Pérez-Míguez, Raquel et al., 2018. A non-targeted metabolomic approach based on reversed-phase liquid chromatography–mass spectrometry to evaluate coffee roasting process. Analytical and bioanalytical chemistry, 410(30), pp.7859–7870.

Available at https://doi.org/10.1007/s00216-018-1405-z

A NON-TARGETED METABOLOMIC APPROACH BASED ON REVERSED-PHASE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY TO EVALUATE COFFEE ROASTING PROCESS

Raquel Pérez-Míguez[1], Elena Sánchez-López[1,2], Merichel Plaza[1,2], María Castro-Puyana[1,2] María Luisa Marina[1,2] *

[1]Departamento de Química Analítica, Química Física e Ingeniería Química, Universidad de Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871 Alcalá de Henares (Madrid), Spain.

[2]Instituto de Investigación Química "Andrés M. del Río" (IQAR), Universidad de Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871 Alcalá de Henares (Madrid), Spain.

Corresponding author: Prof. María Luisa Marina, Departamento de Química Analítica, Química Física e Ingeniería Química. Universidad de Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871, Alcalá de Henares (Madrid), Spain.

E-mail: mluisa.marina@uah.es

Tel./Fax: 34-918854935/34-918854971

Abstract

In this work, a non-targeted metabolomics approach based on the use of reversed-phase liquid chromatography coupled to a high resolution mass spectrometer has been developed to provide the characterization of coffee beans roasted at three different levels (light, medium, and dark). In this way, it was possible to investigate how metabolites change during the roasting process in order to identify those than can be considered as relevant markers. 25 % methanol was selected as extracting solvent since it provided the highest number of molecular features. In addition, the effect of chromatographic and MS parameters was evaluated in order to obtain the most adequate separation and detection conditions. Data were analyzed using both non-supervised and supervised multivariate statistical methods to point out the most significant markers that allow groups discrimination. A total of 24 and 33 compounds in positive and negative ionization modes, respectively, demonstrated to be relevant markers, being most of them from the hydroxycinnamic acids family.

Keywords: Non-targeted metabolomics, liquid chromatography, high resolution mass spectrometry, coffee beans, roasting process.

1. Introduction

Coffee beverage is one of the most consumed drinks in the world. Arabica Coffee, produced by *Coffea arabica* species, is the most consumed and exported coffee variety since it has been considered to have higher sensory properties than other species [1]. Chemical composition of coffee comprises alkaloids, phenolic compounds, carbohydrates, amino acids, proteins and lipids, and some of them are known to present beneficial properties in humans, which makes this beverage a natural source of bioactive compounds. For instance, caffeine, the main alkaloid present in coffee that grants its stimulant nature, has demonstrated, together with chlorogenic acids, to present antioxidant properties [2, 3]. Due to the fact that there are numerous steps in the production of coffee which affect the chemical composition of coffee beans [4], even modifying the organoleptic properties of the drink, it is important to evaluate how coffee chemical composition behaves under the different processes which take place from the growth of its beans until coffee is consumed.

From the different steps carried out during coffee production, the roasting process is notably one of the most important. In this process, several physical and chemical reactions happen with the formation and/or degradation of many compounds responsible for specific organoleptic properties (aroma, flavor and color) that affect the quality of coffee. For instance, total chlorogenic acids composition is reduced during roasting whereas the formation of chlorogenic acid lactones takes place [5, 6], the combination of sugars and amino acids during the Maillard reaction results in the formation of melanoidins [7, 8]. Also, the exposition of carbohydrates to high temperatures affects their composition (they are degraded to lower molecular weight compounds such as mono and oligosaccharides) due to the conversion of part of sugar

in the coffee bean into Maillard reaction and pyrolysis products during roasting, and is responsible for the bean color. Indeed, among the reactions that take place during pyrolysis, pyrazines, molecules known to affect the aroma, are formed by pyrolysis of hydroxy amino acids, while protein pyrolysis together with the degradation of trigonelline are involved in pyridine formation [9]. On the other hand, the composition of other compounds such as lipids or caffeine has demonstrated to be slightly affected by thermal processes [10]. All these changes in the chemical composition of coffee have a great impact in its quality, which makes relevant the search of markers capable of discriminating the changes occurring during the roasting process to ensure the quality and safety of the coffee not only from an industrial point of view but also for the consumers. However, despite the great interest and relevance in studying the changes occurring in coffee during roasting process, the vast majority of reported studies are based on the analysis of a single component or specific class of compounds [11-13]. In this sense, non-targeted metabolomics emerges as a promising tool to obtain the exhaustive and comprehensive analysis of the set of metabolites present in a given system, without prior knowledge on what to look for [14]. Up to date, only few metabolomic works were focused on coffee analysis. On this matter, nuclear magnetic resonance (NMR) [15, 16] and both liquid chromatography mass spectrometry (LC-MS) and gas chromatography mass spectrometry (GC-MS) have been used in the discrimination of different coffee varieties or origins [17-20], being the MS-based techniques also used to discriminate between caffeinated and decaffeinated coffee [21]. However, to the best of our knowledge, the evaluation of the coffee roasting process has only been performed employing a targeted analysis based on the use of ion mobility spectrometry–mass spectrometry (IMS-MS) [22], and using non-targeted metabolomic approaches based on NMR, ambient sonic-spray ionization-mass spectrometry (EASI-

MS) and GC-MS [23- 26]. On the one hand, Wei et al. [23], using a NMR-metabolomics based approach together with a human sensory test, found the chemical substances in roasted coffee bean extracts that could distinguish and predict the different sensations of coffee taste (two degrees of roasted coffee, light and dark were studied). Despite notable advantages of NMR for metabolomic analysis, the major drawback of this technique is its low sensitivity when compared to MS. On the other hand, Santos da Rosa et al. [24] proposed an untargeted and non-volatile approach with EASI coupled to MS to monitor roasting chemical changes in the coffee bean, whereas Sgorbini et al. [25, 26] developed a non-separative headspace solid phase microextraction-mass spectrometry methodology to discriminate volatile compounds among coffee beans submitted to different roasting degrees using GC-MS. Despite LC-MS is the analytical technique most widely used in metabolomics, it has never been applied to study the metabolites changing over the roasting process in coffee. Thus, the aim of this work was to develop, for the first time, a non-targeted metabolomic strategy based on the use of Reversed-Phase Liquid Chromatography (RP-LC) coupled to high resolution MS in order to evaluate changes in the metabolic profiles of coffee samples submitted to different roasting degrees. To accomplish this task, the workflow followed in this study was: (i) the optimization of the sample preparation procedure to obtain the largest number of extracted metabolites, (ii) the optimization of the RP-LC-MS conditions to maximize number of detected peaks, (iii) the metabolic analysis, including data processing and chemometric analysis, and (iv) the identification of the molecular features which show statistical differences along different roasting degrees, i.e. markers of the roasting process.

2. Materials and methods

2.1 Reagents and chemicals

Acetonitrile, ethanol, methanol, and formic acid of MS grade were purchased from Fisher Scientific (Hampton, New Hampshire, USA). Ultrapure water for the chromatographic mobile phase and for preparing coffee extracts was obtained from a Milli-Q system (Millipore, Madrid, Spain). Ammonium formate, sodium borate, phosphoric acid, verapamil, niflumic acid, propranolol, terfenadine, chlorogenic acid, shikimic acid, trans-caffeic acid, paraxanthine, mannose, quinic acid, theobromine, caffeic acid were purchased from Sigma (St. Louis, MO, USA). Neochlorogenic acid, 1,3-dicaffeoylquinic acid, 1,5-dicaffeoylquinic acid were purchased from Plantachem (Pinnow, Germany).

2.2 Coffee beans

Green coffee beans (GCB) of the Arabica variety were roasted to light level at 175 ºC during 12.36 min (LRB), medium level (MRB) at 185 ºC during 14.11 min, and to dark level (DRB) at 195 ºC during 17.06 min. The weight loss of each sample was evaluated in order to control the roasting process being 13% in light coffee, 15% in medium coffee and 17% in dark coffee. All these samples were grounded and provided by "Café Fortaleza" (Madrid, Spain).

2.3 Sample preparation

Grounded coffee samples were extracted using methanol (25 % in water) as extraction solvent (50 mg in 1.5 mL). The solid-liquid extraction procedure was performed using a Thermomixer Compact (Eppendorf AG, Hamburg, Germany) at 700 rpm during 15 min at room temperature (25 ºC). After extraction, the samples were centrifuged at 3500 rpm

for 10 min at 25 ºC. Next, the supernatant fraction was injected in the RP-LC-MS system. Replicate extraction of each group of coffee samples (GCB, LRB, MRB and DRB) (n = 5) were prepared for the metabolomic sequence.

A quality control (QC) sample was prepared by combining equal aliquots from each coffee extract (GCB, LRB, MRB, and DRB). Moreover, a test sample was prepared by adding four standards (verapamil, niflumic acid, propranolol, and terfenadine) at 0.1 µg/mL to the QC sample.

2.4 RP-LC-MS conditions

A 1100 series LC system (Agilent Technologies, Palo Alto, CA, USA) coupled to a 6530 series quadrupole time-of-flight (QTOF) mass spectrometer (Agilent Technologies, Germany) equipped with a Jet Stream thermal orthogonal electrospray ionization (ESI) source was employed to perform the analyses. MS control, data acquisition and data analysis were carried out using the Agilent Mass Hunter Qualitative Analysis software (B.07.00).

Chromatographic separation was performed on a porous-shell fused-core Ascentis Express C18 analytical column (150 x 2.1 mm, particle size 2.7 µm) protected by an Ascentis Express C18 guard column (0.5 cm × 2.1 mm, 2.7 µm particle size), both from Supelco (Bellefonte, PA, USA). LC analyses were performed using mobile phases composed of water containing 0.1 % formic acid (solvent A) and acetonitrile containing 0.1 % formic acid (solvent B) eluted according to the following gradient: 5-100 % B in 45 min; 100 % B during 4 min; 100-5 % B in 2 min; and then the column was re-equilibrated for 15 min using the initial solvent composition. The mobile phase flow

rate was 0.2 mL/min, the column temperature was set to 30 ºC, and the injection volume was 5 µL.

MS analyses were carried out both in positive and negative ESI modes with the mass range set at *m/z* 100-1700 (extended dynamic range) in full scan resolution mode at a scan rate of 2 scans per second (mass resolution greater than 5000 on the 118 m/z and 10000 on the 1522 m/z according to the instrument specifications). ESI parameters for the mass spectrometer were as follows: capillary voltage for positive and negative ionization modes of 3000 V with a nozzle voltage of 0 V; nebulizer pressure at 25 psi; sheath gas of jet stream of 6.5 L/min at 300 ºC; and drying gas of 10 L/min at 300 ºC. The fragmentator voltage was set at 175 V whereas the skimmer and octapole voltages were 60 V at 750 V, respectively. For MS/MS experiments, the selected precursor ions were fragmented by applying voltages between 20 and 40 V in the collision chamber.

In order to obtain proper mass accuracy, spectra were corrected using ions *m/z* 121.0508 ($C_5H_4N_4$) and 922.0097 ($C_{18}H_{18}O_6N_3P_3F_{24}$) in ESI positive, and *m/z* 119.0363 ($C_5H_4N_4$) and 966.0007 ($C_{18}H_{18}O_6N_3P_3F_{24}$ + formate) in ESI negative. To achieve this task, a solution from Agilent Technologies containing those ions was continuously pumped into the ionization source at a 15 µL/min flow rate using a 25 mL Gastight 1000 Series Hamilton syringe (Hamilton Robotics, Bonaduz, Switzerland) on a NE-3000 pump (New Era Pump Systems Inc., Farmingdale, NY, USA).

