CAPILLARY ELECTROPHORESIS-MASS SPECTROMETRY METABOLIC FINGERPRINTING OF GREEN AND ROASTED COFFEE.

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

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

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

Correspondence: PhD. María Castro-Puyana, 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: maria.castrop@uah.es

Tel./Fax: 34-918856430/34-918854971

Abstract

The aim of this work was to develop a capillary electrophoresis-mass spectrometry (CE-ESI-QToF-MS) method to carry out the metabolic fingerprinting of green and roasted coffee samples (*Arabica* variety). To evaluate changes in the metabolic profiles of coffee occurring along the roasting process, green coffee beans were submitted to different roasting degrees. The effect of different parameters concerning the electrophoretic separation (background electrolyte, temperature, voltage, and injection time), the MS detection (temperature and flow of drying gas, sheath gas of jet stream temperature, and capillary, fragmentator, nozzle, skimmer, and octapole voltages) and the sheath liquid (composition and flow rate) was studied to achieve an adequate separation and to obtain the largest number of molecular features. The analyses were carried out in positive ESI mode allowing to detect highly polar cationic metabolites present in coffee beans. Non-supervised and supervised multivariate analyses were performed showing a good discrimination among the different coffee groups. Those features having a high variable importance in the projection values on supervised analyses were selected as significant metabolites for their identification. Thus, 13 compounds were proposed as potential markers of the coffee roasting process, being 7 of them tentatively identified and 2 of them unequivocally identified. Different families of compounds such as pyridines, pyrroles, betaines, or indoles could be pointed out as markers of the coffee roasting process.

Keywords: metabolic fingerprinting, capillary electrophoresis, high resolution mass spectrometry, jet stream, coffee beans, roasting process.

1. Introduction

Coffee production comprises different steps which affect not only its chemical composition but also its organoleptic properties [1]. Among these steps, coffee roasting is crucial for the development of coffee flavor since it involves several physical and chemical reactions that give rise to the formation and/or degradation of many components responsible for aroma, flavor or color. Some coffee components, such as lipids or caffeine, remain practically unaltered during the thermal process [2]. However, this process can lead to the formation of melanoidines due to the combination of sugars and amino acids during the Maillard reaction [3-9], pyridines because of protein hydrolysis and the degradation of trigonelline [10, 11], or even toxic compounds, such as furan derivatives, which have shown carcinogenic activity [12], among others. These changes in the chemical composition of coffee strongly affect its quality so that their evaluation has special importance to guarantee coffee quality both for the coffee industry and for consumers. In the industry, the control of roasting degree is usually carried out through differences in beans color, dry matter loss and/or other changes in sensory characteristics. Therefore, there is a need to develop adequate analytical methodologies enabling to face the discrimination of coffee beans submitted to different roasting degrees.

Different research works have already pointed out the importance to carry out a control of the chemical composition of coffee beans along roasting process [13-15]. Most of the works focused on this topic are based on target (detection of a single component) [16, 17] or profiling (detection of a specific class or components) analyses [18-21]. Thus, compounds such as amino acids, alkylpyrazines, chlorogenic acids or chlorogenic/caffeine ratio have been reported as markers to distinguish the roasting degree of coffee beans [22-24]. However, coffee beans, as well as other foods, are a

very complex matrix that presents hundreds of components so that its fingerprinting analysis (analysis of as many components as possible) during the roasting process will provide a maximum coverage of metabolites that can be simultaneously identified. In this sense, metabolomics, a well-established omics science focused on the study of the metabolome [25], is a powerful tool capable of providing an exhaustive characterization of complex samples which is becoming one of the most relevant procedures to assess food quality, safety and traceability [26, 27]. Up to date, the number of works concerning the metabolomic study of coffee is scarce. The discrimination of coffee varieties or origins [28-31] or between caffeinated and decaffeinated coffee [32] has been performed using as analytical platforms nuclear magnetic resonance (NMR), liquid chromatography-mass spectrometry (LC-MS) or gas chromatography-mass spectrometry (GC-MS). On the other hand, a few works have been focused on the evaluation of coffee roasting process using targeted metabolomic approaches based on ion mobility spectrometry-mass spectrometry (IMS-MS) [10], and employing non-targeted metabolomic approaches based on NMR, ambient sonic-spray ionization-mass spectrometry (EASI-MS) and GC-MS [3, 4, 33, 34]. Recently, our research group developed a non-targeted metabolomics strategy based on reversed-phase liquid chromatography-mass spectrometry (RPLC-MS) to provide the characterization of coffee beans roasted at three different degrees. It enabled to identify 7 and 13 metabolites as markers of roasting process in positive and negative modes, respectively [35].

Although NMR, GC-MS, and LC-MS are well established platforms for metabolomics, capillary electrophoresis (CE) coupled to MS is also a powerful analytical technique for metabolomics research due to its particular characteristics, i.e., it provides fast and efficient separations, requires low consumption of reagents and samples, and has a high

versatility considering its different modes [36]. Moreover, a large number of metabolites are polar and ionic, so they can be detected by CE-MS in contrast to LC-MS or GC-MS in which medium-polar, hydrophobic and volatile compounds are detected. However, even though several works have demonstrated the high potential of CE-MS in metabolomics studies [37-39], up to now, just a few works employed this technique to perform the metabolomics analysis of food matrices [36, 40-42].

