LWT - Food Science and Technology 50 (2013) 496–502

Contents lists available at SciVerse ScienceDirect

LWT - Food Science and Technology

journal homepage: www.elsevier.com/locate/lwt

Discrimination of arabica coffee cultivars by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and chemometrics

Rafael Garrett [a,*], Eduardo M. Schmidt [b], Luiz Filipe P. Pereira [c,d], Cíntia S.G. Kitzberger [c], Maria Brígida S. Scholz [c], Marcos N. Eberlin [b], Claudia M. Rezende [a,*]

[a] Aroma Analysis Laboratory, Institute of Chemistry, Federal University of Rio de Janeiro, 21945-970 Rio de Janeiro, RJ, Brazil
[b] ThoMSon Mass Spectrometry Laboratory, Institute of Chemistry, State University of Campinas, 13083-970 Campinas, SP, Brazil
[c] Instituto Agronômico do Paraná, IAPAR, 86001-970 Paraná, PR, Brazil
[d] Embrapa Café, Brasília, DF, Brazil

ARTICLE INFO

Article history:
Received 5 May 2012
Received in revised form
21 August 2012
Accepted 21 August 2012

Keywords:
Coffea arabica
Genetic background
FT-ICR MS
PCA
PLS-DA

ABSTRACT

Green beans of *Coffea arabica* (arabica coffee) cultivars with Sarchimor and Catuaí genetic background were grown under the same edaphoclimatic conditions, in two different regions, and analysed by direct-infusion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS) followed by a metabolomic approach. A total of 20 coffee metabolites including phenolic compounds, fatty acids, sucrose, and diterpene glycosides were identified by negative ESI FT-ICR MS with a mass error <2 ppm. Furthermore, the multivariate data analysis techniques principal components analysis (PCA) and partial least squares-discriminant analysis (PLS-DA) successfully discriminated the arabica coffee cultivars with Sarchimor from those with Catuaí genetic background, and also revealed the correlations between the coffee metabolites with the arabica cultivars and their growing region. This methodology could be used to identify coffee cultivars according to their genetic background as well as coffees from different growing regions, being a valuable tool for traceability and certification processes.

1. Introduction

Coffea arabica L., commonly known as arabica coffee, is the coffee species most appreciated by consumers. It represents around 70% of the total world coffee production and provides a high-quality brew compared to *Coffea canephora* var. *robusta* due to its intense aroma, low bitterness, and low caffeine content (Lashermes & Anthony, 2007). There are numerous arabica coffee cultivars available and many of them are generated by artificial crosses and mutations to obtain more productive cultures, adapted to the various climates, soil conditions, and resistant to pests and diseases (Sera, 2001). Sarchimor coffee cultivar is derived from the crossing of Villa Sarchi with Timor Hybrid coffee (Silveira et al., 2003). The Timor Hybrid is a natural hybrid of *C. arabica* × *C. canephora* from the island of Timor (Setotaw et al., 2010). The traits inherited from *C. canephora* make Sarchimor derived cultivars resistant to biotic stresses, such as coffee leaf rust (*Hemileia vastatrix* Berk. and Br.), and therefore, more cost efficient than traditional cultivars (Bertrand, Guyot, Anthony, & Lashermes, 2003). Catuaí is a traditional-Brazilian arabica cultivar derived from Caturra x Mundo Novo crossing. It is very productive and account for approximately 40% of all cultivated *C. arabica* in Brazil (Mariuzzo, 2009).

Coffee quality is a demanding analytical task because the genetic background of the plants, associated with agronomic practices and environmental conditions – such as climate, soil type, and altitude – can affect the final chemical composition of the grain (Amorim et al., 2009; Leroy et al., 2006; Vaast, Bertrand, Perriot, Guyot, & Génard, 2006). Furthermore, the genetic variability has been reported to contribute to coffee acidity, sugars, chlorogenic acids, lipids, and caffeine (Guerrero, Suárez, & Moreno, 2001; Kitzberger et al., 2010; Ky et al., 2001).

A powerful way to investigate differences in the profile of secondary metabolites in coffees is the metabolomic approach. Metabolomics is a nontargeted study that involves the characterization of small molecule metabolites in high-dimensional data. By the use of statistical multivariate pattern recognition methods, such as principal components analysis (PCA) and partial least squares-discriminant analysis (PLS-DA), samples can be classified or discriminated, and the biomarkers responsible for samples classification or discrimination can be discovered (Bijlsma et al., 2006; Krastanov, 2010). Ultra-high resolution and mass accuracy mass spectrometers, such as those using Fourier transform ion cyclotron

* Corresponding authors. Tel.: +55 21 2562 7121; fax: +55 21 2562 7106.
E-mail addresses: rafael_garrett@iq.ufrj.br (R. Garrett), crezende@iq.ufrj.br (C.M. Rezende).

0023-6438/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.lwt.2012.08.016

resonance (FT-ICR) analysers, have been successfully applied in metabolomics investigations for many years (Aharoni et al., 2002; Cuadros-Inostroza et al., 2010; Takahashi et al., 2008). FT-ICR MS provides high mass resolution and accuracy with mass errors of 1 ppm or less, which combined with a soft ionization, as electrospray ionization (ESI), enables the determination of molecular formulae from mass measurements only (Corilo et al., 2010; Garrett, Vaz, Hovell, Eberlin, & Rezende, 2012; Marshall & Rodgers, 2004).

Considering the importance of both genetic variability and region of cultivation in coffee quality, and that only a few studies regarding arabica coffee cultivar differentiation based on chemical profile have been described in the literature, we performed the analysis of green beans of arabica coffee cultivars with Sarchimor and Catuaí genetic background grown under the same edaphoclimatic conditions, in two different regions, in Brazil, by direct-infusion ESI(−) FT-ICR MS followed by chemometrics.