2.5. Metabolomics sequence

The metabolomics sequence was designed as follows: blanks and QC sample were injected at the beginning of the metabolic sequence to ensure good stability and repeatability of the chromatographic system. Then, a total of 60 coffee samples (four

groups of samples and five replicates for each group injected in triplicate) were randomly injected and a QC sample was injected every six coffee samples.

Moreover, a QC sample containing four known standards (see section 2.3) was injected eight times during the sequence in order to evaluate the mass accuracy (lower than 4 ppm) and retention time shifting (RSD around 0.2 %).

2.6 Data processing and multivariate analysis

Molecular Feature Extraction (MFE) tool from Mass Hunter Qualitative Analysis (B.07.00) was used to obtain the information related to the molecular features, i.e. chromatographic peaks, present in each sample. The MFE extraction algorithm selected was "small molecules (chromatographic)" using the following parameters: ions ε 500 counts; peak spacing tolerance = 0.0025 m/z, plus 7.0 ppm; isotope model = common organic molecular; and limited assigned change was set to 2. To identify different ion species coming from the same molecular feature, H^+, Na^+, K^+, and NH_4^+ adducts were taken into account in positive ionization, whereas that only the $HCOO^-$ adduct was considered for negative ionization.

Filtering and alignment of the extracted molecular features were performed with Agilent Mass Profiler Professional (MPP) software (B.02.00). Molecular feature filtering was carried out using a minimum absolute abundance of 10.000 counts; number of ions 2 and all charges permitted. Molecular feature alignment was performed using a retention time window of 0.15 min, a mass tolerance of 0.02 Da and a mass window of 15 ppm. To clean data matrix from background signals, only molecular features present in 100 % of all injected QC samples with a coefficient of variation below 30 % were retained for further data analysis.

Multivariate statistical analysis was carried out using SIMCA 14.0 software (MSK Data Analytics Solutions, Umeå, Sweden) where data were centered and divided by the square root of the standard deviation as scaling factor (Pareto scaling). An unsupervised principal component analysis (PCA) was first applied to investigate clustering existing in the analyzed samples. Then, partial least squares discriminant analysis (PLS-DA) was used to discriminate samples according to their roasted degree. The quality of the models was evaluated by the goodness-of-fit parameters R^2X, R^2Y and Q^2.

2.7 Metabolites identification

Molecular features which displayed significant differences in the PLS-DA models were subjected to the identification process. Metabolite identification was performed by matching the obtained accurate mass values and the theoretical mass values (considering an error of 30 ppm in order to increase the number of possible metabolites) in the CEU Mass Mediator [27], which is a tool for searching metabolites in different databases (Kegg, Metlin, LipidMass and HMDB), and in the database FooDB (http://foodb.ca/).

In those cases, in which the standard compounds could be commercially acquired, they were analyzed under the same analytical conditions to obtain their retention time and MS/MS fragmentation in order to confirm the metabolite identity. When standards could not be acquired, experimental MS/MS spectra obtained for each molecular feature were compared to those described both in HMDB database and literature, and/or predicted MS/MS spectra obtained in CFM-ID (cfmid.wishartlab.com).

3. Results and discussion

3.1 Extraction procedure optimization

Sample treatment is probably one of the most crucial steps in metabolomics, especially in non-targeted studies where metabolites of interest are not known *a priori*.

To obtain the greatest number of metabolites from the coffee samples, different solvents, such as methanol (25, 50, 70 and 90 % in water), ethanol (25, 50, 70 and 90 % in water), 100 % water, and aqueous solutions at two different pH (2.0 and 9.0) (obtained by adding to water small amounts of phosphoric acid and sodium borate, respectively) were considered during the extraction procedure performed in a Thermomixer during 30 min at 25 ºC. All the coffee extracts obtained for every extraction solvent were analyzed by RP-LC-MS, both in positive and negative ionization modes, and the number of molecular features obtained by the MFE algorithm was employed to evaluate the extraction efficiency. 25 % methanol was selected as extracting solvent since it provided the highest number of molecular features in both ionization modes, followed by 100 % water, aqueous buffer at pH 2.0 or 9.0 (which resulted in a similar molecular feature content) and 25 % ethanol (see Table S1). Once selected the extracting solvent, the influence of the extraction procedure was evaluated in terms of using an ultrasound bath or a Thermomixer system. Even though the ultrasound bath enabled to obtain a slightly higher number of molecular features for both ESI modes, the use of the Thermomixer system was chosen since it enabled controlling the temperature of the extraction process, an important parameter which severely affects the sample stability. Next, the extraction time was evaluated. No substantial differences in the number of molecular features were observed when 15 or

30 min were used as extraction time, so 15 min was selected to achieve the extraction in both positive and negative ionization modes to short the sample preparation step.

3.2 Optimization of the RP-LC-MS metabolomics method

Optimization of the chromatographic and MS parameters is essential to improve sensitivity, to avoid peak co-elution and, consequently, to reduce possible ion suppression events, which will negatively bias the obtained results. On the one hand, chromatographic parameters were optimized by evaluating the chromatographic peak profiles observed in the total ion chromatogram (TIC) and the based peak chromatogram (BPC). First, the addition of 0.1 % of formic acid or 10 mM ammonium formate to the water/acetonitrile mobile phase was compared in both ionization modes. It was observed that the use of formic acid provided, not only a better chromatographic separation, but also a greater number of molecular features for positive and negative modes. Along with the mobile phase composition, other parameters such as gradient elution, flow rate (ranging from 0.15 to 0.25 mL/min), column temperature (ranging from 30 to 50 ºC) and injection volume (2, 5 and 10 µL) were studied in terms of chromatographic separation, peak efficiency and sensitivity to ensure the detection of the largest number of metabolites in the coffee samples. The optimal conditions were the use of water (solvent A) and acetonitrile (solvent B) both containing 0.1 % formic acid as mobile phases eluting according to the gradient described in experimental section, a flow rate of 0.2 mL/min, a column temperature of 30 ºC, and an injection volume of 5 µL. On the other hand, ESI parameters were also studied: fragmentator voltage (100-200 V), nozzle voltage (0-100V), drying gas temperature (200-350 °C) and sheath gas temperature (250-400 °C). 175 and 0 V for fragmentator and nozzle voltages,

respectively, and 300 ºC for both drying gas and sheath gas temperature, were the selected parameters. Using the optimized parameters, it was possible to detect 1206 and 1184 molecular features for positive and negative ESI modes, respectively. **Figure 1** shows the BPC obtained for GCB, LRB, MRB and DRB coffee samples under optimal conditions.

3.3 Non-targeted analysis of RP-LC-MS data by multivariate statistical analysis

Once optimized the extraction procedure and selected the most adequate RP-LC-MS parameters to carry out the metabolomic analyses, all coffee samples (a total of 60 samples) and QC samples were analyzed following the metabolomic sequence described previously (see section 2.5) and data treatment was performed according to section 2.6. A total of 10450 and 6770 features were obtained for the analyzed samples in positive and negative ionization modes, respectively. After filtering by frequency (only features present in 100 % of QC samples) and variability (features whose RSD was below 30% in QC samples), a total of 280 molecular features were obtained in positive ionization mode whereas 580 were found for the negative ionization mode.

Multivariate data analysis, in particular, principal component analysis (PCA) has been used in most of works concerning metabolomics strategies. The goal of this statistical method is to reduce large volumes of data in order to find out the most relevant variations among groups of samples [28]. Thus, PCA analysis was first employed to evaluate the consistency of the metabolomics sequence using QC sample clustering. **Figures 2A and 2B** show the PCA score plot for all the coffee and QC samples analyzed both in positive and negative ionization modes. As it can be seen in these

figures, a good clustering and high differentiation among groups of samples were obtained for both ionization modes. In addition, QC samples were also tightly clustered and were centered in the score plot, demonstrating good analytical consistency of the data during the whole metabolomics sequence. The score plots of the PCA models without the QC are shown in **Figures 2C and 2D.** The first and second components of the PCA models explained 49 % and 29 % of variance for positive ESI mode and 63 % and 20 % of variance for negative ESI mode, respectively. As the next step, supervised partial least square discriminant analysis (PLS-DA) was used to discriminate group samples. First PLS-DA models were built taking into account the four groups of coffee samples (GCB, LRB, MRB and DRB) which demonstrated that the four groups were grouped separately regardless of the ionization mode used. The quality parameters (R^2X, R^2Y, and Q^2) of both PLS-DA models are shown in **Table 1**. This table also shows the F and p-values of ANOVA test. The high values obtained for F and the low values achieved for p-values, along with the results obtained for permutation test (Q2 and R2 values are below the original values), demonstrated the quality of the model.

Taking into account that the PLS-DA models build on four groups cannot reveal slight differences existing among the group of coffee samples, different PLS-DA models to compare the groups in a pairwise way (GCB vs LRC, GCB vs MRC, and GCB vs DRC) were built to obtain further knowledge on what metabolites are affected by the roasting process. **Figure 3** and **Table 1** show the PLS-DA models and the quality parameters obtained for each pairwise comparison. High Q2 values (> 0.970) were obtained in all models. F and p-values obtained in the cross validated ANOVA of PLS-DA models reinforced the robustness of the proposed models and demonstrated the good separation between groups (F values higher than 151.2 and p-values lower than 3.9×10^{-17}). In order to verify that group sample separation was due to real differences in metabolic

profiles of analyzed samples and not just due to data overfitting, cross validation (CV) and permutation tests were performed. As can be seen in **Figures S1 and S2,** pairwise PLS-DA models were considered as valid and robust since CV score plots showed group separation and slope obtained in the permutation tests was positive for R^2 and Q^2 values. This fact corroborated what is observed in PCA models, i.e. there were real statistical significant differences in the metabolic profiles from the studied groups and therefore it was possible to establish a set of variables responsible for these differences. The selection of these variables was performed according to the variable importance in the projection (VIP) value for each PLS-DA model. Thus a total of 25 and 34 variables with VIP values higher than 1.5 were selected as the most influencing variables in positive and negative ionization modes, respectively.

3.4 Metabolite identification

Metabolite identification usually corresponds to the most laborious step within metabolomics studies, not only for being time-consuming but also due to different factors such as the lack of standards for some compounds (or their high price) or the difficulties in the comparison of the MS/MS spectra for others. Bearing in mind that the identification is based on the accurate mass, isotopic pattern, and MS/MS pathway of each variable selected from the multivariate analysis, a useful and vital tool is the use of tandem high resolution MS to acquire both MS and MS/MS mass spectra. Using the procedure described in section 2.7, the accurate mass information was employed to propose a list of possible metabolites. Among them, only the ones corresponding with compounds whose presence was probable in food and plants were considered. Following this strategy, 24 and 33 molecular features were selected as potential markers

of coffee roasting process for positive and negative ionization mode, respectively. One of these molecular features was found in both ionization modes. This fact highlights the importance of combining both ESI modes in order to maximize the number of detected metabolites.

Tables 2 (positive mode) **and 3** (negative mode) summarize the retention time, the molecular formula, the experimental m/z value, the mass error comparing with the database, the main fragments obtained in MS/MS spectra, the VIP values for the pairwise PLS-DA models, and the trend observed for all significant metabolites along the roasting process of coffee (**Figure S3** of supporting information shows the diagrams of the trends observed for all the tentatively and unequivocally compounds along the coffee roasting process). As it can be seen in both tables, 7 and 13 metabolites were identified in positive and negative ionization modes (16 were tentatively identified and 4 unequivocally identified).

