High-Performance Liquid Chromatography with Fluorescence Detection (HPLC-FLD) Fingerprints as Chemical Descriptors to Authenticate the Origin, Variety and Roasting Degree of Coffee by Multivariate Chemometric Methods.

Running title: HPLC-FLD Fingerprinting for Coffee Authentication

Nerea Núñez[a,*], Clara Martínez[a], Javier Saurina[a,b] and Oscar Núñez[a,b,c,*]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E08028 Barcelona, Spain.

[b] Research Institute in Food Nutrition and Food Safety, University of Barcelona. Av. Prat de la Riba 171, Edifici Recerca (Gaudí), E-08921 Santa Coloma de Gramenet, Barcelona, Spain.

[c] Serra Hunter Fellow. Generalitat de Catalunya, E-08007 Barcelona, Spain.

* Corresponding authors: Ms. Nerea Núñez, Dr. Oscar Núñez

Department of Chemical Engineering and Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E08028 Barcelona, Spain.

Phone: +34-93-4033706

Fax: +34-93-4021286

e-mails: nereant7@gmail.com; oscar.nunez@ub.edu

Abstract

BACKGROUND: Coffee is one of the most popular beverages around the world, consumed as an infusion of ground roasting coffee beans with a characteristic taste and flavor. Two main varieties, Arabica and Robusta, are worldwide produced. Besides, the interest of consumers in quality attributes related to coffee production region and varieties is increasing, being necessary encouraging the development of simple methodologies to authenticate and to guarantee the coffee origin, variety, as well as the roasting degree to prevent fraudulent practices.

RESULTS: C18 high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprints obtained after brewing the coffees without any sample treatment other than filtration (considerably reducing sample manipulation) were employed as sample chemical descriptors for coffee characterization and classification by principal component analysis (PCA) and partial least squares regression-discriminant analysis (PLS-DA). PLS-DA showed good classification capabilities regarding coffee origin, variety and roasting degree when employing HPLC-FLD fingerprints although overlapping for some sample groups occurred. However, the discrimination power increased when selecting HPLC-FLD fingerprinting segments richer in discriminant features, which were deduced from PLS-DA loading plots. In this case, excellent separation was observed and 100% classification rates for both PLS-DA calibrations and predictions were obtained (all samples were correctly classified within their corresponding groups).

CONCLUSION: HPLC-FLD fingerprinting segments resulted to be suitable chemical descriptors to discriminate the origin (country of production), variety (Arabica and Robusta) and roasting degree of coffee. Therefore, HPLC-FLD fingerprinting can be proposed as a feasible, simple and cheap methodology to address coffee authentication, especially for developing coffee production countries.

KEYWORDS: HPLC-FLD fingerprinting; Coffee; Food authentication; Principal component analysis (PCA); Partial least square regression-discriminant analysis (PLS-DA)

Introduction

Coffee is today one of the most popular beverages in the world. More than one billion cups are consumed every day, with an annual consumption per capita over 5 kg, on average, in Europe. Unfortunately, it is one of the most easily adulterable products. Because of the food chain complexity and all the factors involved since food production until its consumption, the adulteration and manipulation of some foodstuffs and beverages are increasing in the last years. Besides, the quality of natural products such as coffee is an issue of great interest in our society, which is increasingly interested in attributes such as the coffee origin and its specific variety. Therefore, it is very important to ensure coffee quality control with the aim of protecting consumers from fraudulent practices as well as to ensure that coffee beverages are fit for human consumption.[1–5]

Coffee is an infusion that comes from roasted and ground coffee beans with a characteristic taste and aroma. The coffee plant belongs to *Coffea* genus from the Rubiaceae family, with more than 70 species, being *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) the most consumed varieties. In general, pure Arabica coffee beans are viewed as superior to Robusta in terms of quality for their higher sensorial properties. In addition, Arabica coffee variety is preferred by consumers because of its less bitter taste than Robusta counterpart. For all these reasons, Arabica coffee usually has a higher price in the international market.[6,7] Coffee is known as a stimulant, a property attributed mainly to caffeine, which has been considered for years to be responsible for the beneficial effects of coffee. However, it is currently known that coffee contains a high number of bioactive substances such as phenolic acids and polyphenols, providing important beneficial health effects such as high antioxidant activity. Besides, its chemical composition and its taste depend also on the coffee variety, the roasting degree, the fermentation degree, the climatic conditions, the storage, and the drying method, among other factors.[6–8]

A large range of analytical methodologies has been developed during the last years to address the characterization and authentication of coffee, most of them in combination with chemometric methods due to the large amount of obtained data from the analyzed samples.[9] Targeted and non-targeted methods are employed and described in the bibliography. For instance, Hatumura et al.[10] proposed a chemometric analysis of ^{1}H NMR fingerprints of *Coffea arabica* cultivated under different conditions. Marek et al.[11] described the use of an electronic nose for the detection of volatile compound profiles in *Arabica coffee* from different regions of origin. In terms of separation techniques to

identify and quantify coffee compounds, liquid chromatography (LC) coupled to ultraviolet (UV)[12,13] and fluorescence (FLD)[14,15] detection has been used. Other separation techniques usually employed are capillary electrophoresis (CE)[16,17] and gas chromatography (GC)[18], both coupled to mass spectrometry (MS). For example, Pérez-Míguez et al.[16] reported the use of capillary electrophoresis-mass spectrometry (CE-MS) to see the differences in coffee fingerprints based on their different roasting degree, as well as for the identification of metabolites. Ongo et al.[18] described the use of gas chromatography coupled to mass spectrometry (GC-MS) for the identification of some volatile metabolites and for the evaluation of the differences between Arabica and Robusta fingerprints.

In a previous work, a high-performance liquid chromatography with ultraviolet detection (HPLC-UV) fingerprinting method was proposed for the characterization and classification of coffee according to different quality attributes such as the region of origin, the variety (Arabica/Robusta) and the roasting degree.[13] Chromatographic separation was addressed by using C18 reversed-phase column under gradient conditions employing 0.1% formic acid aqueous solution and methanol as mobile phase with a total analysis time of 40 min. Although an acceptable discrimination among samples was achieved, overlapping of many groups of samples was still observed by chemometrics. As a result, not all the samples were correctly classified by partial least square regression-discriminant analysis (PLS-DA).

The aim of this work is to evaluate if high-performance liquid chromatography with fluorescence detection (HPLC-FLD) fingerprinting can provide better sample chemical descriptors than those from HPLC-UV for coffee characterization and classification. For that purpose, over one hundred sixty-eight commercially available coffee samples, divided into two sets, were analyzed with proposed HPLC-FLD method. HPLC-FLD fingerprints were then employed as a source of chemical information to address the characterization, classification and authentication of the analyzed coffees according to their origin (country of production), variety (Arabica vs. Robusta), and roasting degree by multivariate chemometric methods such as principal component analysis (PCA) and PLS-DA.