2. Material and methods

2.1. Coffee samples and extraction

Five arabica coffee cultivars with Sarchimor genetic background (IAPAR 59, IPR 97, IPR 98, IPR 99 and IPR 104) and three arabica coffee cultivars with Catuaí genetic background (IPR 100, IPR 101 and IPR 105) were grown under the same edaphoclimatic conditions at the Agricultural Technology Park of Cooperative COCARI, in Mandaguari, Paraná state, Brazil (23°32′52″ S, 51°40′15″ W, 655 m above sea level, and average annual temperatures of 20–21 °C). Cherry fruits were selected from the harvest of May to July 2010, washed and sun-dried in patio. Samples were standardized to sieve size of 6.5 mm, had their defective beans removed, and were frozen at −18 °C until analysis. These arabica coffee cultivars described above were also grown under the same edaphoclimatic conditions at the Experimental Field of IAPAR, in Londrina, Paraná state, Brazil (23°18′36″ S, 51°09′56″ W, 585 m above sea level, and average annual temperatures of 21–22 °C). They were harvested and processed similarly as described for the Mandaguari samples.

Green coffee samples (0.5 g) were grounded to sieve size of 0.1 mm in a mill (IKA A11 Basic, Wilmington, USA) and extracted in triplicate with 10 mL of methanol (HPLC grade) for 20 min using an ultrasonic bath (40 kHz, USC-1400, Unique, São Paulo, Brazil). The methanolic extracts (1.0 mL) were centrifuged at 12,100× *g* for 5 min using a microcentrifuge MiniSpin (Eppendorf, Hamburg, Germany). Then, 100 μL of the extracts were diluted in methanol/deionized water (1:1) and used for MS analysis.

2.2. Mass spectrometry and ion identification

Mass spectra fingerprinting and MS/MS data were acquired using a 7.2T LTQ FT Ultra mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a chip-based direct-infusion nanoelectrospray ionization source (Advion BioSciences, Ithaca, NY, USA) operating in the negative-ion mode at the following conditions: capillary voltage 1.6 kV, tube lens −160 V, temperature 270 °C, and fragmentation energy 15–40 eV. Data acquisition was performed along the 100–1000 *m*/*z* range by the Xcalibur 2.0 software.

Identification of the ions was done comparing the *m*/*z* values and MS/MS data obtained by ESI FT-ICR MS in the negative-ion mode with a homemade library of coffee compounds based on literature search (Alonso-Salces, Guillou, & Berrueta, 2009; Amorim et al., 2009; Clarke & Vitzthum, 2001; Clifford, Johnston, Knight, & Kuhnert, 2003; Jaiswal & Kuhnert, 2010) and standards. We considered a match between the experimental *m*/*z* value and the theoretical *m*/*z* value from our library when the mass error was <3 ppm. The isotope distribution pattern of the ions identified was also considered with the proposed chemical formula.

2.3. Multivariate analysis of data

Tables of *m*/*z* values and relative intensities containing the fifty more abundant ions from each sample (which represented, approximately, a relative intensity higher than 1%) were exported from the Xcalibur software, saved as *.csv* files and uploaded into the MetaboAnalyst web server (http://www.metaboanalyst.ca; Xia, Psychogios, Young, & Wishart, 2009) for multivariate analyses. The data were aligned using a mass tolerance of 0.005 *m*/*z* and treated by Pareto scaling to reduce the differences between large and small relative intensities of *m*/*z* values (van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006). Four matrices, where each line represented a sample and each column a variable, were generated and submitted to principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA). Leave-one-out cross-validation (LOOCV) was employed to prevent overfitting and estimate the quality of the PLS-DA analysis. The matrices were as follow: (a) the arabica coffee cultivars grown in Londrina; (b) the arabica coffee cultivars grown in Mandaguari; (c) The arabica coffee cultivars with Sarchimor genetic background grown in Londrina and Mandaguari; (d) the arabica coffee cultivars with Catuaí genetic background grown in Londrina and Mandaguari.

3. Results and discussion

3.1. ESI(−) FT-ICR MS analysis

Fig. 1 shows the ESI(−) FT-ICR mass spectra for the methanolic extract of sample IAPAR 59 (arabica coffee cultivar with Sarchimor genetic background) and sample IPR 105 (arabica coffee cultivar with Catuaí genetic background). Although basically the same set of ions was detected in all coffee samples, the distinction between the studied cultivars of arabica coffee was achieved due to significant and reproducible differences in relative intensities of ions.

Analysis of green coffee beans was chosen rather than analysis of roasted beans to avoid changes in coffee composition due to the roasting process.

The *m*/*z* values of the data matrices of coffees samples grown in Mandaguari and Londrina generated during multivariate analyses were used for compounds identification. A total of 20 compounds, including phenolics, lipids, and diterpenes were identified with a mass error <2 ppm (Table 1) via comparison of the experimental *m*/*z* values with the theoretical *m*/*z* values from our coffee library. Additional confirmation of the structural assignments was done via comparison of the experimental isotopic patterns of the deprotonated molecules with the theoretical ones generated by the Xcalibur software. A mass accuracy of 3 ppm and 2% of isotopic pattern accuracy can usually remove more than 95% of false candidates and significantly reduce the number of possible molecular formulae (Kind & Fiehn, 2006). In addition, dissociation of the deprotonated molecules via MS/MS experiments was also employed to gain more confidence for the structural identification. A feature that points to the reliability of our approach is that all compounds identified have already been described in the literature for coffee samples.