In the positive ionization mode, the 355 *m/z* ion as $[M+H]^+$ was observed for two different compounds: **compound 4** (t_R = 11.9 min) and **compound 6** (t_R = 13.4 min). The fragment ions obtained from the MS/MS spectra of both compounds were *m/z* 163 and *m/z* 145, corresponding to the caffeoyl residue and its loss of water. **Compound 4** was unequivocally identified as chlorogenic acid based on the comparison of its retention time and MS/MS fragmentation pattern to those obtained for the commercial standard. **Compound 6** was tentatively identified as a chlorogenic acid isomer-1. Interestingly, chlorogenic acid (compound 4) levels decreased with the roasting process whereas that the isomer (compound 6) is a clear marker of DRC roasting as its level only decreased when comparing GCB vs DRC. **Compounds 12** (t_R = 23.5 min), **13** (t_R = 23.8 min), and **14** (t_R = 24.5 min) all exhibited $[M+H]^+$ ions at *m/z* 517. MS/MS

fragmentation of these ions gave a fragment ion at 355 *m/z* suggesting the loss of a caffeoyl residue (163 Da). By matching the MS/MS spectra obtained to those reported in the literature [29, 30], these compounds were tentatively identified as dicaffeoylquinic acid isomer. Compound 13 is the same as compound 25 obtained in negative ionization mode which was unequivocally identified as 1,5-dicaffeoylquinic acid by comparison of its retention time and MS/MS fragmentation pattern with those of the standard. In this negative mode, the MS/MS spectrum of this compound shows fragment ions at *m/z* 353 ([M-H-caffeoyl]), *m/z* 191, ([M-H-caffeoyl-caffeoyl]), and *m/z* 179 ([M-H-caffeoyl-quinic]) which were in agreement with those fragments described previously in the literature for this compound [29, 30]. The trend of 1,5-dicaffeoylquinic acid is to decrease as roasting process increases.

As it can be seen in **Table 3**, for negative ionization mode, the $[M-H]^-$ ions at *m/z* 353 and *m/z* 515 are also highlighted as markers, corresponding to compounds 6 and 25, respectively. Standards of neochlorogenic acid (t_R = 8.0 min) and 1,3-dicaffeoylquinic acid (t_R = 19.7 min) were analyzed in order to obtain their retention times and MS/MS pattern to compare with those pointed out in the tables. However, the possibility of corresponding any of these standards with a metabolite was discarded by differences in the retention time. The MS/MS spectra of **compound 6** (t_R = 6.8 min) showed a fragment ion *m/z* 191 corresponding to [M-H-caffeoyl], and *m/z* 179 corresponding to [M-H-quinic] [29, 30]. According to that, this compound could be tentatively identified as a chlorogenic acid isomer-2 (different from compounds 4 and 6 in positive ionization mode because they had different retention times (see **Tables 2** and **3**)). The levels of this chlorogenic acid isomer-2 significantly increased from GCB to LRC but then gradually decreased with further roasting.

Compound 22 (t_R = 20.8 min) obtained in negative mode displayed the deprotonated molecular ion at *m/z* 335. Its MS/MS fragmentation showed the characteristics fragments of chlorogenic acids (i.e *m/z* 179 m/z corresponding to [M-H-caffeoyl-quinic], *m/z* 161 corresponding to [M-H-caffeoyl-quinic-H_2O], *m/z* 135 [M-H-caffeoyl-quinic-CO_2]), except by the absence of signal *m/z* 191. This is an important feature which helps to distinguish between chlorogenic acids and chlorogenic acid lactones, as reported elsewhere [31]. Moreover, the increasing trend that can be observed for this compound during roasting (**Figure S3**) suggested that this compound could be a chlorogenic acid lactone isomer-1 instead of a shikimic acid derivative which presents the same fragmentation pattern (see below).

Other hydroxycinnamic acids (coumaric acid, cinnamic acid, ferulic acid, caftaric, caffeic acid, etc.) were also tentatively identified in positive and negative ionization mode. In positive mode, **compound 21** (t_R = 28.4 min, $[M-H]^+$ = 147 had MS/MS fragment ions at *m/z* 65 and *m/z* 91 that, according to Melo et al. correspond to coumarin fragmentation pattern [32]. In negative ESI mode belonging to the hydroxycinnaminc acids, **compounds 11** (t_R = 13.2 min) **and 18** (t_R = 19.0 min) displayed roughly the same $[M-H]^-$ ion at *m/z* 367. The MS/MS spectra of these ions showed the main fragment at *m/z* 193 (compound 11) and *m/z* 191 (compound 18). According to the literature, these fragments indicated the presence of feruloyl and caffeoyl groups, respectively [31]. Therefore, there were assigned tentatively as feruloylquinic acid isomer-1 (compound 11) and caffeoyl-methylquinic acid isomer-1 (compound 18). These compounds showed a different behavior when increasing the roasting level. On the one hand, feruloylquinic acid isomer-1 had significant differences in the two first roasting levels (light and medium) when compared to GCB, however, no

differences were found when comparing GCB to DRC. On the other hand, interestingly feruloylquinic acid isomer-1 has a similar trend to the previously commented chlorogenic acid isomer-2 (compound 6 in negative mode), i.e. its level increased from GCB to LRC and then decreased with further roasting.

Also in negative ionization mode, two different compounds (**15 and 16**) shared the same $[M-H]^-$ ion at *m/z* 337 and had similar fragmentation; both show fragment ions at *m/z* 163 ($[M-H\text{-quinic}]^-$) and *m/z* 119 ($[M-H\text{-quinic-}CO_2]^-$). Moreover, an ion fragment at *m/z* 191 for compound 15 corresponding to quinic acid residue and *m/z* 173 for compound 16 corresponding to [quinic acid-H_2O-H]$^-$ were also observed. These compounds, according to the fragmentation described in literature by Clifford et al. were tentatively identified as coumaroylquinic acid isomers-1 and 2 [33].

Under MS/MS fragmentation of **compound 17** (t_R = 17.5 min, $[M-H]^-$ = 335) in negative ionization mode, ion *m/z* 179 was obtained, which, as previously reported suggested the presence of caffeic acid, and *m/z* 161 and *m/z* 135 were formed from loss of H_2O and CO_2, respectively [29]. According to that, this compound was tentatively identified as a caffeoylshikimic acid isomer-1. Other hydroxycinnamic acids, such as **compound 5** (t_R = 6.8 min), **compound 27** (t_R = 25.3 min) and **compound 30** (t_R = 27.0 min) that displayed a $[M-H]^-$ ion at *m/z* 179, *m/z* 529 and *m/z* 365 respectively, in negative ionization mode, were shown to have differences in their levels along the roasting process. Compound 5 was tentatively identified as a caffeic acid isomer-1 since its MS/MS fragmentation pattern corresponded to the fragmentation obtained for caffeic acid and trans-caffeic acid standards (*m/z* 135 and *m/z* 134), but whose retention time differed from those of these compounds. According to the fragmentation reported by Lukáš et al. compound 27 was identified as a caffeoyl-feruloyquinic acid isomer-1

because of the presence of MS/MS fragments at *m/z* 367 ([M-H-caffeoyl]$^-$), *m/z* 193 ([M-H-caffeoyl-quinic]$^-$), and *m/z* 335 ([M-H-ferulic acidl]$^-$) characteristics of this family of compounds [31]. Finally, compound 30 gave a MS/MS fragmentation at *m/z* 135 and *m/z* 161, corresponding with caffeoyl family and *m/z* 203 which corresponded to the theoretical mass of tryptophan suggesting that this compound could be tentatively identified as N-caffeoyltryptophan. Regarding the behavior of these compounds during the roasting process, it was observed, on the one hand, that caffeoyl-feruloyquinic acid isomer-1 (compound 27) and N-caffeoyltryptophan (compound 30) had an opposite trend, whereas the former decreased during the roasting process, the latter increased. On the other hand, the caffeic acid isomer-1 (compound 5) decreased during roasting disappearing in DRC.

Moreover, other compounds proposed as markers are **compounds 1 and 2** from Table 3. These compounds, with the same retention time (t_R = 1.8), gave a [M-H]$^-$ at *m/z* 191 and *m/z* 179, respectively. In this case, their MS/MS fragmentation patterns suggested that they could be quinic acid and hexose. Compound 1 was unequivocally identified as quinic acid by comparison with the standard. Paraxanthine and theobromine standards were also analyzed since their mass was in agreement with that found for compound 2. However, they showed a different retention time and MS/MS spectra than the ones experimentally obtained. For both compounds (1 and 2) their levels increased gradually with the roasting process.

The last metabolite tentatively identified in positive ionization mode (**compound 11,** t_R = 22.6 min) is of a great interest. It gave a [M+H]$^+$ ion at *m/z* 509, presented the highest VIP values, and its composition decreased with roasting. The experimental mass of compound 11 is in agreement with theoretical mass and the MS/MS fragmentation

pattern of mozambioside, which has previously been reported by Roman et al., as interesting compound in coffee beans that exhibit a bitter taste [34]. In that work authors quantified this compound in Arabica coffee samples demonstrating its degradation from green to roasted coffee. In summary, the proposed RP-LC-MS methodology enables to identify markers of the roasting process that had not previously been identified using non-targeted metabolomic approaches. To obtain a more comprehensive analysis it should be necessary to apply the developed metabolomics platform to a larger number of samples.

4. Conclusions

The present study describes the development of an untargeted metabolomics strategy based on RP-LC-MS to investigate changes occurring in coffee samples submitted to three different roasting degrees (light, medium, and dark coffee). By using this method, it was possible to obtain a list of 24 and 33 compounds that have demonstrated to be potential markers of roasting process of coffee for positive and negative ionization modes, respectively. Only one of these compounds appeared as significant in both ionization modes, highlighting the importance of using both ESI modes to carry out the metabolomics analysis. A total of 7 and 13 metabolites were identified as markers of roasting process in positive and negative modes, respectively. Most of these compounds belong to the group of hydroxycinnamic acids. In general, in most cases these metabolites decreased with the roasting process, although quinic acid, hexose, chlorogenic acid lactone and N-caffeoyltryptophan showed a different trend, i.e. they increased at high roasted degrees. Finally, an interesting compound belonging to the naphthofurans family, has demonstrated to be an important biomarker since it presents one of the highest VIP values in the list of markers and its levels decrease during

roasting. Thus, the developed metabolomics strategy demonstrated not only to be a useful tool to differentiate coffee beans submitted up to three different roasting degrees but also to highlight potential markers of the roasting process.

Acknowledgements

Authors thank financial support from the Comunidad of Madrid (Spain) and European funding from FEDER program (project S2013/ABI-3028, AVANSECAL-CM) and the University of Alcalá (project CCG2015/EXP-032). R.P.M. thanks the University of Alcalá for her pre-doctoral contract. M.C.P. and M.P. thank MINECO for their "Ramón y Cajal" (RYC-2013-12688) and "Juan de la Cierva" (IJCI-2014-22143) research contracts, respectively. Authors gratefully acknowledge "Café Fortaleza" for providing coffee samples. Authors thank Dr. F.J. Morales (ICTAN, CSIC, Madrid, Spain) for fruitful scientific discussions.

References

[1] Esquivela P, Jiménez VM. Functional properties of coffee and coffee by-products. Food Res Int. 2012; 46: 488-95.

[2] Wei F, Furihata K, Hu F, Miyakawa T, Tanokura M. Two-dimensional 1H–13C nuclear magnetic resonance (NMR)-based comprehensive analysis of roasted coffee bean extract. J Agric Food Chem. 2011; 59: 9065-73.

[3] Jeszka-Skowron M, Zgoła-Grześkowiak A, Grześkowiak T. Analytical methods applied for the characterization and the determination of bioactive compounds in coffee. Eur Food Res Technol 2015; 240: 19-31.