4

Materials and Methods

Samples and sample treatment

The proposed HPLC-FLD fingerprinting method was applied to the analysis of 186 commercially available coffee samples, grouped in two different sets (commercial name, number of samples, and coffee production region, variety and roasting degree are described in Table 1). The first set of samples included a total of 120 Nespresso® coffee samples purchased from several supermarkets in Barcelona (Spain). There were six types of samples differing in the coffee variety (Arabica and Arabica-Robusta mixtures), the region of origin (Colombia, Ethiopia, India, Nicaragua, Indonesia, and a group of samples with an unknown origin), and the roasting degree (increasing from 1 to 5). The second set contained a total of 66 coffee samples obtained from commercial brands in Cambodia and Vietnam supermarkets. Samples were divided according their coffee variety and their production region into five groups. The roasting degree was unknown for this second set of samples.

Samples were analyzed without any sample treatment apart from brewing the coffee with mineral water. For the first set of samples, coffees were directly brewed in an espresso machine (Nespresso), always using the same time of brewing to achieve the same final volume. For the second set of samples, coffees were brewed in an Italian coffee maker using 400 mL of mineral water measured with a test tube and using the same amount of coffee after grinding the coffee beans with a grinder (when it was necessary). Then, the coffee was prepared with the help of a Bunsen burner to carry out the coffee lixiviation. Finally, all brewed coffees (sets 1 and 2) were filtered with 0.45 µm nylon filters into amber glass vials of 2 mL which were stored at -4 ℃ until their analysis.

Furthermore, in order to evaluate the repeatability of the proposed fingerprinting method and the robustness of the chemometric results, a quality control (QC) solution was prepared for each set of samples by mixing 50 µL of each corresponding sample extracts.

To prevent signal tendencies attributed to the sample sequence analysis, all coffee samples were analyzed randomly with the proposed HPLC-FLD method. Besides, blanks of methanol and QCs were injected every 10 randomly analyzed samples.

Chemicals

Mobile phase was composed of methanol (HPLC grade) purchased from PanReac AppliChem (Barcelona, Spain), Milli-Q water and formic acid (≥98%) obtained from

Sigma-Aldrich (St. Louis, MO, USA). Mobile phase water was purified filtering it through a 0.22 µm nylon membrane integrated into the Milli-Q system. Mineral water obtained from Eroski supermarket (Barcelona, Spain) was employed for brewing the coffees.

Experimental

A HPLC instrument from Agilent HPLC 1100 Series (Waldbronn, Germany), equipped with a binary pump (G1312A), an automatic sample injector (WPALS G1367A), a fluorescence detector (G1321A) and a PC with the Agilent Chemstation software, was used to obtain the HPLC-FLD chromatographic fingerprints. Chromatographic separation was performed using a Kinetex® C18 reversed-phase (100 × 4.6 mm i.d., 2.6 µm particle size) column obtained from Phenomenex (Torrance, California, USA) under gradient elution conditions employing 0.1% formic acid aqueous solution (solvent A) and methanol (solvent B) as mobile phase components. The elution program employed is indicated in Table 2. The injection volume was 5 µL and the mobile phase flow rate was 0.4 mL/min. Chromatographic fingerprints were recorded at an excitation wavelength of 310 nm and an emission wavelength of 410 nm.

Data analysis

The obtained HPLC-FLD chromatograms were exported creating different fingerprinting data matrices to be analyzed by principal component analysis (PCA) and partial least squares regression-discriminant analysis (PLS-DA) chemometric methods. Stand Alone Chemometrics Software (SOLO) from Eigenvector Research was used for calculations.[19] A detailed description of the theoretical background of chemometric methods is given elsewhere.[20] In all the chemometric methods, data was autoscaled to provide the same weight to each variable by suppressing differences in their magnitude and amplitude scales. PLS-DA models were stablished and validated using 70% of the samples (randomly selected) as the calibration set and the remaining 30% of the samples as the prediction set. The optimal number of LVs in the PLS-DA models was established by considering the first significant minimum point of the cross-validation (CV) error from a Venetian blind approach.

Results and discussion

Non-targeted HPLC-FLD fingerprints of coffees

The authentication of the origin (country of production), the variety and the roasting degree of commercially available coffee samples is gaining relevance. In a previous work,[13] a non-targeted HPLC-UV fingerprinting method for the characterization and authentication of coffee samples was developed. It was concluded that HPLC-UV fingerprints were good chemical descriptors for classifying certain groups of coffee samples, although they could not classify all the analyzed coffees (classification rates higher than 89.3% in some evaluated pairs of samples). Recent works have demonstrated that the use of more selective sample chemical descriptors, such as the ones obtained by FLD, may improve the classification capabilities of the proposed authentication strategies.[21] Therefore, the main objective of this work was to evaluate the ability of HPLC-FLD fingerprints as an alternative to obtain better chemical descriptors to address the classification and authentication of commercially available coffee samples. Although this approach relies on non-targeted analysis, hydroxycinnamic acids are one of the most relevant families of coffee components, including caffeic acid and their derivatives. The method is mainly focused on the detection of this family so FLD is adjusted to provide the best sensitivity and selectivity for caffeic acid species (excitation at 310 nm / emission at 410 nm).

As an example, the chromatograms of the six groups of coffees belonging to the first set of samples are depicted in Figure 1. Representative HPLC-FLD fingerprints of each group of coffee samples belonging to the second set of samples are provided in Figure S1 (Supporting information). As can be seen, chromatographic fingerprints show important differences among samples regarding both the number of peak signals detected, as well as the peak intensities. For example, the depicted coffee samples from Colombia and Ethiopia (Figures 1a and b, respectively), both of the Arabica variety, displayed much higher intensities in comparison to the other samples due to the presence of a high intensity signal at a retention time of ~17 min, which is attributed to the presence of caffeic acid.[13] Regarding the second set of samples (Figure S1), signal intensities are more comparable among them, although they tend to be less intense than the ones obtained for the coffees of the first set of samples. Besides, signal intensities among the different detected compounds are more similar within the same chromatogram, although important

differences in the number of peaks and signal intensities are again observed among the different analyzed samples.

The obtained HPLC-FLD fingerprints also seem to be richer in comparison to HPLC-UV fingerprints regarding the number of extracted compounds that can be detected.[13] Moreover, these fingerprints are also very reproducible within the coffee samples of the same type. Therefore, they will be evaluated as potential chemical descriptors for the characterization and classification of commercially available coffee samples by PCA and PLS-DA as the chemometric methods.