Six diterpene glycosides, known as atractyloside (ATR) and carboxyatractyloside (CATR) analogues were found in all coffee cultivars. Basically, this type of diterpenoid glycosides consists of an aglycone with a perhydrophenanthrene structure and a glycoside moiety made up of glucose with sulphate and/or isovalerate (Obatomi & Bach, 1998). ATRs were first isolated from the roots of *Atractylis gummifera* L. (Asteraceae) and can also be found in other

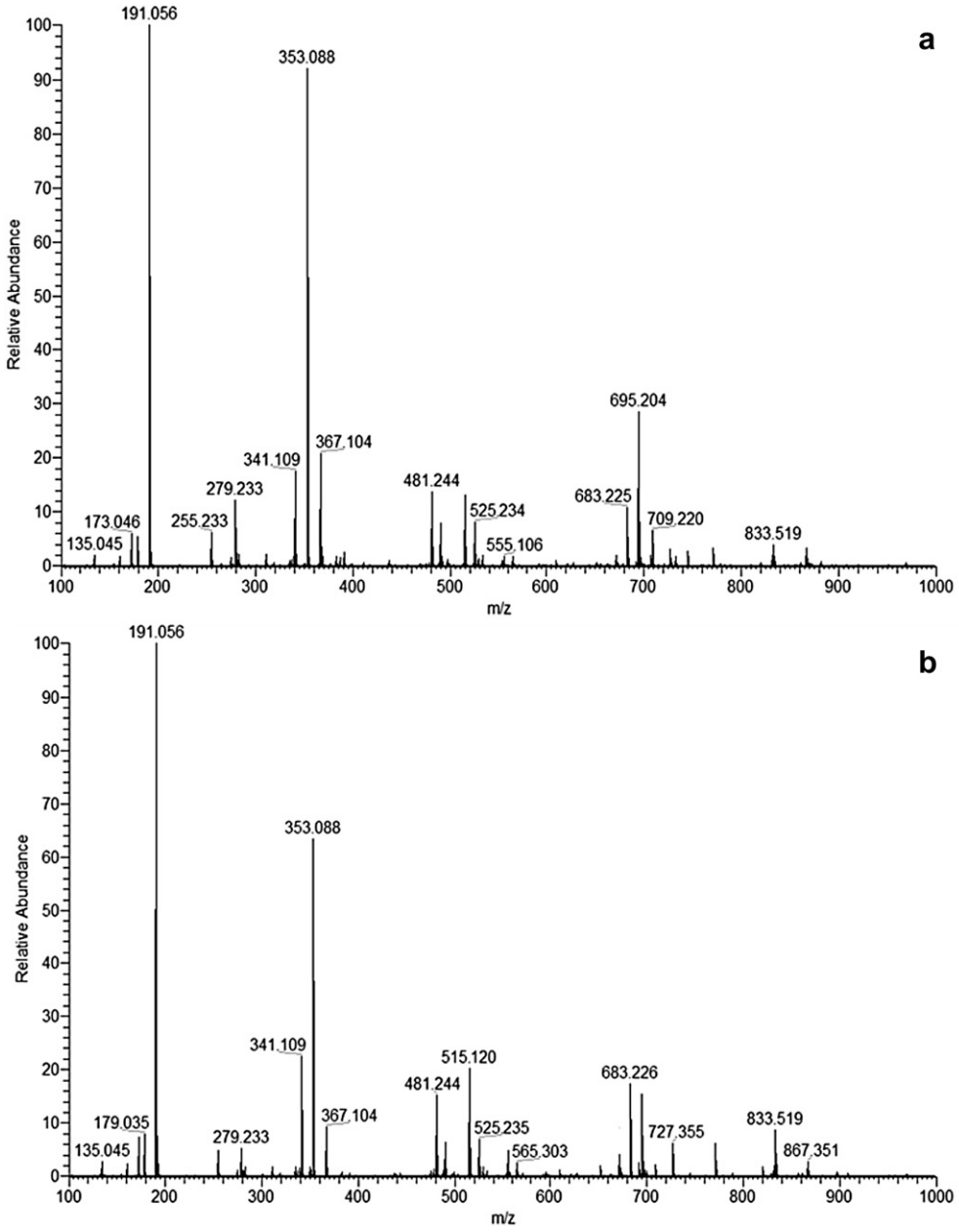


Fig. 1. ESI(−) FT-ICR mass spectra for the methanolic extract of samples: (a) IAPAR 59 (arabica coffee cultivar with Sarchimor genetic background); (b) IPR 105 (arabica coffee cultivar with Catuaí genetic background).

plants as *Callilepis laureola* DC. and *Xanthium strumarium* L. (Obatomi & Bach, 1998). The occurrence of ATR and CATR analogues in the extracts of green and roasted coffees was first described in the 1970s (Obermann & Spiteller, 1976; Richter & Spiteller, 1978). Fig. 2 shows the structure of the ATR and CATR analogues identified in the arabica coffee cultivars and the main fragments of their deprotonated molecules observed by the MS/MS experiments. When the CATRs were fragmented, they generated their respective ATRs by decarboxylation. For the ATR analogues, fragmentation led to the breakdown of the glycosidic linkage and the isovaleric ester linkage.

Two compounds with *m*/*z* values of 695.20442 and 833.51922 were indicated by multivariate analyses (see Section 3.2) as very important for discriminating the coffee cultivars. However, they could not be fully characterized. MS/MS experiments with these two deprotonated molecules revealed the presence of fragments of *m*/*z* 353 (caffeoylquinic acid or quinic acid hexoside) and 191 (quinic acid), which points to derivatives of such acids. All the attempts made to identify these two compounds failed when a mass error <3 ppm and an isotopic pattern error of $<10\%$ were considered.

3.2. Multivariate analyses

PCA was employed as a first approach in multivariate analysis to get an overview of the data and to identify possible outliers. PCA is an unsupervised method that aims to find the maximum variation within the dataset (*X*) without referring to any class labels (*Y*). The score plot generated by PCA displays the principal components (PCs) responsible for the maximum variation within the data and

Table 1
Compounds identified in the methanolic extracts of all arabica coffee cultivars analysed by ESI(−) FT-ICR MS.