[4] Joëta T, Laffargue A, Descroix F, Doulbeau S, Bertrand B, De kochko A, Dussert S. Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. Food Chem. 2010; 118: 693-701.

[5] Moon JK, Yoo HS, Shibamoto T. Role of roasting conditions in the level of chlorogenic acid content in coffee beans: correlation with coffee acidity. J Agric Food Chem. 2009; 57: 5365-69.

[6] Ginz M, Engelhardt U. Analysis of bitter fractions of roasted coffee by LC-ESI-MS-new chlorogenic acid derivatives. Association Scientifique Internationale du Café (ASIC). 2001; 1-5.

[7] Koen Bekedam E, Schols HA, Cämmerer B, Kroh LW, Van Boekel MAJS, Smit G. Electron spin resonance (ESR) studies on the formation of roasting-induced antioxidative structures in coffee brews at different degrees of roast. J Agric Food Chem. 2008; 56: 4597-604.

[8] Delgado-Andrade C, Morales FJ. Unraveling the contribution of melanoidins to the antioxidant activity of coffee brews. J Agric Food Chem. 2005; 53: 1403-7.

[9] De Maria CAB, Trugo LC, Aquino Neto FR, Moreira RFA, Alviano CS. Composition of green coffee water-soluble fractions and identification of volatiles formed during roasting. Food Chem. 1996; 55: 203-7.

[10] Mussatto SI, Machado EMS, Martins S, Teixeira JA. Production, composition, and application of coffee and its industrial residues. Food Bioprocess Technol. 2011; 4: 661-72.

[11] Casal S, Oliveira MBPP, Alves MR, Ferreira MA. Discriminate analysis of roasted coffee varieties for trigonelline, nicotinic acid, and caffeine Content. J Agric Food Chem. 2008; 48: 3420-24.

[12] Garrett R, Vaz BG, Hovell AMC, Eberlin MN, Rezende CM. Arabica and robusta coffees: Identification of major polar compounds and quantification of blends by direct-infusion electrospray ionization–mass spectrometry. J Agric Food Chem. 2012; 60: 4253-58.

[13] Toledo BR, Hantao LW, Ho TD, Augusto F, Anderson JL. A chemometric approach toward the detection and quantification of coffee adulteration by solid-phase microextraction using polymeric ionic liquid sorbent coatings. J Chromatogr A. 2014; 1346: 1-7.

[14] Fiehn O. Metabolomics-the link between genotypes and phenotypes. In: Town C Eds.). Functional Genomics, Springer. Dordrecht: 2002. pp. 155-71.

[15] Wei F, Furihata K, Koda M, Hu F, Kato R. ^{13}C NMR-Based Metabolomics for the classification of green coffee Beans According to Variety and Origin. J Agric Food Chem. 2012; 60: 10118-25.

[16] Consonni R, Ruth L, Cogliati CC. NMR based geographical characterization of roasted coffee. Talanta. 2012; 88: 420-26.

[17] Garrett R, Schmidt EM, Pereira LFP, Kitzberger CSG, Scholz MBS, Berlin MNE, Rezende CM. Discrimination of Arabica Coffee cultivars by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and chemometrics. Food Sci Technol. 2013; 50: 496-502.

[18] Setoyama D, Iwasa K, Seta H, Shimizu H, Fujimura Y, Miura D, Wariishi H, Nagai C, Nakahara K. High-throughput metabolic profiling of diverse green coffea arabica beans identified tryptophan as a universal discrimination factor for immature beans. PLOS ONE. 2013; 9: e70098.

[19] Choi MY, Choi W, Park JH, Lim J, Kwon SW. Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. Food Chem. 2010; 121:1260-68.

[20] Da Silva Taveira JH, Meira Borém F, Pereira Figueiredo L, Reis N, Franca AS, Harding SA, Tsai CJ. Potential markers of coffee genotypes grown in different Brazilian regions: A metabolomics approach. Food Res Int. 2014; 61: 75-82.

[21] Chang KL, Ho PC. Gas chromatography time-of-flight mass spectrometry (GC-TOF-MS)-based metabolomics for comparison of caffeinated and decaffeinated coffee and its implications for Alzheimer's disease. PLOS ONE. 2014; 9 e104621.

[22] Gloess AN, Yeretzian C, Knochenmuss R, Groessl M. On-line analysis of coffee roasting with ion mobility spectrometry (IMS-MS). Int J Mass Spectrom. 2018; 424: 49-57.

[23] Wei F, Furihata K, Koda M, Hu F, Miyakawa T, Tanokura M. Roasting process of coffee beans as studied by nuclear magnetic resonance: time course of changes in composition. J Agric Food Chem. 2012; 60: 1005-12.

[24] Santos da Rosa J, Freitas-Silva O, Costa Rouws JR, Gonçalves da Silva Moreira I, Moreira Novaes FJ, De Almeida Azevedo D, Schwab N, Luiz de Oliveira Godoy R, Nogueira Eberlin M, Moraes de Rezende C. Mass spectrometry screening of Arabica coffee roasting: A non-target and non-volatile approach by EASI-MS and ESI-MS. Food Res Int. 2016; 89: 967-75.

[25] Liberto E, Ruosi MR, Cordero C, Rubiolo P, Bicchi C, Sgorbini B. Non-separative headspace solid phase microextraction-mass spectrometry profile as a marker to monitor coffee roasting degree. J Agric Chem. 2013; 61: 1652-60.

[26] Ruosi MR, Cordero C, Cagliero C, Rubiolo P, Bicchi C, Sgorbini B, Liberto E. A further tool to monitor the coffee roasting process: aroma composition and chemical indices. J Agric Chem. 2012; 60: 11283-91.

[27] Gil de la Fuente A, Godzien J, Fernández-López M, Rupérez FJ, Barbas C, Otero A. Knowledge-based metabolite annotation tool: CEU Mass Mediator. J. Pharm. Biomed. Anal. 2018; 154: 138-149.

[28] Deshpande S, El-Abassy RM, Jaiswal R, Eravuchira P, Von der Kammer B, Materny A, Kuhnert N. Which spectroscopic technique allows the best differentiation of coffee varieties: comparing principal component analysis using data derived from CD-,

NMR- and IR-spectroscopies and LC-MS in the analysis of the chlorogenic acid fraction in green coffee beans. Anal Methods. 2014; 6: 3268-76.

[29] Gouveia S, Castilho PC. Characterisation of phenolic acid derivatives and flavonoids from different morphological parts of Helichrysum obconicum by a RP-HPLC–DAD-(-)–ESI-MS^n method. Food Chem. 2011; 129: 333-44.

[30] Bondia-Pons I, Savolainen O, Törrönen R, Martinez JA, Poutanen K, Hanhineva K. Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. Food Res Int. 2014; 63: 132-38.

[31] Kučera L, Papoušek R, Kurka O, Barták P, Bednár P. Study of composition of espresso coffee prepared from various roast degrees of *Coffea arabica* L. coffee beans. Food Chemistry. 2016; 199: 727-35.

[32] De Melo LV, Sawaya ACHF. UHPLC–MS quantification of coumarin and chlorogenic acid in extracts of the medicinal plants known as guaco (*Mikania glomerata* and *Mikania laevigata*). Rev Bras farmacogn. 2015; 25: 105-10.

[33] Clifford MIN, Knight S, Kuhner N. Discriminating between the Six Isomers of Dicaffeoylquinic Acid by LC-MS^n. J Agric Food Chem. 2005; 53: 3821-32.

[34] Lang R, Klade S, Beusch A, Dunkel A, Hofmann T. Mozambioside is an arabica-specific bitter-tasting furokaurane glucoside in coffee beans. J Agric Food Chem. 2015; 63: 10492-99.

Figure captions

Figure 1. Base peak chromatograms (BPC) obtained in negative ionization mode for green coffee (GCB) (A); light coffee (LRC) (B); medium coffee (MRC) (C); and dark coffee (DRC) (D) under optimal separation conditions. RP-LC-MS conditions are summarized in Section 2.4.

Figure 2. Principal component analysis (PCA) score plot obtained in positive (A) and negative (B) ionization modes, for the four studied coffee groups (GCB, LRC, MRC and DRC) submitted to different roasting degree with QC samples, and PCA score plot obtained in positive (C) and negative (D) ionization modes for coffee sample submitted to different roasting degree without QC samples.

Figure 3. Partial least squares discriminant analysis (PLS-DA) score plots of LRC, MRC and DRC compared with GCB in negative ionization mode (A, B and C, respectively) and in positive ionization mode (D, E and F, respectively).

Figure 1.

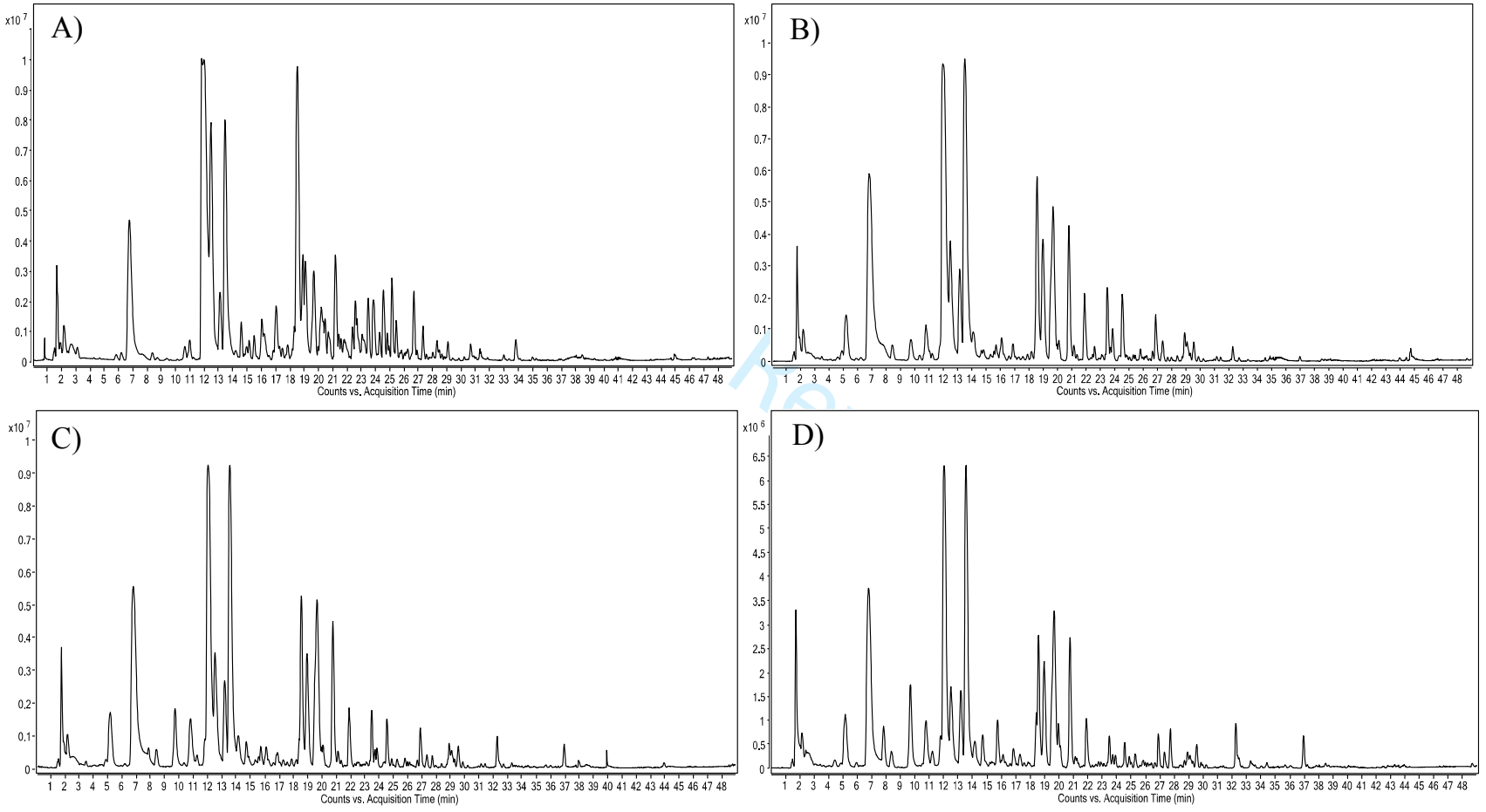


Figure 2.