Non-supervised sample exploration by PCA

As a first approach, HPLC-FLD fingerprints of the analyzed coffee samples, together with the analyzed QCs, were evaluated by a non-supervised exploratory method such as PCA for the discrimination of samples regarding their origin (country of production). This first study is relevant to determine the behavior of QCs to establish the reproducibility of the proposed method but, more important, the robustness of the chemometric results. Therefore, X-data matrices for the two coffee sets of samples (including the corresponding QC for each set) were created. Matrices consisted of the fluorescence intensity signals registered as a function of retention time, providing matrix dimensions of 133 x 5554 and 72 x 5554 for the first and second set of samples, respectively. Autoscaling data pretreatment was employed to give similar weight to all the fingerprinting variables. As examples, the best 2D PCA plots of scores for both coffee sets of samples regarding the coffee origin are shown in Figure 2. The PCA models were able to retain a 89.73% of the total variance (60.89% from PC1, 13.15% from PC2, 7.04% from PC3, and 2.46% from PC4) for the first set of samples (Figure 2a), and 89.73% of the total variance (67.07% from PC1, 13.15% from PC2, 7.04% from PC3, and 2.46% from PC4) for the second set of samples (Figure 2b), with 4 PCs. The influence plot of Hotelling T2 versus Q residuals evidenced the absence of outlier samples. Good reproducibility of the proposed HPLC-FLD methodology and robustness of the chemometric results was assessed as can be observed with the QCs samples that appeared perfectly clustered in the center area of PCA plots. Therefore, the obtained chemometric results can be employed to evaluate the ability of HPLC-FLD fingerprints to address coffee sample classification and authentication issues. As can be seen in Figure 2a for the first set of samples, certain sample discrimination is observed when plotting PC1 vs. PC4, although perfect separation among the different groups of samples is not accomplished.

8

Coffees produced in India and those of unknown origin appeared completely mixed at the right of the plot, being differentiated for the other groups of coffees by presenting positive PC1 scores. Indonesian coffees are clustered at the center of the plot, together with the QCs, and the other three groups of coffees are clustered at the left are of the plot being differentiated by both PC1 and PC4, although some overlapping between groups is observed. When studying the second set of coffee samples (Figure 2b), no discrimination at all was observed between Vietnamese and Cambodian coffees with both groups completely overlapped. However, Cambodian coffees are clustered in small group in comparison to Vietnamese coffees, which is expected taking into consideration the variability of other sample attribute within the Vietnamese samples (see Table 1).

PCA results suggested that HPLC-FLD fingerprints contained some features with potential ability to distinguish among coffee classes. For that reason, data was also submitted to a supervised classification chemometric method like PLS-DA.

Supervised sample classification by PLS-DA

Supervised PLS-DA was employed for the classification of the analyzed commercially available coffee samples according of three attributes: the coffee region of origin (country of production), the coffee variety (Arabia and Robusta), and the coffee roasting degree (from low roasted coffees, 1/5, to highly roasted coffees, 5/5). For that purpose, X-data matrix consisted of the fluorescence intensity signals registered as a function of retention time for the coffee samples and the corresponding QCs of each sample set, while the Y-data matrix defined each sample class (origin, variety or roasting degree, depending on the case). Figure 3 shows the best 2D or 3D PLS-DA score plots obtained for each set of samples for the three attributes under study (origin, variety and roasting degree) when using HPLC-FLD fingerprints as sample chemical descriptors. The total number of LVs employed to build each PLS-DA model, calculated by considering the first significant minimum point of the CV error from a Venetian blind approach, are also indicated in the Figure caption. Again, QC samples are closely grouped in all the depicted score plots, and close to the center area of the plots, thus verifying the robustness of the obtained chemometric results. Besides, sample classification according to the different studied attributes improved by PLS-DA in comparison with those obtained by PCA, as expected. It can be clearly observed that samples tend to be well clustered according to the production region. For example, regarding the first set of samples (Figure 3a), similar distribution among the sample groups than the one obtained by PCA (Figure

9

2a) is observed, although better discrimination of Nicaraguan coffees respect to those from Colombian, Ethiopian and Indonesian was attained. It should be pointed out that the sample distribution obtained when using HPLC-FLD fingerprints differs completely from HPLC-UV results previously reported,[13] enhancing the differences on the obtained fingerprints depending on the detection system employed. However, regarding the sample discrimination for this first set of samples as a function of the sample origin, results are not showing a clear improvement when comparing FLD and UV detection, and similar sample discrimination was observed.

Sample discrimination depending on the production region clearly improved when studying Vietnamese and Cambodian coffees by PLS-DA (Figure 3b). In that case, separation of Cambodian respect to Vietnamese coffees was also obtained. Besides, sample discrimination clearly improved respect to the one obtained when using HPLC-UV fingerprints as sample chemical descriptors (Figure 2b).[13] Therefore, in that specific case, HPLC-FLD fingerprints seem to be better chemical descriptors than HPLC-UV fingerprints to address the classification and authentication of these two groups of coffee samples.

The evaluated fingerprints also showed their performance for the classification of the studied samples according to the coffee variety. Regarding the first set of coffees, samples are perfectly clustered in two groups (Figure 3c), with coffees of Arabica variety showing positive values of LV1, and the coffees blended with mixtures of Arabica and Robusta varieties exhibiting negative values of LV1. This discrimination was better than the one previously observed by employing HPLC-UV fingerprints with more separation between the two groups of samples.[13] Regarding the samples coming from Vietnam and Cambodia, only those where the variety was clearly labeled (Arabica and Robusta) were subjected to PLS-DA. In that case, separation of the coffees regarding their variety was not fully accomplished, as can be observed in Figure 3d, although certain discrimination between the two groups of samples is obtained. In addition, results were similar to the ones obtained when using HPLC-UV fingerprints.[13]

Finally, the classification of coffees belonging to the first set of samples was also evaluated regarding their different roasting degree, and the 3D PLS-DA results are shown in Figure 3e (this sample attribute was unavailable for Vietnamese and Cambodian coffees). In general, similar sample distribution was observed either using HPLC-FLD or HPLC-UV fingerprints as chemical descriptors,[13] with the samples following a circle through the plot showing that all LVs were related to the coffee roasting degree, although

several sample groups tend to be more clustered when employing HPLC-FLD fingerprints. Despite that certain discrimination depending on the sample roasting degree was obtained, important overlapping of the sample groups is still present.

In general, it was observed that the proposed HPLC-FLD fingerprints resulted to be acceptable chemical descriptors to address the classification and authentication of commercially available coffee samples, showing results at least similar, but in some cases better, than the ones previously obtained by employing HPLC-UV fingerprints. Nevertheless, complete sample discrimination was not accomplished in some cases and data need to be refined.