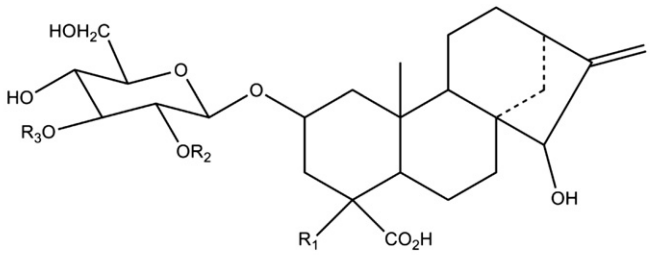
Compounds	Formula	Theoretical *m/z*	Experimental *m/z*	Error (ppm)	RI[c] (Mean ± SD)
Caffeic acid	$C_9H_8O_4$	179.03498	179.03502	0.22	6.33 ± 1.19
Ferulic acid	$C_{10}H_{10}O_4$	193.05063	193.05069	0.31	2.43 ± 0.23
Quinic acid	$C_7H_{12}O_6$	191.05611	191.05613	0.10	100.00 ± 0.00
Caffeoylquinic acid[a]	$C_{16}H_{18}O_9$	353.08781	353.08779	−0.06	76.20 ± 7.55
Feruloylquinic acid	$C_{17}H_{20}O_9$	367.10346	367.10360	0.38	16.25 ± 3.33
diCaffeoylquinic acid	$C_{25}H_{24}O_{12}$	515.11950	515.11980	0.58	17.59 ± 3.45
Feruloylcaffeoylquinic acid	$C_{26}H_{26}O_{12}$	529.13515	529.13568	1.00	2.01 ± 0.36
Palmitic acid	$C_{16}H_{32}O_2$	255.23295	255.23306	0.43	9.50 ± 4.57
Linoleic acid	$C_{18}H_{32}O_2$	279.23295	279.23303	0.29	16.15 ± 8.32
Oleic acid	$C_{18}H_{34}O_2$	281.24860	281.24875	0.53	4.13 ± 1.77
Stearic acid	$C_{18}H_{36}O_2$	283.26425	283.26439	0.49	3.32 ± 1.09
Arachidic acid	$C_{20}H_{40}O_2$	311.29555	311.29572	0.55	3.04 ± 0.95
Behenic acid	$C_{22}H_{44}O_2$	339.32685	339.32708	0.68	3.07 ± 0.85
Sucrose[b]	$C_{12}H_{22}O_{11}$	341.10893	341.10904	0.32	21.42 ± 6.59
Atractyloside II	$C_{25}H_{38}O_9$	481.24431	481.24446	0.31	14.44 ± 2.16
Carboxyatractyloside II	$C_{26}H_{38}O_{11}$	525.23414	525.23461	0.90	7.40 ± 0.87
Atractyloside III	$C_{30}H_{46}O_{10}$	565.30182	565.30252	1.24	3.37 ± 1.93
Carboxyatractyloside III	$C_{31}H_{46}O_{12}$	609.29165	609.29241	1.25	2.37 ± 0.35
Atractyloside I	$C_{36}H_{56}O_{15}$	727.35464	727.35541	1.06	5.03 ± 2.13
Carboxyatractyloside I	$C_{37}H_{56}O_{17}$	771.34447	771.34496	0.64	5.31 ± 1.95

[a] $[2M - H]^- = 707.18362$ (error: 1.05 ppm).
[b] $[2M - H]^- = 683.22553$ (error: 0.57 ppm).
[c] RI: relative intensities of ions in the coffee samples.

highlights clustering or outliers, while the loading plot stresses the influence of the variables in the PCs. Variables with high absolute loading values on each PC have a strong contribution to that PC (Takami, Mahmoudi, Dahlquist, & Lindenmo, 2011).

PLS-DA was used to investigate the differences between the arabica coffee cultivars and the growing region of each cultivar. Unlike PCA, PLS-DA is a supervised method that explains the maximum variation in the dataset (*X*) using a dummy matrix (*Y*) that describes the class of samples (Barker & Rayens, 2003). Variable importance in projection (VIP), a weighted sum of squares of the PLS loadings taking into account the amount of explained *Y*-variation in each dimension, was employed to indicate the most important variables (*m/z* values), with VIP values larger than 1, responsible for differences in the groups studied (the arabica coffee samples with Sarchimor genetic background *vs.* the arabica coffee samples with Catuaí genetic background; Londrina *vs.* Mandaguari) (Mazzara et al., 2011).

From a matrix of 23 samples and 47 variables, the variables with high absolute loading values, and thus, being the most important for the separation achieved in the PCA analysis of coffees with Sarchimor and Catuaí genetic background grown under the same edaphoclimatic conditions, in Londrina, were the ions of *m/z* 255.23295, 279.23295, 341.10893, 353.08781, 515.11950, 695.20442, and 833.51922 (Fig. 3b). The PCA score plot revealed that PC2, which explained 24.8% of the total data variation, was responsible to distinguish the coffees with Catuaí genetic background from those with Sarchimor genetic background (Fig. 3a). Samples IAPAR 59 and IPR 104 were positively correlated with variables 353.08781 and 695.20442, but negatively correlated with variables 515.11950 and 833.51922. This means that when the levels of caffeoylquinic acid (CQA; 353.08781) and the deprotonated molecule $[M - H]^-$ of *m/z* 695.20442 (not identified) are high in these samples, the levels of dicaffeoylquinic acid (diCQA; 515.11950) and the $[M - H]^-$ of *m/z* 833.51922 (not identified) are low. The opposite was observed for samples IPR 101 and 105, which showed high levels of the $[M - H]^-$ of *m/z* 833.51922 and diCQA, but low levels of CQA and the $[M - H]^-$ of *m/z* 695.20442. Samples IPR 97 and 99 were positively correlated with variables 255.23295,



Compounds	R_1	R_2	R_3	$[M-H]^-$	MS/MS		Other fragments
CATR I	CO_2H	$COCH_2CH(CH_3)_2$	Glucose	771	$[M-H-CO_2H]^-$	727	391
ATR I	H	$COCH_2CH(CH_3)_2$	Glucose	727	$[M-H-COCH_2CH(CH_3)_2]^-$	643	625,481
CATR III	CO_2H	$COCH_2CH(CH_3)_2$	H	609	$[M-H-CO_2H]^-$	565	481
ATR III	H	$COCH_2CH(CH_3)_2$	H	565	$[M-H-COCH_2CH(CH_3)_2]^-$	481	463, 301
CATR II	CO_2H	H	H	525	$[M-H-CO_2H]^-$	481	396
ATR II	H	H	H	481	[M-H-Glucose]$^-$	319	301,179,161

Note: ATR: Atractyloside; CATR: Carboxyatractyloside

Fig. 2. Structure of the atractyloside (ATR) and carboxyatractyloside (CATR) analogues found in the arabica coffee cultivars and the main fragments observed by ESI(−) MS/MS.