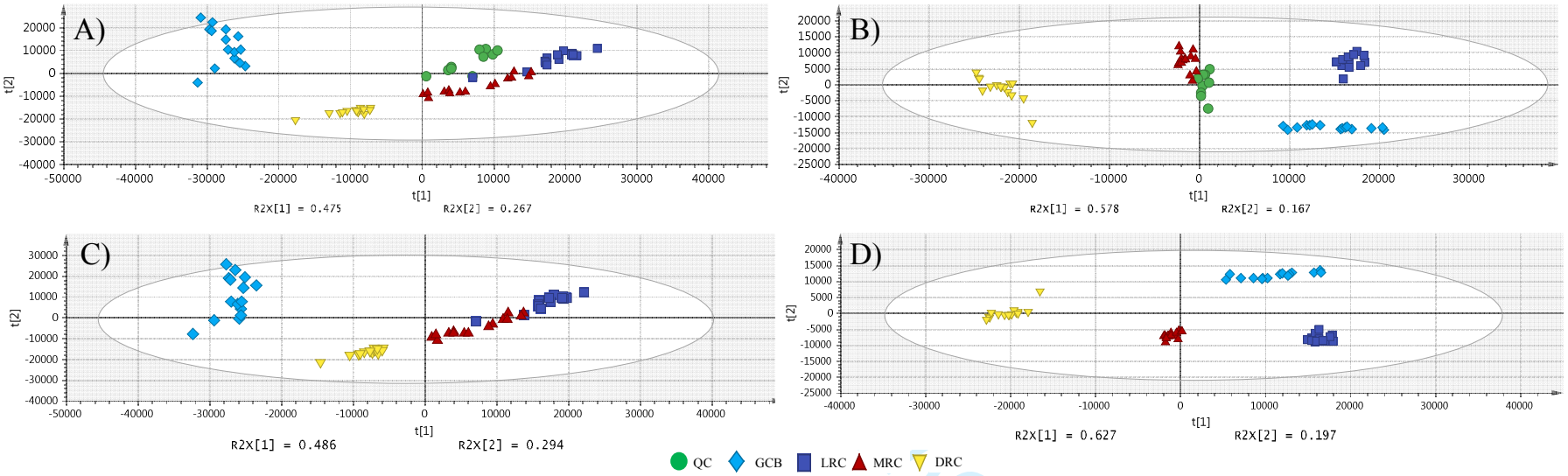

A)
B)
C)
D)
t[1]
t[2]
R2X[1] = 0.475
R2X[2] = 0.267
R2X[1] = 0.578
R2X[2] = 0.167
R2X[1] = 0.486
R2X[2] = 0.294
R2X[1] = 0.627
R2X[2] = 0.197
QC
GCB
LRC
MRC
DRC

Figure 3.

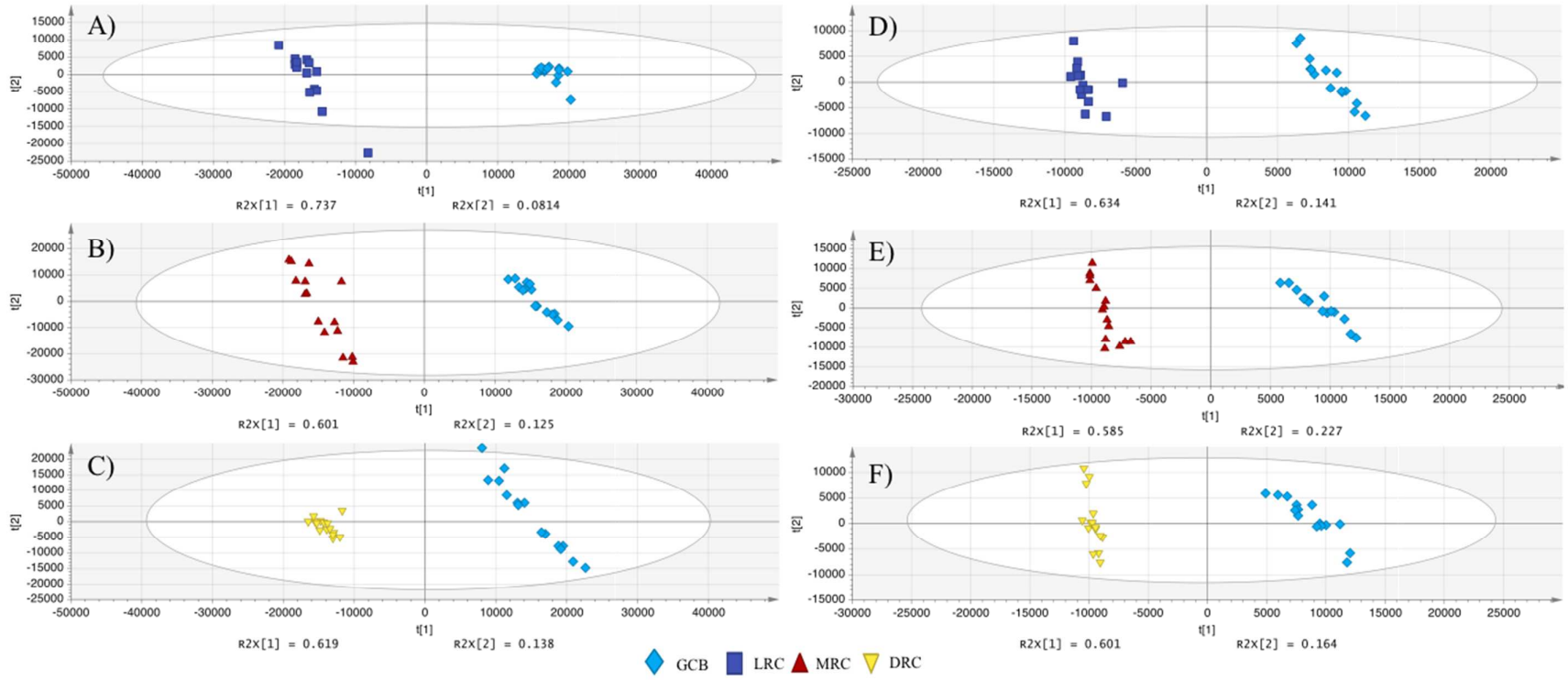

A)
R2X[1] = 0.737
R2X[2] = 0.0814
B)
R2X[1] = 0.601
R2X[2] = 0.125
C)
R2X[1] = 0.619
R2X[2] = 0.138
D)
R2X[1] = 0.634
R2X[2] = 0.141
E)
R2X[1] = 0.585
R2X[2] = 0.227
F)
R2X[1] = 0.601
R2X[2] = 0.164
t[1]
t[2]
GCB
LRC
MRC
DRC

Table 1. PLS-DA models for samples submitted to different roasting degree compared with green coffee.

	ESI+		ESI-	
	Quality parameters	**F (and P-values) of cross-validated ANOVA**	**Quality parameters**	**F (and P-values) of cross-validated ANOVA**
GCB, LRC, MRC, DRC	R2X = 0.931 R2Y = 0.976 Q2 = 0.951	71.4 (0)	R2X = 0.947 R2Y = 0.966 Q2 = 0.895	12.9 (1.5×10^{-30})
GCB vs LRC	R2X = 0.775 R2Y = 0.992 Q2 = 0.987	394.5 (3.4×10^{-22})	R2X = 0.819 R2Y = 0.996 Q2 = 0.984	428.0 (1.26×10^{-22})
GCB vs MRC	R2X = 0.812 R2Y = 0.989 Q2 = 0.977	234.1 (2.0×10^{-19})	R2X = 0.726 R2Y = 0.991 Q2 = 0.982	286.5 (1.7×10^{-20})
GCB vs DRC	R2X = 0.765 R2Y = 0.984 Q2 = 0.970	151.2 (3.9×10^{-17})	R2X = 0.756 R2Y = 0.992 Q2 = 0.978	202.2 (1.2×10^{-18})

Table 2. MS/MS fragmentation of the selected compounds in positive ionization mode

#	RT	Molecular formula	Tentative identification	$[M+H]^+$	Mass error (ppm)	Main MS/MS fragments	VIP values			Roasting trend
							GCB vs LRC	GCB vs MRC	GCB vs DRC	
1	1.6		Unknown	104.1046		58.0647 44.0495 42.0338	4.7782	5.29545	2.93603	↓
2	1.7		Unknown	176.0057		118.0626 132.0802	2.56381	1.4932	1.6759	—
3	2.6		Unknown	453.1327		435.144 211.9747 140.9606	2.00091	2.0663	2.13435	↓
4	11.9	$C_{16}H_{18}O_9$	Chlorogenic acid*	355.0998	7	163. 0376 145.0236	4.30062	6.38721	8.13866	↓
5	12.0		Unknown	370.1474		147.0416 208.0939	1.4292	1.74676	2.30126	↓
6	13.4	$C_{16}H_{18}O_9$	Chlorogenic acid isomer-1	355.0990	10	163.0367 145.0261	0.3627	0.66204	2.30032	—
7	18.5		Unknown	759.2100		391.0958	1.92238	2.94479	3.25095	—
8	21.9		Unknown	303.1966		285.1718 211.1442 131.0828 119.0821	6.13668	5.15103	4.99282	↑
9	21.9		Unknown	500.2074		163.0235 239.1502	4.51142	3.74334	3.23861	↑
10	21.9		Unknown	505.2418		152.1014 180.097	1.68956	1.4904	1.4254	↑
11	22.6	$C_{26}H_{36}O_{10}$	Mozambioside	509.2402	4	147.0393 329.1696 347.1804	8.09174	8.58909	8.71963	↓
12	23.5	$C_{25}H_{24}O_{12}$	Dicaffeoylquinic acid isomer-1	517.1352	2	163.0338 353.1451	3.42601	0.3839	2.42987	↓
13	23.8	$C_{25}H_{24}O_{12}$	1.5-Dicaffeoylquinic acid*	517.1372	6	163.0278 355.1305	4.62276	5.77454	3.88 E-8	↓
14	24.5	$C_{25}H_{24}O_{12}$	Dicaffeoylquinic acid isomer-2	517.1365	5	163.0338 355.1666	3.54144	0.8314	3.10321	↓
15	25.7		Unknown	513.1409		177.0479 163.0313	2.24824	1.33E-08	1.05E-08	↓
16	25.8		Unknown	421.1526		163.0353 241.1036	1.64952	1.0751	0.4846	—
17	26.4		Unknown	421.1532		163.0316 241.0982	1.2018	1.73713	1.98771	↓
18	26.9		Unknown	247.1192		171.0746 115.0453	1.8825	1.56312	1.32956	↑
19	26.9		Unknown	751.3548		Not clear MS/MS	1.60329	7.36E-10	0.9139	↑
20	27.3		Unknown	421.1531		163.0306 241.097	1.7343	2.00181	1.84203	↓
21	28.4	$C_9H_6O_2$	Coumarin	147.0442	1	65.0378 91.0531	1.0416	1.3966	1.67408	↓
22	34.9		Unknown	277.1806		137.0577	2.04446	1.95653	1.92496	↓
23	35.7		Unknown	668.4022		331.1739	1.67226	1.58306	1.40786	↑
24	39.4		Unknown	415.2130		119.0843	0.0922	0.2322	1.51976	—

* Confirmed with standard.