Supervised sample classification by PLS-DA using discriminant HPLC-FLD fingerprinting segments

To improve the sample classification, PLS-DA loadings plots were evaluated in order to select the most discriminant ranges from the whole fingerprints. As an example, Figure S2 (supporting information) shows the PLS-DA loading plot for the classification of coffee set of samples 1 regarding the coffee production origin. As can be seen, PLS-DA loading plots allow to detect those fingerprinting segments that are more discriminant as the ones containing the group of variables more separated from the center area of the loading plot. Studies with various fingerprinting segments and combinations were evaluated. In the case of the coffee set of samples 1, the best results were obtained when three chromatographic windows (represented in Figure S3a, supporting information) were selected: from 2-4.5 min, 8-27 min and 36.5-38 min segments. For the Vietnamese and Cambodian coffee samples, the best results were achieved when two HPLC-FLD fingerprinting segments (Figure S3b, supporting information) were used: from 2-5 min and 8-27 min segments. X-data matrices were built comprising the fluorescence intensity signals of each time range for the coffee samples and the corresponding QCs of each sample set, while the Y-data matrices defined each sample class (origin, variety or roasting degree, depending on the case). The best 2D or 3D PLS-DA score plots obtained for each set of samples for the three sample attributes under study (origin, variety and roasting degree) are depicted in Figure 4. The total number of LVs employed to build each PLS-DA model by Venetian blind CV approach are also indicated in the Figure caption. The use of the selected data ranges clearly improved the sample classification obtained in comparison to the use of the complete HPLC-FLD fingerprint. This is clearly enhanced in the classification set of samples 1 regarding the coffee production region.

11

Perfect separation and discrimination between all the sample groups was shown in the 3D PLS-DA score plot (Figure 4a), and only Indonesian and Nicaraguan coffee samples are clustered close, although well separated. The separation improvement is also very representative when studying this same group of samples in relation to their roasting degree (Figure 4e), with a huge decrease in the sample overlapping in comparison to the use of the whole fingerprint as chemical descriptor (Figure 3e). In all the other cases under study, similar or slightly better results were observed when working under selected time windows.

The results obtained in this study clearly demonstrate that selected fingerprinting segments can be proposed as good sample chemical descriptors to address the characterization and classification coffee samples according to the coffee region of origin (production country), the coffee variety (Arabica vs. Robusta), and the coffee roasting degree.

Coffee classification by PLS-DA

With the aim of demonstrating the applicability of the proposed methodology based on selected HPLC-FLD fingerprinting segments as chemical descriptors of the analyzed samples, the classification rate was studied for some paired PLS-DA models: (i) Indonesian vs. Nicaraguan coffees, (ii) Colombian vs. Nicaraguan coffees, (iii) Indonesian vs. Indian coffees, (iv) Colombian vs. Ethiopian coffees, (v) Vietnamese vs. Cambodian coffees, and (vi) Vietnamese Arabica vs. Vietnamese Robusta coffees. To validate the PLS-DA models and their prediction rates, each paired system was evaluated by using 70% of randomly selected samples as calibration set, while the remaining 30% of the samples were used as an "unknown" group of samples for validation/prediction purposes. Figure 5 depicts the classification plots for the six paired PLS-DA models evaluated. The number of LVs employed to build each classificatory model, as well as the classification rates for both model and prediction, are also indicated in the figure.

As can be seen, in all cases all the samples were correctly classified with 100% classification rates for both calibration and prediction steps. This good classification rate performance is even achieved with groups of samples that appeared clustered remarkably close in the PLS-DA classification study such as the case of coffees produced in India and Indonesia (Figure 4a). Besides, these results are clearly better than the ones previously obtained by employing HPLC-UV fingerprints as chemical descriptors[13] for the classification between Vietnamese Arabica and Robusta samples. In the case of

12

HPLC-UV fingerprints, the classification rate for prediction was only of 91.7% (3 out of 15 samples were not correctly classified, one Arabica and two Robusta samples), while with the proposed approach based on selected time windows, all the samples are perfectly assigned to its corresponding group.

Conclusions

In the present work, HPLC-FLD fingerprints, obtained by C18 reversed-phase chromatography from the direct injection of brewed coffees without any sample treatment other than filtration (reducing sample manipulation), provided excellent chemical descriptors to authenticate the origin (country of production), the variety (Arabica vs. Robusta), and the roasting degree of coffees by chemometrics.

In a first approach, whole HPLC-FLD fingerprints provided acceptable sample classification by PLS-DA, although some overlapping occurred between several groups of samples, especially among production region for the coffee set of samples 1, among Vietnamese Arabica and Robusta samples, and among the coffee roasting degrees. Some groups were even completely overlapped, such as the Indian coffees with those of unknown origin. The evaluation of the PLS-DA loading plots allowed the selection of more specific fingerprinting segments with enhanced discrimination capabilities. The discrimination and classification of samples by PLS-DA considerably improved by employing HPLC-FLD fingerprinting segments as sample chemical descriptors, especially for the coffees set of samples 1 (with all the sample groups completely separated) and for Vietnamese Arabica and Robusta samples. Besides, PLS-DA provided excellent classification rates, with all the samples correctly classified within its corresponding group.

Therefore, the proposed HPLC-FLD fingerprinting methodology by selecting discriminant chromatographic segments is a simple and relatively cheap methodology to address coffee authentication regarding origin, variety and roasting degree.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

References

1   Esquivel P and Jiménez VM. Functional properties of coffee and coffee by-products. *Food Res Int* Elsevier Ltd; **46**:488–495 (2012).

2   Campmajó G, Núñez N, and Núñez O. The Role of Liquid Chromatography-Mass Spectrometry in Food Integrity and Authenticity, in: Kamble, G.S. (ed.) Mass Spectrometry - Future Perceptions and Applications. London: IntechOpen; p. 3–20 2019.

3   Moore JC, Spink J, and Lipp M. Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. *J Food Sci* **77**:R118–R126 (2012).

4   Gonzalvez A, Armenta S, and Guardia M De. Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin. *Trends Anal Chem* Elsevier Ltd; **28**:1295–1311 (2009).

5   Kamiloglu S. Authenticity and traceability in beverages. *Food Chem*; **277**:12–24 (2019).

6   Naranjo M, Vélez ILT, Benjamín II, and Iii AR. Actividad antioxidante de café colombiano de diferentes calidades Antioxidant activity of different grades of Colombian coffee. *Rev Cubana Plant Med*; **16**:164–173 (2011).

7   Crozier A, Ahihara H, and Tomás-Barbéran F (eds.). Teas, Cocoa and Coffee. Plant Secondary Metabolites and Health. Oxford, UK: Wiley-Blackwell; 2012.

8   Herawati D, Giriwono PE, Nur F, Dewi A, Kashiwagi T, and Andarwulan N. Critical roasting level determines bioactive content and antioxidant activity of Robusta coffee beans. *Food Sci Biotechnol* Springer Singapore; **28**:7–14 (2019).

9   Toci AT, Farah A, Pezza HR, and Pezza L. Coffee Adulteration: More than Two Decades of Research. *Crit Rev Anal Chem* Taylor & Francis; **46**:83–92 (2016).

10   Hatumura PH, Oliveira GS de, Marcheafave GG, Rakocevic M, Bruns RE, Scarminio IS, and Terrile AE. Chemometric Analysis of 1H NMR Fingerprints of Coffea arabica Green Bean Extracts Cultivated under Different Planting Densities. *Food Anal Methods* Food Analytical Methods; **11**:1906–1914 (2018).