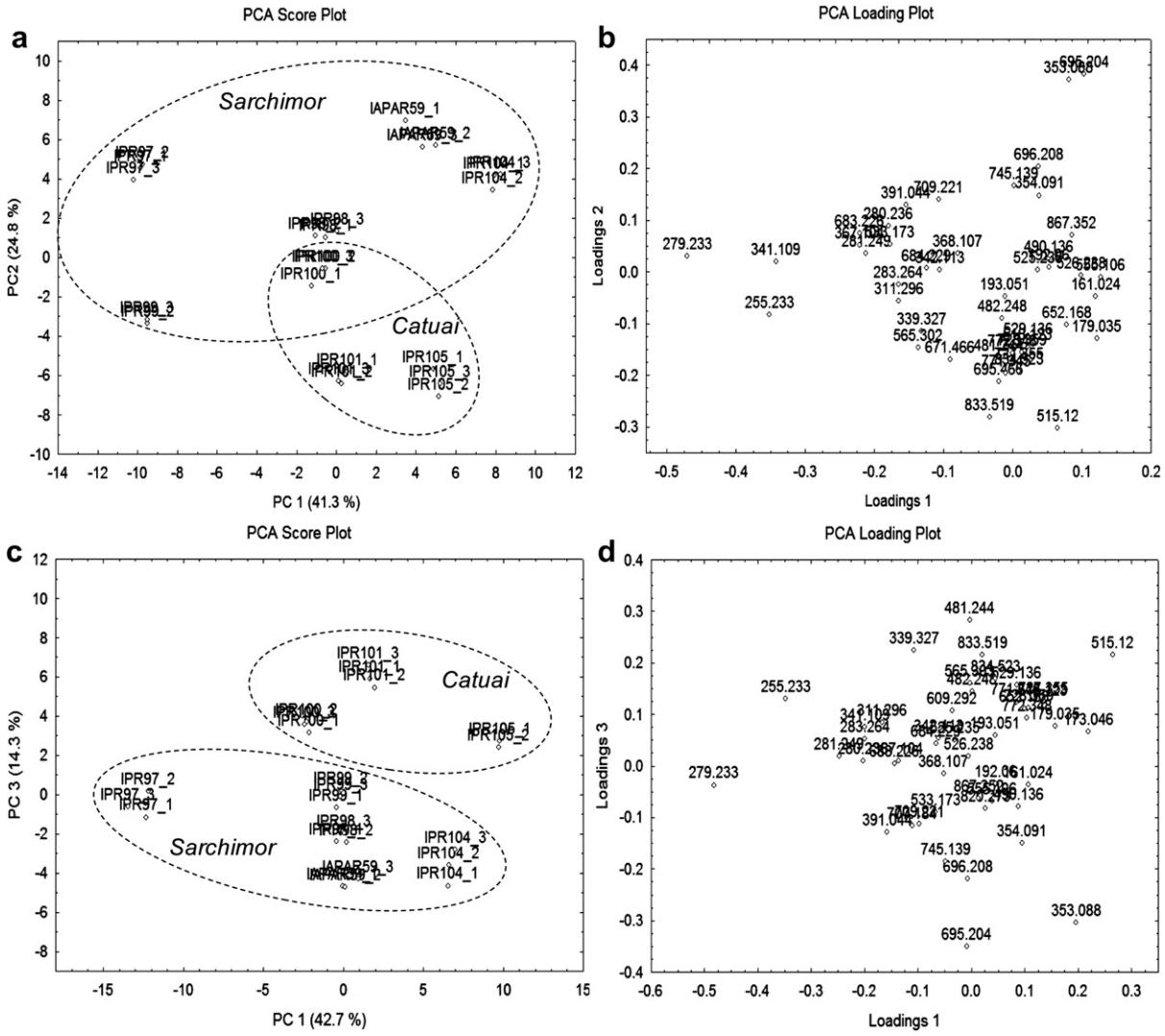


Fig. 3. PCA score and loading plots of arabica coffee cultivars with Sarchimor and Catuaí genetic background grown under the same edaphoclimatic in (a,b) Londrina and (c,d) Mandaguari regions.

279.23295, and 341.10893 by PC1, showing high levels of palmitic acid, linoleic acid and sucrose, respectively.

PLS-DA analysis improved the discrimination between the arabica coffee cultivars grown in Londrina (Fig. 4a). The VIP scores indicated the $[M - H]^-$ of m/z 695.20442, CQA, feruloylquinic acid, and sucrose as the most important compounds for the coffees with Sarchimor genetic background. Investigations on the original dataset confirmed that these compounds had the highest relative intensities in all coffee samples with Sarchimor genetic background. For the coffees with Catuaí genetic background, the most important compounds indicated by VIP scores were diCQA, the $[M - H]^-$ of m/z 833.51922, atractyloside analogue II (ATR II), and caffeic acid. The $[M - H]^-$ of m/z 695.20442 and 833.51922, CQA, and diCQA have already been recognized by the PCA analysis as important for the discrimination between the arabica coffee cultivars grown in Londrina. The $R2$ and $Q2$ values obtained by leave-one-out cross-validation (LOOCV) using 6 components were 0.990 and 0.968, respectively.

