. *Critical Reviews in Biomedical Engineering, 41*(3), 205–221.

Kroonenberg, P. M. (2008). *Applied multiway analysis.* New York: John Wiley & Sons Ltd.

Liu, J.-Y., Li, N., Yang, J., Li, N., Qiu, H., Ai, D., ... D, B. (2010). Metabolic profiling of murine plasma reveals an unexpected biomarker in rofecoxib-mediated cardiovascular events. *Proceedings of the National Academy of Sciences, 107*(39), 17017–17022.

López-Martínez, L., López-de-Alba, P. L., García-Campos, R., & De León-Rodríguez, L. M. (2003). Simultaneous determination of methylxanthines in coffees and teas by UV-Vis spectrophotometry and partial least squares. *Analytica Chimica Acta, 493*(1), 83–94.

Martín, M. J., Pablos, F., & González, A. G. (1998). Discrimination between arabica and robusta green coffee varieties according to their chemical composition. *Talanta, 46*(6), 1259–1264.

Matera, J. A., Cruz, A. G., Raices, R. S. L., Silva, M. C., Nogueira, L. C., Quitério, S. L., ... Conte Júnior, C. A. (2014). Discrimination of Brazilian artisanal and inspected pork sausages: Application of unsupervised, linear and non-linear supervised chemometric methods. *Food Research International, 64*, 380–386.

Matlab 6.5 (2000). *Natick.* Massachusetts: The MathWorks Inc.

Moeenfard, M., Erny, G. L., & Alves, A. (2016). Variability of some diterpene esters in coffee beverages as influenced by brewing procedures. *Journal of Food Science and Technology, 53*(11), 3916–3927.

Montagnon, C., Marraccini, P., & Bertrand, B. (2012). Breeding for coffee quality. In T. Oberthür, P. Läderac, H. A. J. Pohlan, & J. H. Cock (Eds.). *Specialty coffee: Managing quality* (pp. 347). (1st ed.). Penang: International Plant Nutrition Institute.

Moreira, I., & Scarminio, I. S. (2013). Chemometric discrimination of genetically modified *Coffea arabica* cultivars using spectroscopic and chromatographic fingerprints. *Talanta, 107*, 416–422.

Mortera, P., Zuljan, F. A., Magni, C., Bortolato, S. A., & Alarcón, S. H. (2018). Multivariate analysis of organic acids in fermented food from reversed-phase high-performance liquid chromatography data. *Talanta, 178*, 15–23.

Nicholson, J. K., & Lindon, J. C. (2008). System biology: Metabolomics. *Nature, 455*, 1054–1056.

Novaes, F. J. M., Oigman, S. S., De Souza, R. O. M. A., Rezende, C. M., & De Aquino Neto, F. R. (2015). New approaches on the analyses of thermolabile coffee diterpenes by gas chromatography and its relationship with cup quality. *Talanta, 139*, 159–166.

Olivieri, A., & Escandar, G. (2014). *Practical three-way calibration* (1st ed.). Amsterdam: Elsevier.

Oms-Oliu, G., Odriozola-Serrano, I., & Martín-Belloso, O. (2013). Metabolomics for assessing safety and quality of plant-derived food. *Food Research International, 54*(1), 1172–1183.

Passari, L. M. Z. G., Scarminio, I. S., & Bruns, R. E. (2014). Experimental designs characterizing seasonal variations and solvent effects on the quantities of coumarin and related metabolites from *Mikania laevigata. Analytica Chimica Acta, 821*, 89–96.

Passari, L. M. Z. G., Scarminio, I. S., Marcheafave, G. G., & Bruns, R. E. (2017). Seasonal changes and solvent effects on fractionated functional food component yields from *Mikania laevigata* leaves. *Food Chemistry*. http://dx.doi.org/10.1016/j.foodchem. 2017.12.075.