11   Marek G, Dobrzański B, Oniszczuk T, Combrzyński M, Ćwikła D, and Rusinek

R. Detection and Differentiation of Volatile Compound Profiles in Roasted Coffee Arabica Beans from Different Countries Using an Electronic Nose and GC-MS. *Sensors (Basel)* **20** (2020).

12 Restuccia D, Spizzirri UG, Parisi OI, Cirillo G, and Picci N. Brewing effect on levels of biogenic amines in different coffee samples as determined by LC-UV. *Food Chem* Elsevier Ltd; **175**:143–150 (2015).

13 Núñez N, Collado X, Martínez C, Saurina J, and Núñez O. Authentication of the Origin, Variety and Roasting Degree of Coffee Samples by Non-Targeted HPLC-UV Fingerprinting and Chemometrics. Application to the Detection and Quantitation of Adulterated Coffee Samples. *Foods* **9**:378 (2020).

14 Shi Y, Wu H, Wang C, Guo X, Du J, and Du L. Determination of polycyclic aromatic hydrocarbons in coffee and tea samples by magnetic solid-phase extraction coupled with HPLC-FLD. *Food Chem* Elsevier Ltd; **199**:75–80 (2016).

15 Moez E, Noel D, Brice S, Benjamin G, Pascaline A, and Didier M. Aptamer assisted ultrafiltration cleanup with high performance liquid chromatography-fluorescence detector for the determination of OTA in green coffee. *Food Chem* Elsevier; **310**:125851 (2020).

16 Pérez-Míguez R, Sánchez-López E, Plaza M, Marina ML, and Castro-Puyana M. Capillary electrophoresis-mass spectrometry metabolic fingerprinting of green and roasted coffee. *J Chromatogr A* **1605** (2019).

17 Daniel D, Lopes FS, Santos VB dos, and Lago CL do. Detection of coffee adulteration with soybean and corn by capillary electrophoresis-tandem mass spectrometry. *Food Chem* Elsevier; **243**:305–310 (2018).

18 Ongo EA, Montevecchi G, Antonelli A, Sberveglieri V, and Sevilla F. Metabolomics fingerprint of Philippine coffee by SPME-GC-MS for geographical and varietal classification. *Food Res Int* **134** (2020).

19 Eigenvector Research Incorporated. Powerful Resources for Intelligent Data Analysis. Available online: http://www.eigenvector.com/software/solo.htm (accessed on 15 January 2019). .

20 Massart DL, Vandeginste BGM, Buydens LMC, Jong S de, Lewi PJ, and Smeyers-Verbeke J. Handbook of chemometrics and qualimetrics. Amsterdam, The Netherlands: Elsevier; 1997.

21 Campmajó G, Saez-Vigo R, Saurina J, and Núñez O. High-performance liquid

chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations. *Food Control* **114**:107265 (2020).

Figure captions

Figure 1. Non-targeted HPLC-FLD fingerprints (λ_{exc} 310 nm; λ_{em} 410 nm) of the six groups of commercially available coffee samples belonging to the first set of samples. (a) Arabica coffee from Colombia, (b) Arabica coffee from Ethiopia, (c) Arabica-Robusta coffee from India, (d) Arabica coffee Nicaragua, (e) Arabica coffee from Indonesia, and (f) Arabica-Robusta coffee of Unknown origin. r.f.u: relative fluorescence units.

Figure 2. PCA scatter plot of scores for the classification of coffee samples regarding the origin (country of production). (a) PC1 vs. PC4 score plot for sample set 1 and (b) PC1 vs. PC2 for sample set 2.

Figure 3. PLS-DA scatter plot of scores when using HPLC-FLD fingerprints as chemical descriptors for the classification of the analyzed coffee samples. (a) 2D PLS-DA score plot (LV1 vs. LV2), build with 6 LVs, for coffee set of samples 1 regarding the coffee origin; (b) 2D PLS-DA score plot (LV1 vs. LV2), build with 3 LVs, for coffee set of samples 2 regarding the coffee origin; (c) 2D PLS-DA score plot (LV1 vs. LV2), build with 3 LVs, for coffee set of samples 1 regarding the coffee variety; (d) 3D PLS-DA score plot (LV1 vs. LV2 vs. LV3), build with 5 LVs, for coffee set of samples 2 regarding the coffee variety; and (e) 3D PLS-DA score plot (LV1 vs. LV2 vs. LV3), build with 8 LVs, for coffee set of samples 1 regarding the coffee roasting degree.

Figure 4. PLS-DA scatter plot of scores when using HPLC-FLD fingerprinting segments as chemical descriptors for the classification of the analyzed coffee samples. (a) 3D PLS-DA score plot (LV1 vs. LV2 vs. LV3), build with 5 LVs, for coffee set of samples 1 regarding the coffee origin; (b) 2D PLS-DA score plot (LV1 vs. LV2), build with 3 LVs, for coffee set of samples 2 regarding the coffee origin; (c) 2D PLS-DA score plot (LV1 vs. LV2), build with 4 LVs, for coffee set of samples 1 regarding the coffee variety; (d) 3D PLS-DA score plot (LV1 vs. LV2 vs. LV3), build with 3 LVs, for coffee set of samples 2 regarding the coffee variety; and (e) 3D PLS-DA score plot (LV1 vs. LV2 vs. LV3), build with 6 LVs, for coffee set of samples 1 regarding the coffee roasting degree.

Figure 5. Samples vs. Y predicted 1 Scores plot for (a) Indonesian vs. Nicaraguan coffees, (b) Colombian vs. Nicaraguan coffees, (c) Indonesian vs. Indian coffees, (d) Colombian vs. Ethiopia coffees, (e) Vietnamese vs. Cambodian coffees, and (f) Vietnamese Arabica vs. Vietnamese Robusta coffees. Filled and empty symbols correspond to calibration and validation sets, respectively. The number of LVs employed to generate each classificatory model and sample classification rates are also indicated.

Table 1. Description of the analyzed commercially available coffee samples.