From a matrix of 23 samples and 50 variables, discrimination between the arabica coffee cultivars grown under same edaphoclimatic conditions, in Mandaguari, was only achieved by PC3, which explained 14.31% of the total data variation (Fig. 3c). The variables with high absolute loading values were 255.23295, 279.23295, 353.08781, 481.24431, 515.11950, and 695.20442 (Fig. 3d). Again, samples IAPAR 59 and IPR 104 were positively correlated with variables 353.08781(CQA) and 695.20442; sample IPR 97 was positively correlated with variables 255.23295 (palmitic acid) and 279.23295 (linoleic acid); and sample IPR 105 was positively correlated with variable 515.11950 (diCQA). Differently from Londrina, in Mandaguari, samples IPR 100 and 101 were positively correlated with variable 481.24431 (ATR II), but negatively correlated with variables 353.08781(CQA) and 695.20442.

Good discrimination between the arabica coffee cultivars grown in Mandaguari was achieved by PLS-DA (Fig. 4b). VIP scores indicated the $[M - H]^-$ of m/z 695.20442, CQA, sucrose, and linoleic acid as the most important compounds for the coffees with Sarchimor genetic background, whereas diCQA, ATR II and feruloylcaffeoylquinic acid were the most important compounds for the coffees with Catuaí genetic background. As seen in the PLS-DA analysis of samples grown in Londrina, the $[M - H]^-$ of m/z 695.20442, compounds CQA, sucrose, diCQA, and ATR II were the most important compounds responsible for the discrimination between the arabica coffee cultivars grown in Mandaguari. The $R2$ and $Q2$ values obtained using 5 components in LOOCV were 0.991 and 0.966, respectively.

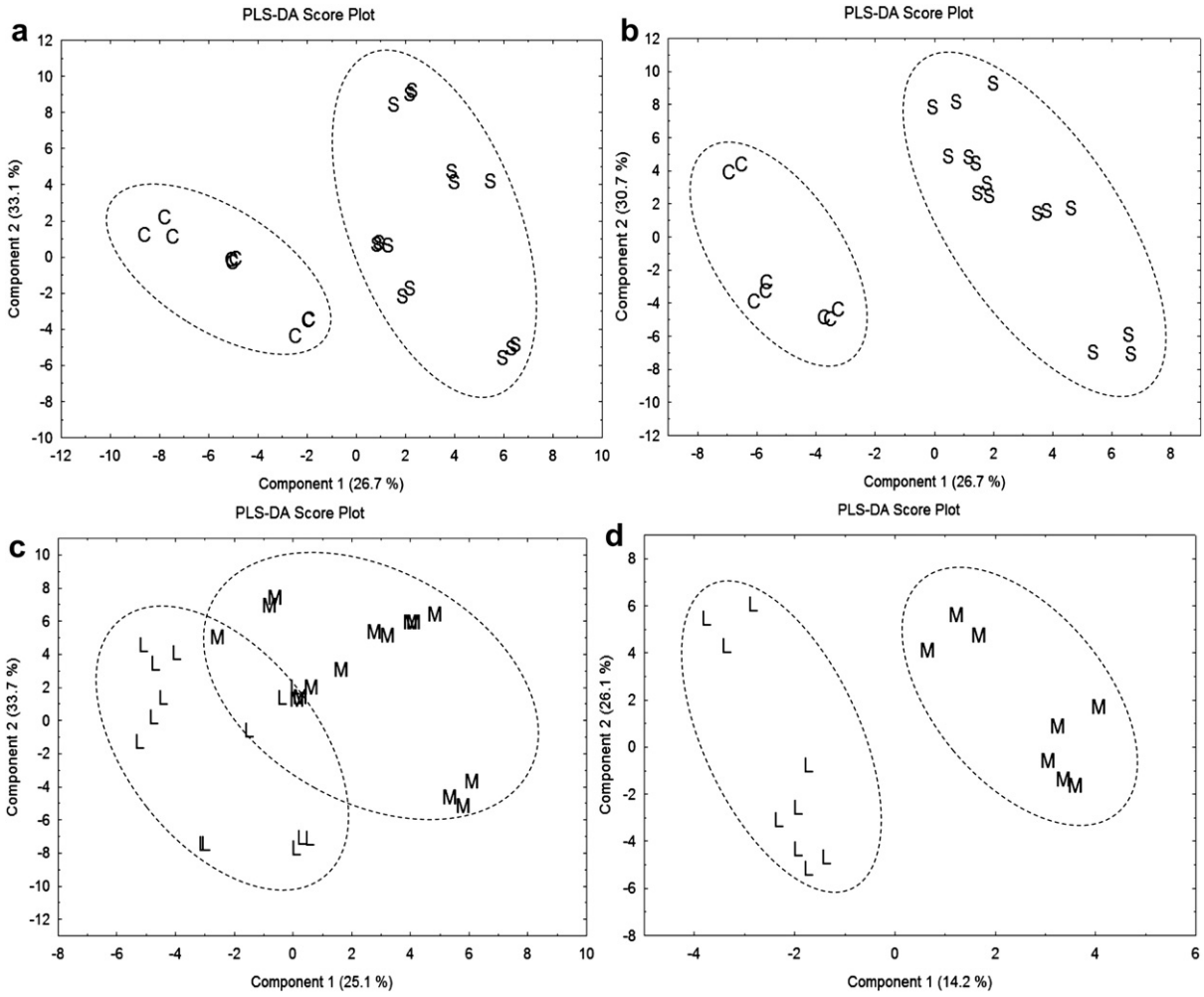


Fig. 4. PLS-DA score plots of arabica coffee cultivars with Sarchimor (**S**) and Catuaí (**C**) genetic background grown under the same edaphoclimatic in: (a) Londrina; (b) Mandaguari regions. PLS-DA score plots of arabica coffee cultivars with Sarchimor (c) and Catuaí (d) genetic background grown in both Londrina (**L**) and Mandaguari (**M**) regions.