↑ The level of the compound increases with roasting.
↓ The level of the compound decreases with roasting.
— Indicates a different trend from the previously described.

Table 3. MS/MS fragmentation of the selected compounds in negative ionization mode

#	RT	Molecular formula	Tentative identification	[M-H]$^-$	Mass error (ppm)	Main MS/MS fragments	VIP values GCB vs LRC	GCB vs MRC	GCB vs DRC	Roasting trend
1	1.8	$C_7H_{12}O_6$	Quinic acid*	191.0586	13	108.022 109.0294 127. 0384	3.37525	3.89183	5.45095	↑
2	1.8	$C_6H_{12}O_6$	Hexose	179.0585	13	134.0348 117.0325	0.890106	1.06703	1.49571	↑
3	2.1		Unknown	341.1090		119.031 101.021 113.0204	1.68073	1.90 E-8	1.52 E-8	↓
4	3.1		Unknown	371.0977		191.0533 173.0440	2.06996	2.63868	2.77613	↓
5	6.8	$C_9H_8O_4$	Caffeic acid isomer-1	179.0346	2	135.0404 134.0333	1.50793	1.23632	0.291829	—
6	6.8	$C_{16}H_{18}O_9$	Chlorogenic acid isomer-2	353.0871	2	191.0528 179.0317 135.0419	11.4834	8.53885	3.02571	—
7	10.7		Unknown	705.1616		353.081 179.0283	1.60495	0.56905	0.900738	—
8	12.3		Unknown	857.2334		503.1273 353.0793 191.0514	1.75214	2.26314	2.5821	—
9	12.5		Unknown	707.1760		353.0864 191.0548	8.16811	10.0931	12.0855	↓
10	13.2		Unknown	193.0446		134.0348 117.0325	1.61384	1.17043	0.14827	—
11	13.2	$C_{17}H_{20}O_9$	Feruloylquinic acid isomer-1	367.0978	15	193.0456 134.0325	3.6265	2.36628	0.14504	—
12	13.8		Unknown	379.1584		113.0233 119.0322 191.0555	1.26436	1.63921	1.73346	↓
13	15.2		Unknown	597.1870		173.0706 132.0279 481.1830	2.31266	2.66958	2.83305	↓
14	16.0		Unknown	241.1114		197.1291 141.1019	0.56638	1.02896	1.57448	↑
15	16.1	$C_{16}H_{18}O_8$	Coumaroylquinic isomer-1	337.0835	28	191.0534 163.0377 119.0467	2.07409	2.90015	3.56737	↓
16	16.9	$C_{16}H_{18}O_8$	Coumaroylquinic isomer-2	337.0824	31#	173.0444 163.0393 119.0495	1.82156	1.4922	0.877095	—
17	17.5	$C_{16}H_{16}O_8$	Caffeoylshikimic acid isomer-1	335.0712	18	179.0340 135.0444 161.0242	1.85381	2.15126	0.69881	↓
18	19.0	$C_{17}H_{20}O_9$	Caffeoyl-methylquinic acid isomer-1	367.1021	4	191.0534 173.0425	4.75652	0.840603	5.60833	—
19	19.0		Unknown	173.0422		108.0177 109.0287	1.48052	0.70964	0.501266	—
20	20.0		Unknown	569.2246		Not clear MS/MS	1.88176	1.32728	1.46 E-8	↓
21	20.5		Unknown	559.2758		351.2166	3.40067	6.38 E-8	3.43 E-8	↓
22	20.8	$C_{16}H_{16}O_8$	Caffeoyl-quinolactone isomer-1	335.0737	10	161.0217 179.0319 135.0422	8.27339	9.37878	7.51789	↑

#	RT	Molecular formula	Tentative identification	[M-H]⁻	Mass error (ppm)	Main MS/MS fragments	VIP values GCB vs LRC	GCB vs MRC	GCB vs DRC	Roasting trend
23	22.6		Unknown	553.2254		101.0229 119.0335 113.0227	3.59581	4.48496	4.86973	↓
24	23.2		Unknown	553.2262		507.2202 179.054	1.64501	2.12151	2.3711	↓
25	23.9	$C_{25}H_{24}O_{12}$	1.5-Dicaffeoylquinic acid*	515.1185	2	353.0754 173.0411 179.0305 191.0515	4.93467	6.55407	7.39515	↓
26	24.9		Unknown	319.1887		275.1927	1.10073	1.54124	1.92031	↑
27	25.3	$C_{26}H_{26}O_{12}$	Caffeoyl-feruloylquinic acid isomer-1	529.1351	0	173.0438 367.1022 193.0499 335.0749	0.84749	1.5228	1.81611	↓
28	26.5		Unknown	437.1446		101.06 275.1117	0.837543	1.30399	1.53752	↓
29	26.9		Unknown	727.3560		643.2947 113.0234 625.2838 565.2973	5.01431	4.80522	4.03683	↑
30	27.0	$C_{20}H_{18}N_2O_5$	N-caffeoyltryptophan	365.1141	0	135.0434 161.0221 229.0602 186.0543 203.0797	1.80289	1.50784	1.06821	↑
31	27.3		Unknown	437.1447		173.0441 275.1122	4.36024	5.2307	5.74657	↓
32	29.0		Unknown	565.3017		101.0596 113.0232 115.0758 463.2291	3.06446	2.88513	2.88513	—
33	29.5		Unknown	565.3020		113.0231 463.2342 101.059	2.88252	3.2899	3.5306	↑

* Confirmed with standard.
# Metabolite identification for this compound was performed considering an error of 31 ppm.
↑ The level of the compound increases with roasting.
↓ The level of the compound decreases with roasting.
— Indicates a different trend from the previously describe

Supplementary data for:

A NON-TARGETED METABOLOMIC APPROACH BASED ON REVERSED-PHASE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY TO EVALUATE COFFEE ROASTING PROCESS

Raquel Pérez-Míguez, Elena Sánchez-López, Merichel Plaza, María Castro-Puyana

María Luisa Marina*

Departamento de Química Analítica, Química Física e Ingeniería Química, Universidad de Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871 Alcalá de Henares (Madrid), Spain.

Instituto de Investigación Química "Andrés M. del Río" (IQAR), Universidad de Alcalá, Ctra. Madrid-Barcelona Km. 33.600, 28871 Alcalá de Henares (Madrid), Spain.

Contents:

compounds both in positive and negative ionization mode along the coffee roasting process.

Table S1.

SOLVENT	ESI +	ESI -
Methanol 25 %	1625	1523
Methanol 50 %	1555	1483
Methanol 70 %	1421	1252
Methanol 90 %	1614	1216
Ethanol 25 %	1221	1047
Ethanol 50 %	1025	975
Ethanol 70 %	994	1082
Ethanol 90 %	933	723
Water 100 %	1434	1336
Aqueous solution at pH 2.0	1329	1340
Aqueous solution at pH 9.0	1387	1305

Figure S1.

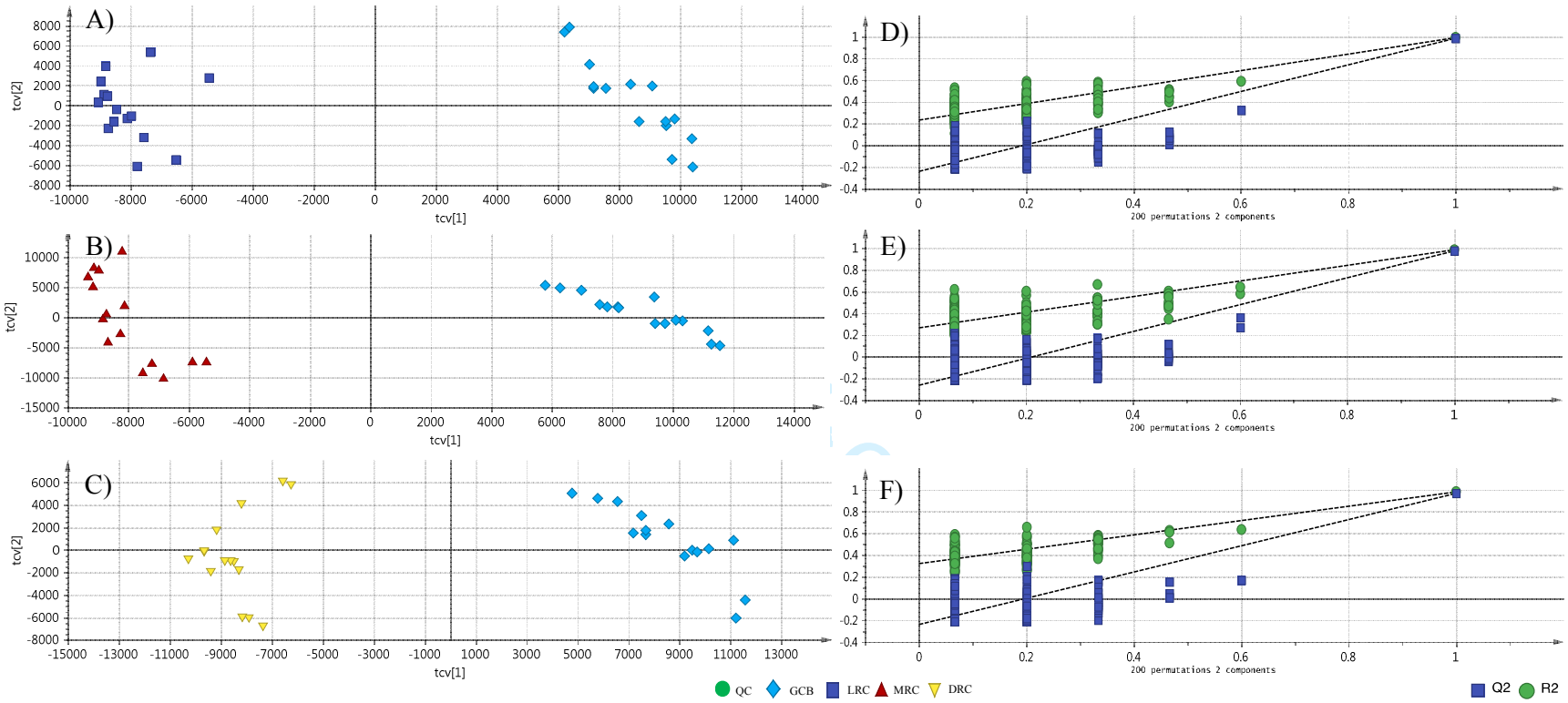

A)
B)
C)
D)
E)
F)
tcv[1]
tcv[2]
200 permutations 2 components
QC
GCB
LRC
MRC
DRC
Q2
R2