Commercial Name	Number of Samples	Coffee variety	Origin Region	Roasting degree
Set of samples 1				
Master Origin Colombia	20	Arabica	Colombia	3/5
Master Origin Ethiopia	20	Arabica	Ethiopia	2/5
Master Origin India	20	Arabica-Robusta Mixture	India	5/5
Master Origin Nicaragua	20	Arabica	Nicaragua	2/5
Master Origin Indonesia	20	Arabica	Indonesia	4/5
Paris Black	20	Arabica-Robusta Mixture	Unknown origin	4/5
Set of samples 2				
-	20	Arabica	Vietnam	Unknown
-	20	Robusta	Vietnam	Unknown
-	10	Arabica-Robusta Mixture	Vietnam	Unknown
-	6	Unknown	Vietnam	Unknown
-	10	Unknown	Cambodia	Unknown

Table 2. Elution gradient used for the chromatographic separation

Time (min)	Solvent B (%)	Elution mode
0-30	3 → 75	Lineal gradient
30-32	75 → 95	Lineal gradient
32-34	95	Isocratic
34-34.2	95 → 3	Back to initial conditions
34.2-40	3	Isocratic

Figure 1

Figure 2
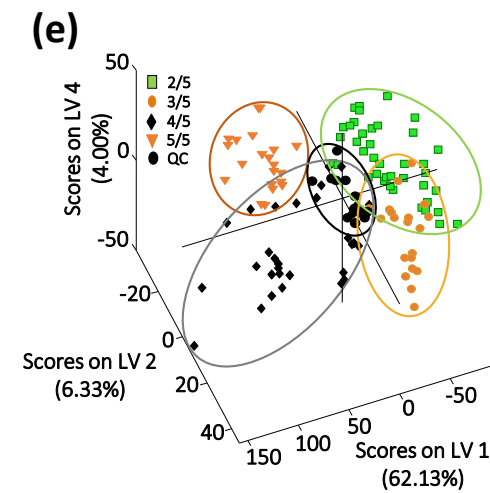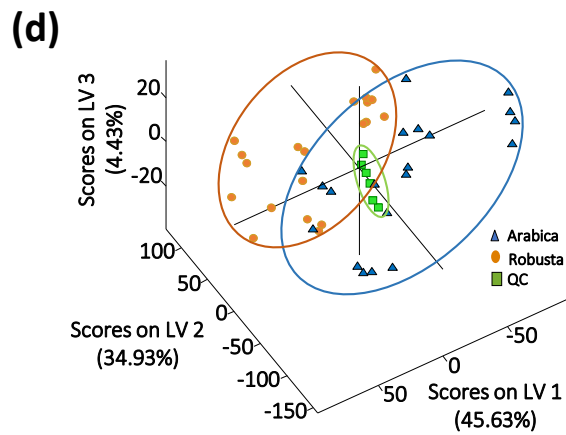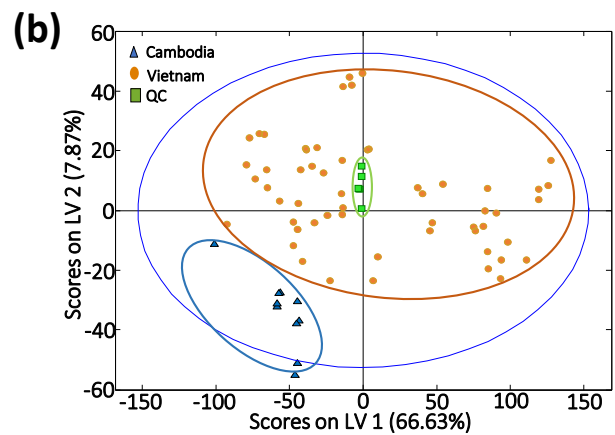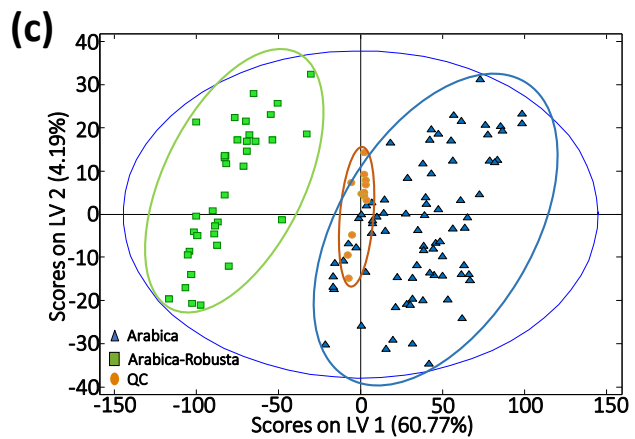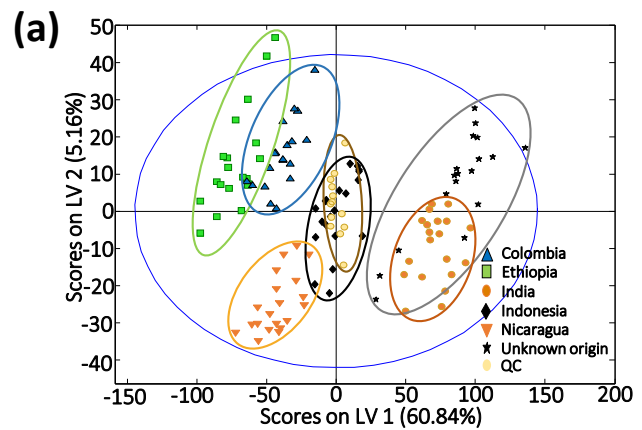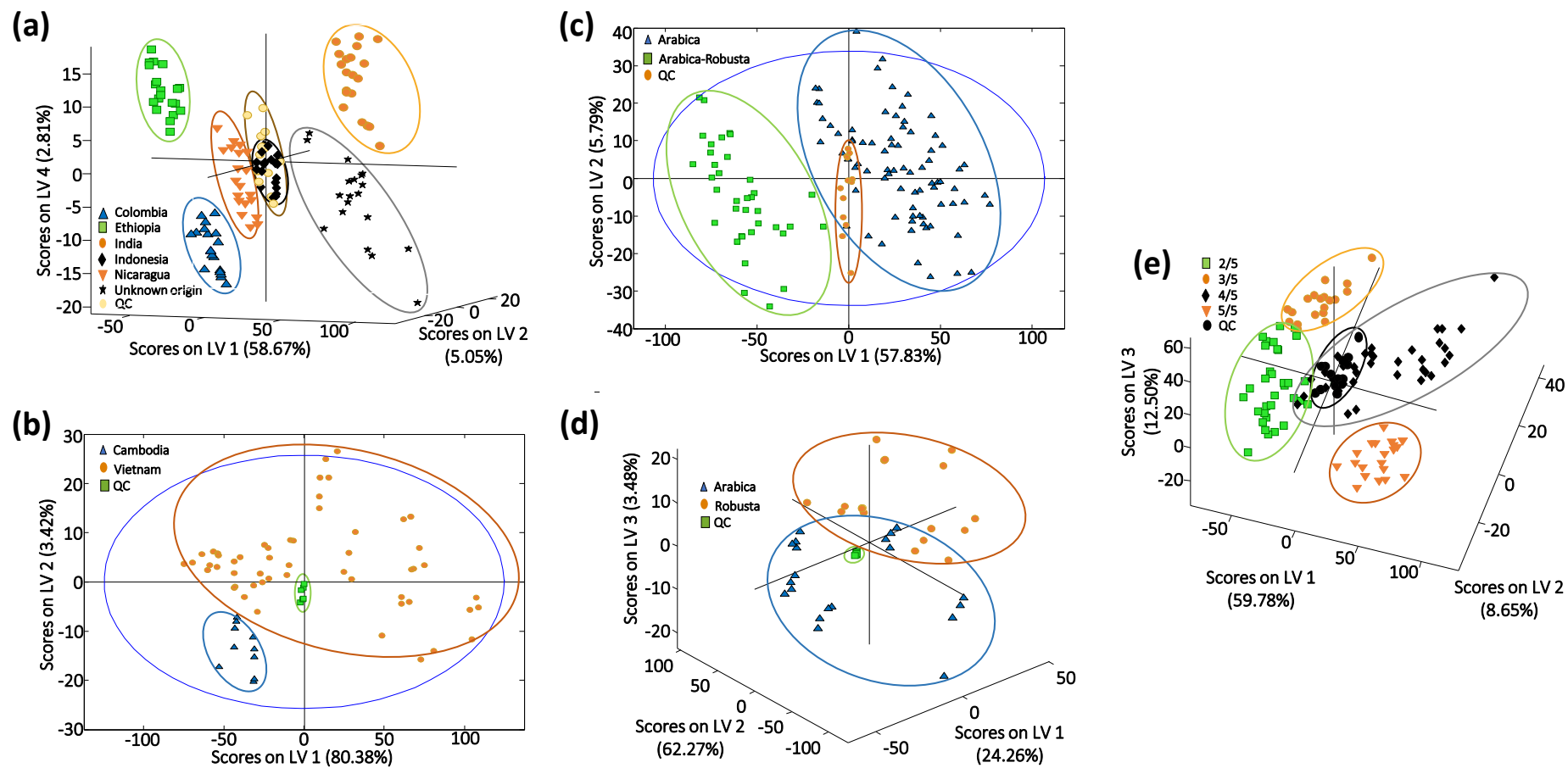
(a)