In both Londrina and Mandaguari regions, samples were grown under the same edaphoclimatic conditions, harvested and processed in the same way. Differences in the chemical compositions of coffee samples were expected to be mostly due to the genetic variability of the arabica coffees. The arabica coffee cultivars with Sarchimor genetic background showed higher levels of CQA and sucrose compared to those cultivars with Catuaí genetic background. While high levels of sucrose are expected for *C. arabica*, high levels of CQA found in the coffees with Sarchimor genetic background may be due to its traits inherited from *C. canephora*, which is recognized for the high levels of chlorogenic acids (mainly CQA). Furthermore, it is known that the introgression of *C. canephora* genes into *C. arabica* via the Timor Hybrid may result in coffees with high levels of CQA, which causes a drop in the beverage quality (Bertrand et al., 2003). In contrast, high levels of diCQA as found in the arabica coffee cultivars with Catuaí genetic background have been associated with good beverage quality (Farah, Monteiro, Calado, Franca, & Trugo, 2006).

When the influence of coffee growing region was investigated, PCA analysis failed to discriminate the same coffee cultivar grown in both Londrina and Mandaguari. This objective was only achieved by the supervised technique PLS-DA (Fig. 4c, d). The less effective separation observed in the PLS-DA score plot between the coffees with Sarchimor genetic background grown in both Londrina and Mandaguari (Fig. 4c), compared to the coffees with Catuaí genetic background grown in both Londrina and Mandaguari (Fig. 4d), may indicate that the Sarchimor cultivars are less influenced by the growing region.

The PLS-DA VIP scores indicated that for the coffees with Sarchimor genetic background, Londrina region provided higher levels of the $[M-H]^-$ of m/z 695.20442, sucrose, CQA, diCQA, and feruloylquinic acid; whereas Mandaguari region provided higher levels of ATR II. The $R2$ and $Q2$ obtained using 7 components were 0.961 and 0.763, respectively. For the coffees with Catuaí genetic background, Londrina region provided higher levels of the $[M-H]^-$ of m/z 695.20442 and 833.51922, sucrose, and CQA; while Mandaguari region provided higher levels of ATR II and CATR II. Using 5 components in LOOCV, the $R2 = 0.989$ and $Q2 = 0.949$.

The lower altitude and warmer climatic conditions in Londrina, compared to Mandaguari, usually promotes a faster coffee bean development and fruit ripening, and thus, may explain the high levels of sucrose and chlorogenic acids of its coffees (Vaast et al., 2006).

The most important compounds, responsible for discriminations in all PLS-DA analyses were: the unidentified $[M-H]^-$ of m/z 695.20442, ATR II, sucrose, and the chlorogenic acids CQA and diCQA. The diterpene ATR II was highly related to coffees with Catuaí genetic background and Mandaguari place, while sucrose,

CQA, and the $[M - H]^-$ of *m/z* 695.20442 were highly related to coffees with Sarchimor genetic background and Londrina place. Compound diCQA was highly related to coffees with Catuaí genetic background and Londrina place.

4. Conclusion

A simple and rapid coffee extraction employing ultrasonic bath, and the direct analysis of the extracts by ESI(−) FT-ICR MS allowed the identification of 20 coffee metabolites, including phenolic compounds, fatty acids, sucrose, and diterpene glycosides. The metabolomic approach using PCA and PLS-DA multivariate analysis techniques successfully distinguish the arabica coffee cultivars with Sarchimor from those with Catuaí genetic background grown under the same edaphoclimatic conditions, and also the coffee growing region, with around 100 km of distance only. In addition, PLS-DA VIP scores indicated which compounds were responsible for the discriminations.

Association of these data with gene expression profile can be very useful to better understand the coffee metabolism and to select cultivars with desirable features for a good coffee brew. Besides, with the actual and future coffee markets requirement, where discussion for "terroir" characterization are underway, this methodology, associated to sensory analysis, has the potential to be an extremely value tool for origin certification processes as well as for traceability of coffee beans.

Acknowledgements

The authors thank Dr. Tumoru Sera and Dr. João Batista G. D. Silva for providing access to plant materials, and the Brazilian Coffee Research Consortium, CNPq, CAPES, FAPERJ, and ANP for financial support.

References

Aharoni, A., Ric de Vos, C. H., Verhoeven, H. A., Maliepaard, C. A., Kruppa, G., Bino, R., et al. (2002). Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. *OMICS: A Journal of Integrative Biology, 6*, 217–234.

Alonso-Salces, R. M., Guillou, C., & Berrueta, L. A. (2009). Liquid chromatography coupled with ultraviolet absorbance detection, electrospray ionization, collision-induced dissociation and tandem mass spectrometry on a triple quadrupole for the on-line characterization of polyphenols and methylxanthines in green coffee beans. *Rapid Communications in Mass Spectrometry, 23*, 363–383.

Amorim, A. C. L., Hovell, A. M. C., Pinto, A. C., Eberlin, M. N., Arruda, N. P., Pereira, E. J., et al. (2009). Green and roasted arabica coffees differentiated by ripeness, process and cup quality via electrospray ionization mass spectrometry fingerprinting. *Journal of the Brazilian Chemical Society, 20*, 313–321.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics, 17*, 166–173.

van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

Bertrand, B., Guyot, B., Anthony, F., & Lashermes, P. (2003). Impact of the *Coffea canephora* gene introgression on beverage quality of *C. arabica*. *Theoretical and Applied Genetics, 107*, 387–394.

Bijlsma, S., Bobeldijk, I., Verheij, E. R., Ramaker, R., Kochhar, S., Macdonald, I. A., et al. (2006). Large-scale human metabolomics studies: a strategy for data (pre-) processing and validation. *Analytical Chemistry, 78*, 567–574.