Peres, R. G., Tonin, F. G., Tavares, M. F. M., & Rodriguez-Amaya, D. B. (2013). HPLC-DAD-ESI/MS identification and quantification of phenolic compounds in *Ilex paraguariensis* beverages and on-line evaluation of individual antioxidant activity. *Molecules, 18*, 3859–3871.

Pompelli, F., Pompelli, M., de Oliveira, G. F. M., & Antunes, C. (2013). The effect of light and nitrogen availability on the caffeine, theophylline and allantoin contents in the leaves of *Coffea arabica* L. *Aims Environmental Science, 1*(1), 1–11.

Scholz, M. B.d. S., da Silva, J. V. N., de Figueiredo, V. R. G., & Kitzberger, C. S. G. (2013). Atributos sensoriais e características físico-químicas de bebida de cultivares de café do IAPAR. *Coffee Science, 8*(1), 6–16.

Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way analysis: Applications in the chemical sciences.* New York: John Wiley & Sons Ltd.

Soares, P. K., Bruns, R. E., & Scarminio, I. S. (2012). Principal component and Tucker3 analyses of high performance liquid chromatography with diode-array detection fingerprints of crude extracts of *Erythrina speciosa* Andrews leaves. *Analytica Chimica Acta, 736*, 36–44.

Soares, P. K., & Scarminio, I. S. (2008). Multivariate chromatographic fingerprint preparation and authentication of plant material from the genus *Bauhinia. Phytochemical Analysis, 19*(1), 78–85.

Souza, S. S., Cruz, A. G., Walter, E. H. M., Faria, J. A. F., Celeghini, R. M. S., Ferreira, M. M. C., ... Sant'Ana, A.d. S. (2011). Monitoring the authenticity of Brazilian UHT milk: A chemometric approach. *Food Chemistry, 124*, 692–695.

Stanimirova, I., Kita, A., Malkowski, E., John, E., & Walczak, B. (2009). N-way exploration of environmental data obtained from sequential extraction procedure. *Chemometrics and Intelligent Laboratory Systems, 96*(2), 203–209.

Stitt, M., & Fernie, A. R. (2003). From measurements of metabolites to metabolomics: An "on the fly" perspective illustrated by recent studies of carbon-nitrogen interactions. *Current Opinion in Biotechnology, 14*(2), 136–144.

Tang, H., Wang, X., Xu, L., Ran, X., Li, X., Chen, L., ... Liu, X. (2016). Establishment of local searching methods for orbitrap-based high throughput metabolomics analysis. *Talanta, 156–157*, 163–171.

Terrile, A. E., Marcheafave, G. G., Oliveira, G. S., Rakocevic, M., Bruns, R. E., & Scarminio, I. S. (2016). Chemometric analysis of UV characteristic profile and infrared fingerprint variations of *Coffea arabica* green beans under different space management treatments. *Journal of the Brazilian Chemical Society, 27*(7), 1254–1263.

Worku, M., de Meulenaer, B., Duchateau, L., & Boeckx, P. (2018). Effect of altitude on biochemical composition and quality of green arabica coffee beans can be affected by shade and postharvest processing method. *Food Research International, 105*, 278–285.

Zhang, A., Sun, H., Wang, P., Han, Y., & Wang, X. (2012). Modern analytical techniques in metabolomics analysis. *Analyst, 137*(2), 293–300.

Zhang, J.-l., Cui, M., He, Y., Yu, H., & Guo, D. (2005). Chemical fingerprint and metabolic fingerprint analysis of Danshen injection by HPLC–UV and HPLC–MS methods. *Journal of Pharmaceutical and Biomedical Analysis, 36*(5), 1029–1035.

Zhao, M., Nian, Y., Allen, P., Downey, G., Kerry, J. P., & O'Donnell, C. P. (2018). Application of Raman spectroscopy and chemometric techniques to assess sensory characteristics of young dairy bull beef. *Food Research International, 107*, 27–40.