Scores on PC 4 (3.41%) vs Scores on PC 1 (60.86%)

- ▲ Colombia
- ▢ Ethiopia
- ● India
- ◆ Indonesia
- ▼ Nicaragua
- ★ Unknown origin
- ● QC

(b)

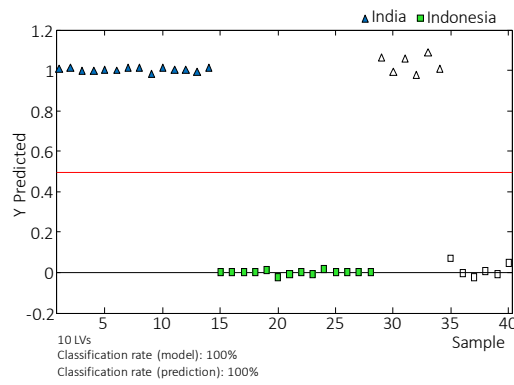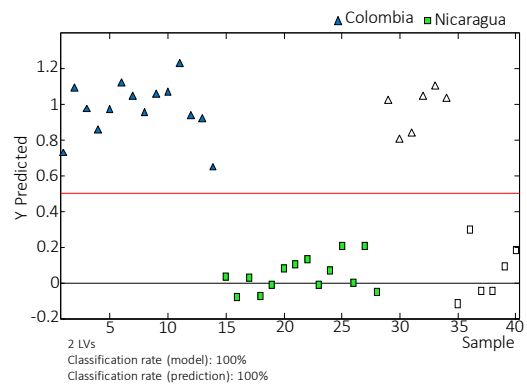Scores on PC 2 (13.15%) vs Scores on PC 1 (67.07%)

- ▲ Cambodia
- ● Vietnam
- ▢ QC

Figure 3

Figure 4
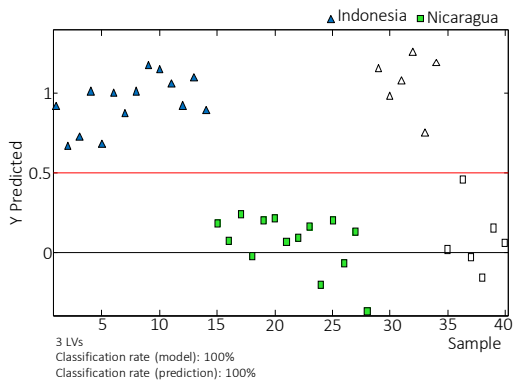
Figure 5
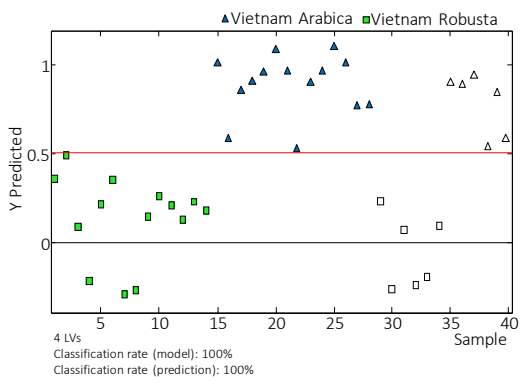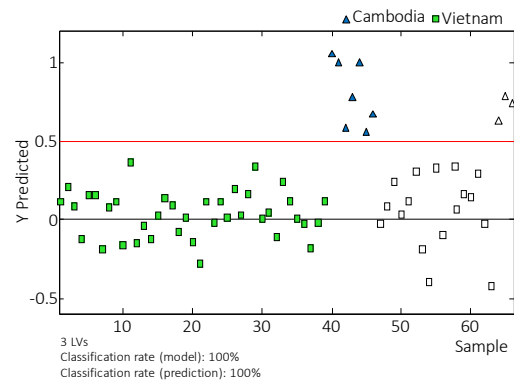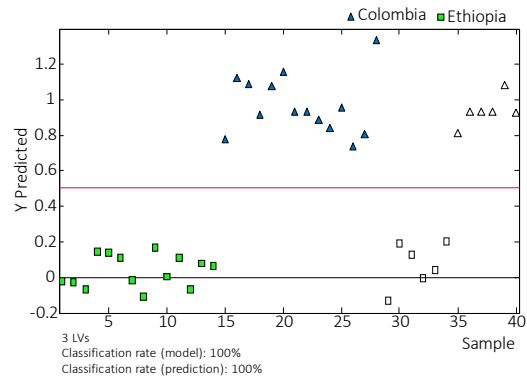
High-Performance Liquid Chromatography with Fluorescence Detection (HPLC-FLD) Fingerprints as Chemical Descriptors to Authenticate the Origin, Variety and Roasting Degree of Coffee by Multivariate Chemometric Methods.
Running title: HPLC-FL Fingerprinting for Coffee Authentication

Nerea Núñez[a,*], Clara Martínez[a], Javier Saurina[a,b] and Oscar Núñez[a,b,c,*]