Figure S2

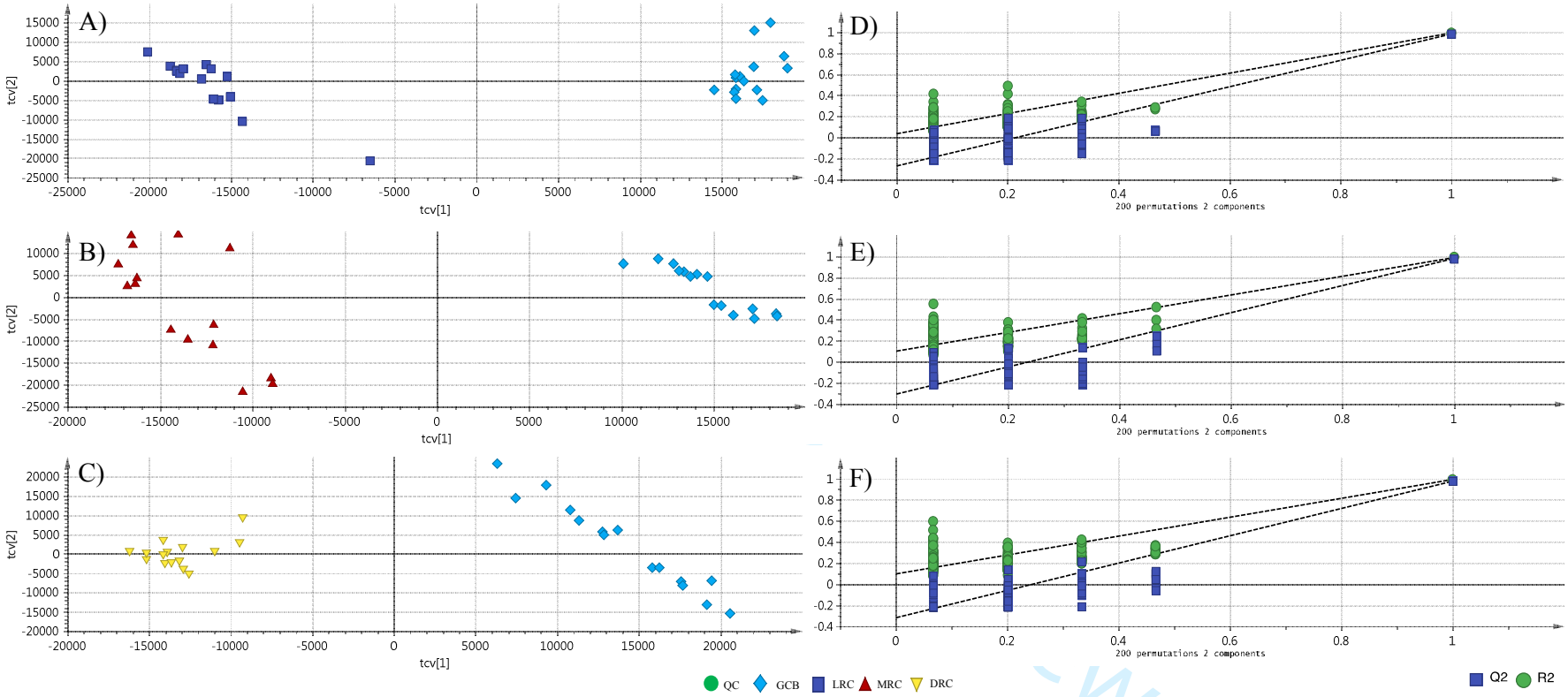

A)
B)
C)
D)
E)
F)
tcv[1]
tcv[2]
200 permutations 2 components
QC
GCB
LRC
MRC
DRC
Q2
R2

Figure S3

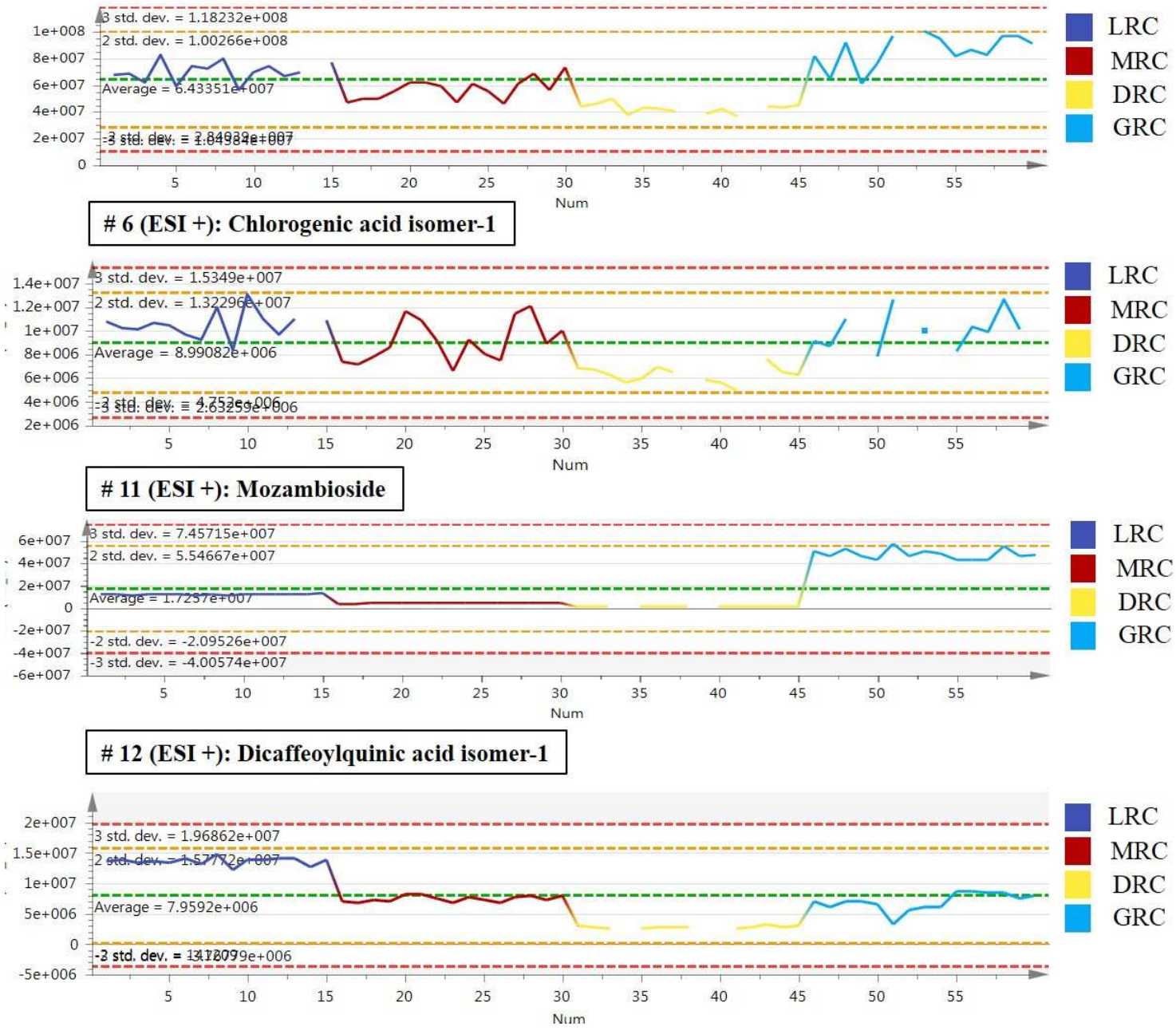

3 std. dev. = 1.18232e+008
2 std. dev. = 1.00266e+008
Average = 6.43351e+007
-2 std. dev. = 2.84939e+007
-3 std. dev. = 1.84584e+007
Num
LRC
MRC
DRC
GRC
6 (ESI +): Chlorogenic acid isomer-1
3 std. dev. = 1.5349e+007
2 std. dev. = 1.32296e+007
Average = 8.99082e+006
-2 std. dev. = 4.752e+006
-3 std. dev. = 2.63259e+006
Num
LRC
MRC
DRC
GRC
11 (ESI +): Mozambioside
3 std. dev. = 7.45715e+007
2 std. dev. = 5.54667e+007
Average = 1.7257e+007
-2 std. dev. = -2.09526e+007
-3 std. dev. = -4.00574e+007
Num
LRC
MRC
DRC
GRC
12 (ESI +): Dicaffeoylquinic acid isomer-1
3 std. dev. = 1.96862e+007
2 std. dev. = 1.57772e+007
Average = 7.9592e+006
-2 std. dev. = -3.41209e+006
Num
LRC
MRC
DRC
GRC

Figure S3 (cont.)

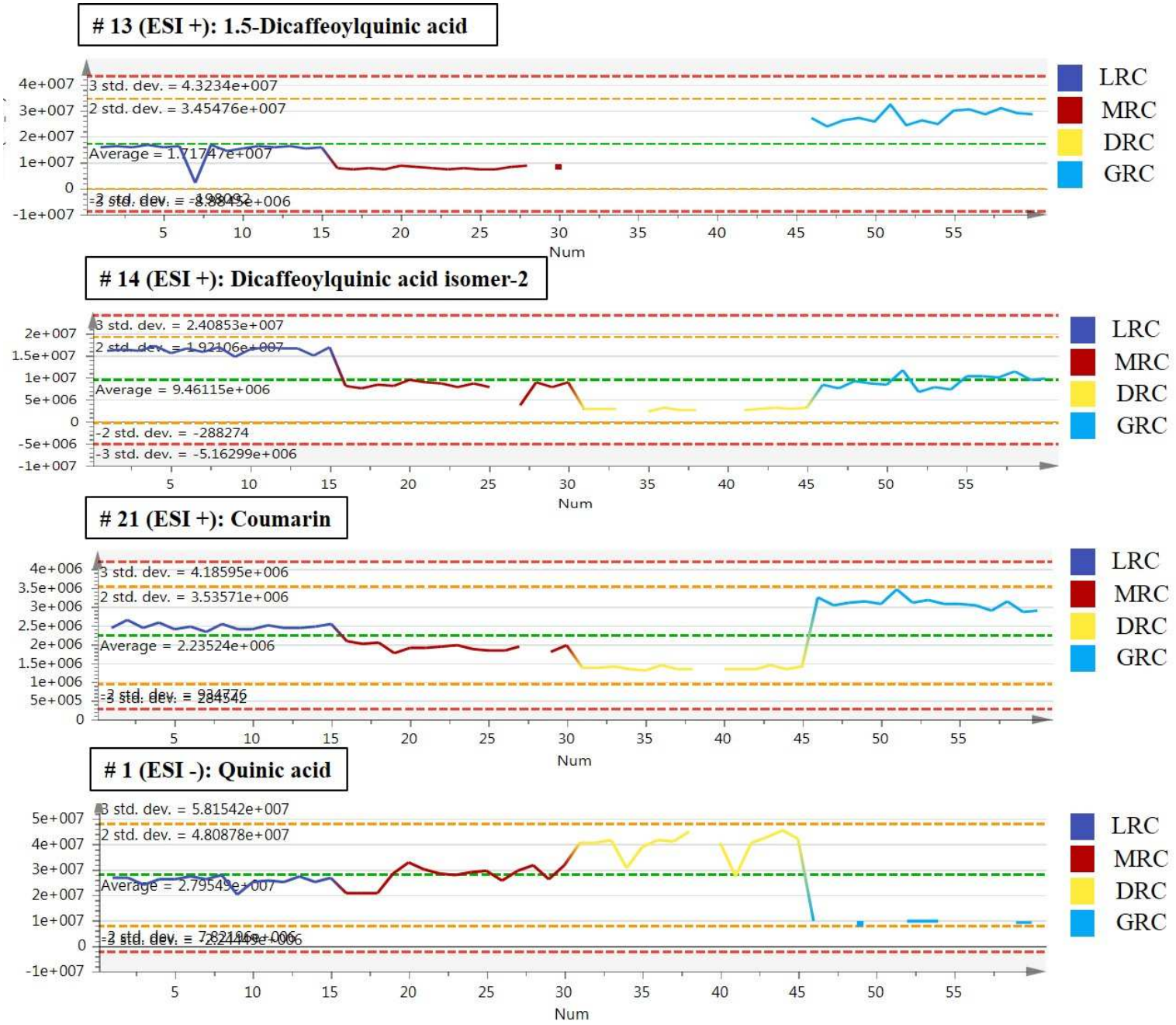


Figure S3 (cont.)