Clarke, R. J., & Vitzthum, O. G. (2001). *Coffee, recent developments*. Oxford: Blackwell Science.

Clifford, M. N., Johnston, K. L., Knight, S., & Kuhnert, N. (2003). Hierarchical scheme for LC–MS^n identification of chlorogenic acids. *Journal of Agricultural and Food Chemistry, 51*, 2900–2911.

Corilo, Y. E., Vaz, B. G., Simas, R. C., Lopes Nascimento, H. D., Klitzke, C. F., Pereira, R. C. L., et al. (2010). Petroleomics by EASI(±) FT-ICR MS. *Analytical Chemistry, 82*, 3990–3996.

Cuadros-Inostroza, A., Giavalisco, P., Hummel, J., Eckardt, A., Willmitzer, L., & Peña-Cortés, H. (2010). Discrimination of wine attributes by metabolome analysis. *Analytical Chemistry, 82*, 3573–3580.

Farah, A., Monteiro, M. C., Calado, V., Franca, A. S., & Trugo, L. C. (2006). Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry, 98*, 373–380.

Garrett, R., Vaz, B. G., Hovell, A. M. C., Eberlin, M. N., & Rezende, C. M. (2012). Arabica and robusta coffees: identification of major polar compounds and quantification of blends by direct-infusion electrospray ionization-mass spectrometry. *Journal of Agricultural and Food Chemistry, 60*, 4253–4258.

Guerrero, G., Suárez, M., & Moreno, G. (2001). Chlorogenic acids as a potential criterion in coffee genotype selections. *Journal of Agricultural and Food Chemistry, 49*, 2454–2458.

Jaiswal, R., & Kuhnert, N. (2010). Hierarchical scheme for liquid chromatography/multi-stage spectrometric identification of 3,4,5-triacylchlorogenic acids in green Robusta coffee beans. *Rapid Communications in Mass Spectrometry, 24*, 2283–2294.

Kind, T., & Fiehn, O. (2006). Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics, 7*, 234.

Kitzberger, C. S. G., Scholz, M. B. S., Pereira, L. F. P., Vieira, L. G. E., Sera, T., Silva, J. B. G. D., et al. (2010). Analysis of diterpens in green and roasted coffee of *Coffea arabica* cultivars growing in the same edapho-climatic conditions. *Association for Science and Information on Coffee, 1*, 1–7.

Krastanov, A. (2010). Metabolomics – the state of art. *Biotechnology & Biotechnological Equipment., 24*, 1537–1543.

Ky, C.-L., Louarn, J., Dussert, S., Guyot, B., Hamon, S., & Noirot, M. (2001). Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. *Food Chemistry, 75*, 223–230.

Lashermes, P., & Anthony, F. (2007). Coffee. In C. Kole (Ed.), *Technical crops: Genome mapping and molecular breeding in plants* (pp. 109–118). Berlin, Heidelberg: Springer.

Leroy, T., Ribeyre, F., Bertrand, B., Charmetant, P., Dufour, M., Montagnon, C., et al. (2006). Genetics of coffee quality. *Journal of Plant Physiology, 18*, 229–242.

Mariuzzo, P. (2009). Coffee made in Brazil is grown by the world's largest producers. *Ciência e Cultura, 61*, 14–16.

Marshall, A. G., & Rodgers, R. P. (2004). Petroleomics: the next grand challenge for chemical analysis. *Accounts of Chemical Research, 37*, 53–59.

Mazzara, S., Cerutti, S., Iannaccone, S., Conti, A., Olivieri, S., Alessio, M., et al. (2011). Application of multivariate data analysis for the classification of two dimensional gel images in neuroproteomics. *Journal of Proteomics & Bioinformatics, 4*, 016–021.

Obatomi, D. K., & Bach, P. H. (1998). Biochemistry and toxicology of the diterpenoid glycoside atractyloside. *Food and Chemical Toxicology, 36*, 335–346.

Obermann, H., & Spiteller, G. (1976). The structures of the "Coffee Atractylosides". *Chemische Berichte, 109*, 3450–3461.

Richter, H., & Spiteller, G. (1978). A new atractyligenine glycoside from green coffee-beans. *Chemische Berichte, 111*, 3506–3509.

Sera, T. (2001). Coffee genetic breeding at IAPAR. *Crop Breeding and Applied Biotechnology, 1*, 179–199.

Setotaw, T. A., Caixeta, E. T., Pena, G. F., Zambolim, E. M., Pereira, A. A., & Sakiyama, N. S. (2010). Breeding potential and genetic diversity of "Híbrido do Timor" coffee evaluated by molecular markers. *Crop Breeding and Applied Biotechnology, 10*, 298–304.

Silveira, S. R., Ruas, P. M., Ruas, C. F., Sera, T., Carvalho, V. P., & Coelho, A. S. G. (2003). Assessment of genetic variability within and among coffee progenies and cultivars using RAPD markers. *Genetics and Molecular Biology, 26*, 329–336.

Takahashi, H., Kai, K., Shinbo, Y., Tanaka, K., Ohta, D., Oshima, T., et al. (2008). Metabolomics approach for determining growth-specific metabolites based on Fourier transform ion cyclotron resonance mass spectrometry. *Analytical and Bioanalytical Chemistry, 391*, 2769–2782.

Takami, K. M., Mahmoudi, J., Dahlquist, E., & Lindenmo, M. (2011). Multivariable data analysis of a cold rolling control system to minimise defects. *The International Journal of Advanced Manufacturing Technology, 54*, 553–565.

Vaast, P., Bertrand, B., Perriot, J.-J., Guyot, B., & Génard, M. (2006). Fruit thinning and shade improve bean characteristics and beverage quality of coffee (*Coffea arabica* L.) under optimal conditions. *Journal of the Science of Food and Agriculture, 86*, 197–204.

Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research, 37*, W652–W660.