[a] Department of Chemical Engineering and Analytical Chemistry, University of Barcelona. Martí i Franquès 1-11, E08028 Barcelona, Spain.
[b] Research Institute in Food Nutrition and Food Safety, University of Barcelona. Av. Prat de la Riba 171, Edifici Recerca (Gaudí), E-08921 Santa Coloma de Gramenet, Barcelona, Spain.
[c] Serra Hunter Fellow. Generalitat de Catalunya, E-08007 Barcelona, Spain.
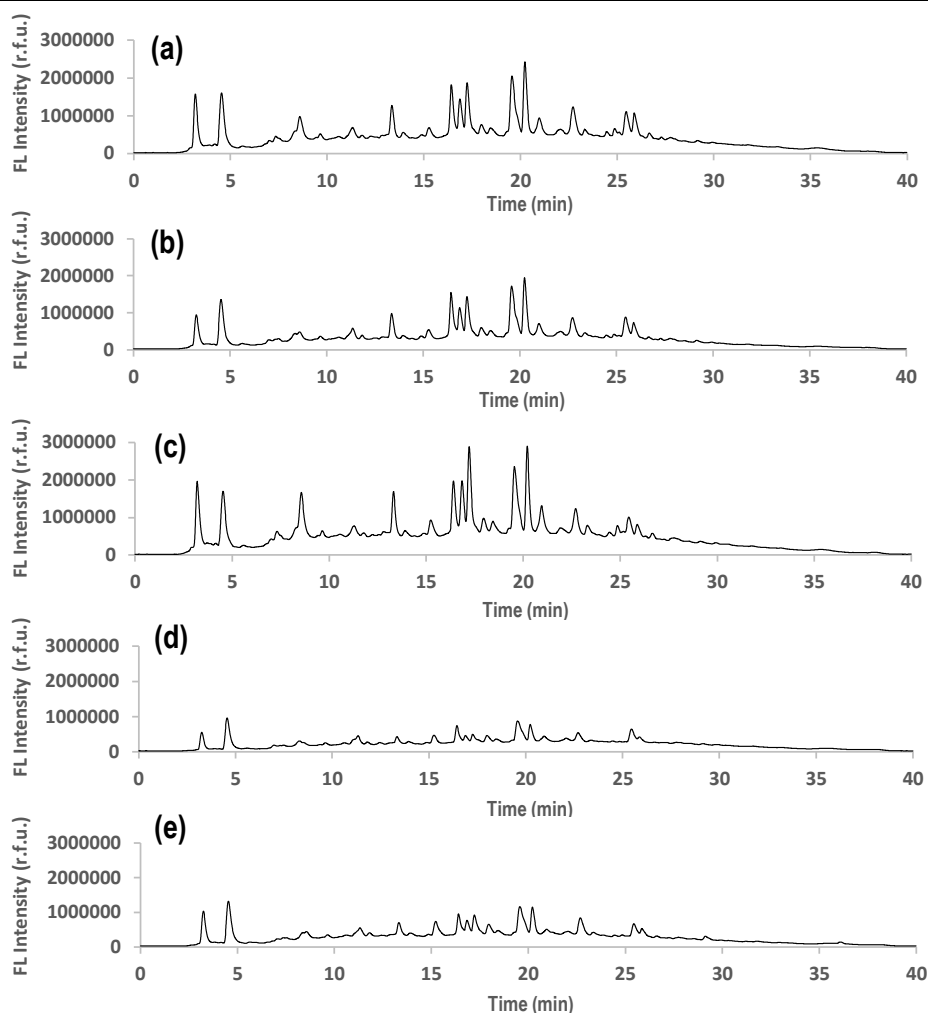[*] Corresponding authors

Figure S1. Non-targeted HPLC-FLD fingerprints (λ_{exc} 310 nm; λ_{em} 440 nm) of the five groups of commercially available coffee samples belonging to the second set of samples. (a) Arabica coffee from Vietnam, (b) Robusta coffee from Vietnam, (c) Mixture variety coffee from Vietnam, (d) Coffee from Vietnam (unknown variety), and (e) Coffee from Cambodia (unknown variety). r.f.u.: relative fluorescence units.
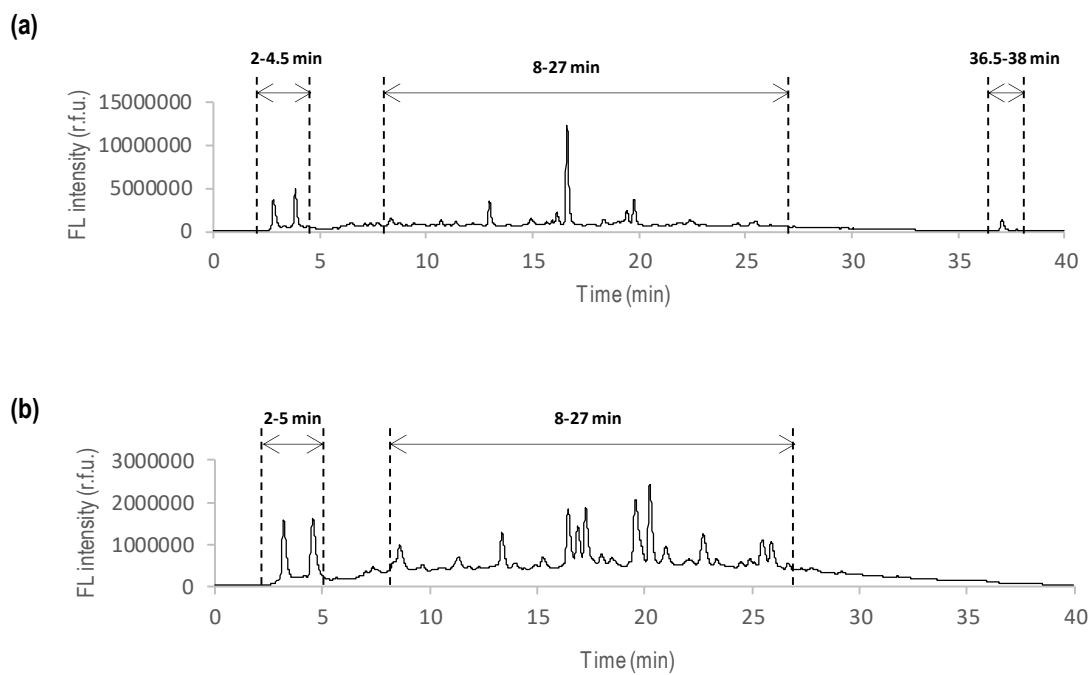
Figure S2. PLS-DA loadings plot (LV1 vs. LV2) when using HPLC-FLD fingerprints as sample chemical descriptors for the classification of coffee set of samples 1 regarding the coffee origin (country of production).

Figure S3. HPLC-FLD coffee chromatograms depicting the selected fingerprinting segments. (a) Colombia Arabica coffee sample, (b) Vietnam Arabica coffee sample. r.f.u.: relative fluorescence units.