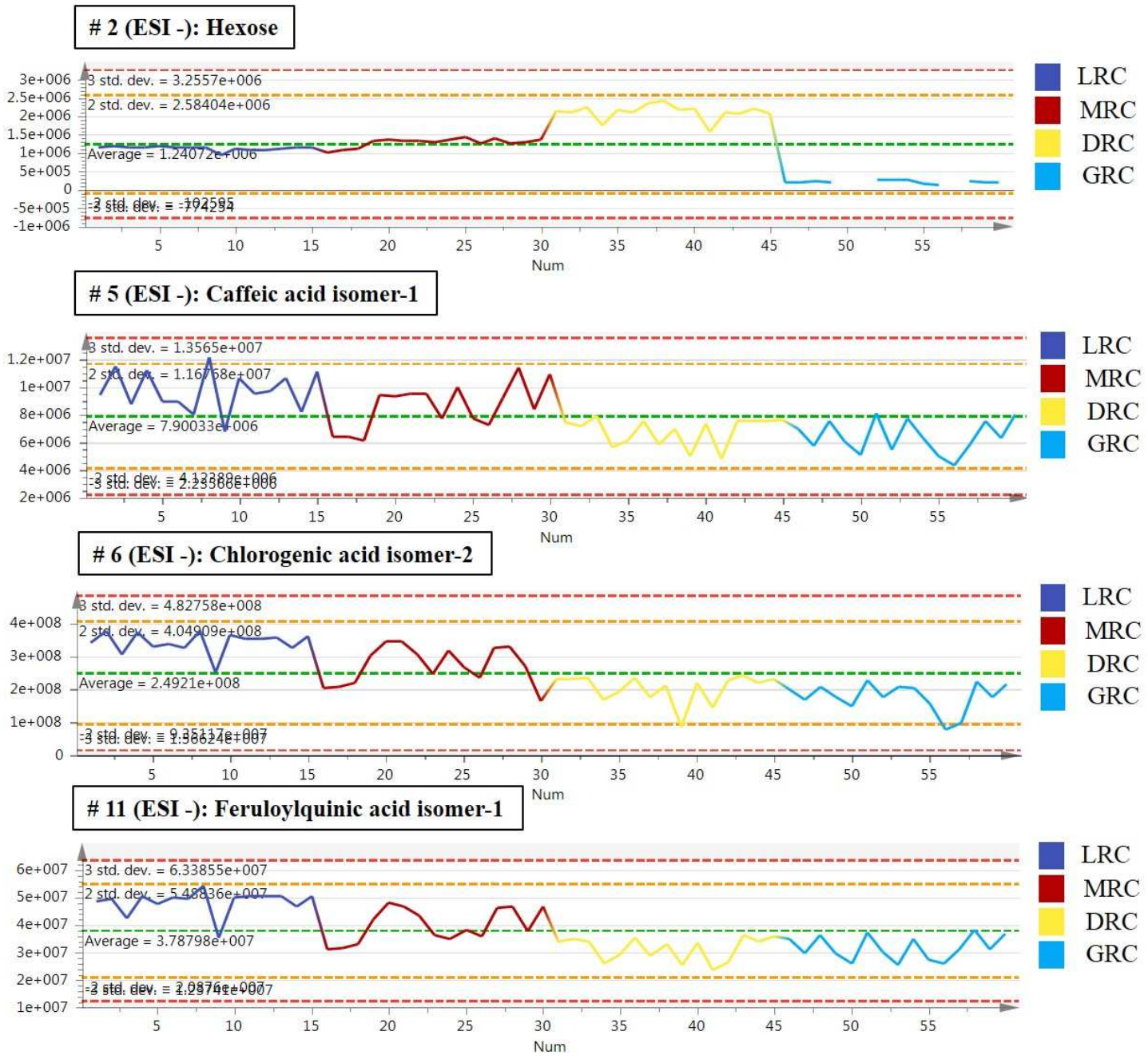


Figure S3 (cont.)

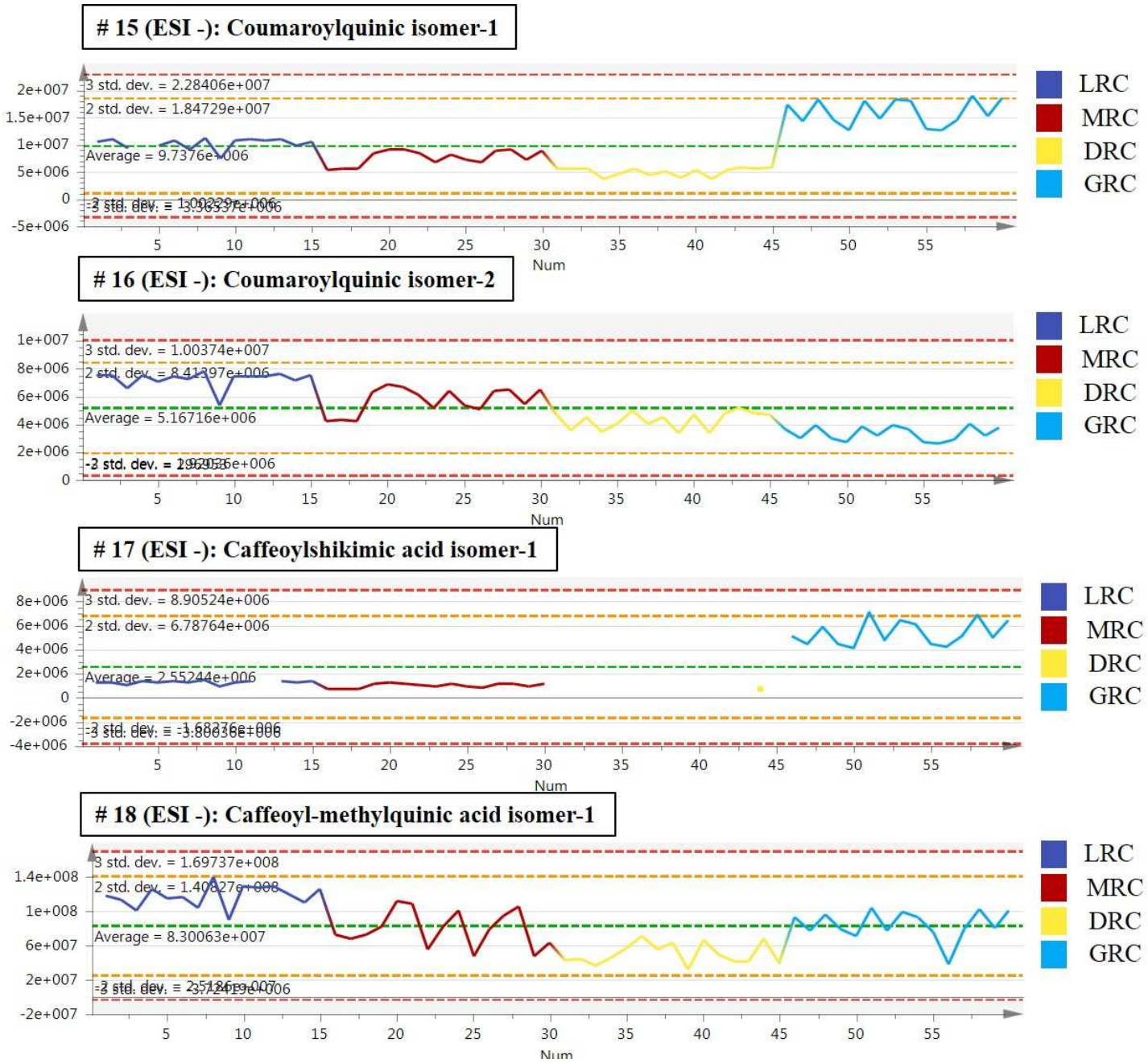
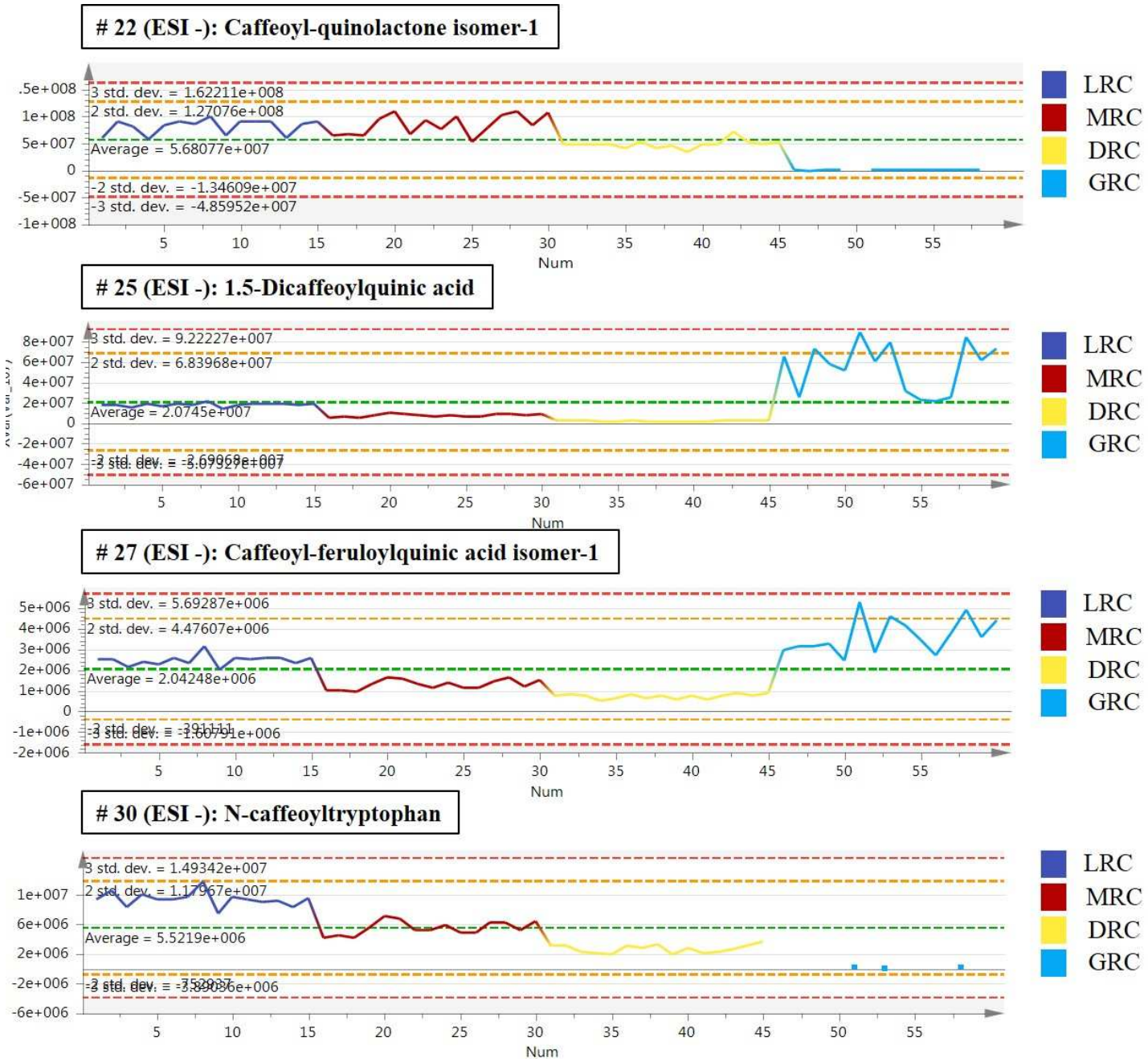


Figure S3 (cont.)



X