In this work, an analytical method, based on the use of CE coupled to high resolution MS equipped with a Jet Stream thermal orthogonal electrospray ionization source, was developed in order to carry out the fingerprinting of green and roasted coffee and to evaluate changes in the metabolic profiles of coffee samples (*Arabica* variety) submitted to different roasting degrees. With this aim, different parameters concerning the electrophoretic separation and MS detection were optimized in order to maximize the number of detected peaks. In addition, the metabolic analysis, including data processing and chemometric analysis using PCA and PLS-DA models was applied to discriminate coffee beans according to their roasting degree. Finally, the identification of the significant metabolites along different roasting levels was performed.

2. Materials and methods

2.1 Chemicals and coffee samples

MS-grade methanol, acetic acid, and formic acid and HPLC-grade isopropanol were purchased from Fisher Scientific (Hampton, NH, USA). Ammonium formate and ammonium acetate of MS grade were from Sigma (St. Louis, MO, USA).

Calcium acetate, 3-ethylpyridine, 2-acetylpyrrole, 1-methyl-2-pyrrolecarboxaldehyde, L-(+)-arabinose, methyl anthranilate, indole-3-butyric acid, choline, and duloxetine were purchased from Sigma (St. Louis, MO, USA). Water employed to prepare the

running buffer and the coffee extracts was purified through a Milli-Q system from Millipore (Millipore, Madrid, Spain).

Arabica green coffee beans (GCB) were roasted and provided by the company "Café Fortaleza" (Vitoria, Spain). Coffee beans were roasted at three different levels: light (LRC), medium (MRC) and dark (DRC) using temperatures of 175ºC (during 12.36 min), 185ºC (for 14.11 min), and 195ºC (during 17.06 min), respectively. The roasting process was controlled in terms of the weight loss of each sample being 13% in LRC, 15% in MRC and 17% in DRC.

2.2 Preparation of coffee samples

The metabolite extraction procedure from coffee samples was performed using an extraction protocol previously optimized by our research group [35]. Briefly, 5 mg of grounded coffee samples were extracted with 1.5 mL of 25% (v/v) methanol in water. The solid-liquid extraction was carried out in a Thermomixer Compact (Eppendorf AG, Hamburg, Germany) at 700 rpm during 15 min at 25ºC. After centrifugation (3500 rpm, 10 min, 25 ºC) the supernatant fraction was collected and directly analyzed by CE-MS. Five replicate extractions for each group of coffee samples (GCB, LRC, MRC and DRC) were prepared and analyzed in the metabolomic sequence. Quality control samples (QC) were prepared by pooling equal aliquots of each coffee sample, which lets monitoring instrumental drifts throughout the analysis [43]. Duloxetine was used as internal standard (IS) at a final concentration of 10 μg/mL in all the analyzed samples including QCs.

2.3 CE-MS analysis

Metabolic fingerprinting of coffee samples was carried out using a 7100 CE system from Agilent Technologies (Waldbronn, Germany) coupled to a 6530 quadrupole time-of-flight mass spectrometer from Agilent Technologies (Waldbronn, Germany)

147 equipped with a Jet Stream thermal orthogonal electrospray ionization (ESI) source.
148 Coupling was performed via a sheath liquid interface with a CE-ESI co-axial sprayer
149 (G1607 model from Agilent Technologies). A sheath liquid composed of
150 methanol:water (50:50 v/v) containing 1 M acetic acid was delivered into the ESI
151 source at a 8 μL/min flow rate by means of a NE-3000 pump (New Era Pump Systems
152 Inc., Farmingdale, NY, USA). Sheath liquid also included two reference standards from
153 Agilent Technologies (purine (*m/z* 121.0508) and HP921 (*m/z* 922.0097)) to allow mass
154 accuracy monitoring. Agilent Mass Hunter Qualitative Analysis software (B.07.00) was
155 employed for MS control and data acquisition.
156 Separations took place in uncoated fused-silica capillaries of 50 μm ID with a total
157 length of 100 cm (Polymicro Technologies, Phoenix, AZ, USA) using a solution of 1 M
158 formic acid (pH 1.8) as background electrolyte (BGE). Before first use, new capillaries
159 were rinsed (applying 1 bar) with 1 M sodium hydroxide for 30 min, followed by 5 min
160 with Milli-Q water and conditioned with BGE for 60 min. Between injections, the
161 capillary was preconditioned with BGE for 5 min. Then, the samples were injected
162 applying a pressure of 50 mbar for 80 s. Finally, the electrophoretic separation was
163 achieved applying +30 kV at a working temperature of 20 ºC.
164 MS operate in positive ESI mode and data were acquired within 100-1700 *m/z* range
165 (extended dynamic range) in full scan resolution mode at a scan rate of 2 scans per s.
166 Capillary voltage was set to 3000 V with a nozzle voltage of 0 V, a nebulizer pressure
167 of 10 psi, and a sheath gas of jet stream at 3 L/min and 150 ºC. Drying gas was supplied
168 at 5 L/min at 180 ºC. The fragmentator voltage was set at 175 V whereas the skimmer
169 and octapole voltages were 60 V and 750 V, respectively.

MS/MS experiments were performed to assist in the metabolite identification. The voltage employed in MS/MS analyses ranged from 20 to 40 V depending on each analyte.

2.4 Metabolomic sequence

Blank and QC samples were injected in the CE-MS system at the beginning of the metabolomic sequence to ensure a good repeatability. Then, a total of 60 coffee samples (five replicates of each group (GCB, LRC, MRC, and DRC) injected in triplicate) were randomly injected and a QC sample was injected every six coffee samples and at the end of the sequence.

2.5 Data processing and analysis

Molecular features from the raw data were obtained using the Molecular Feature Extraction (MFE) tool from Mass Hunter Qualitative Analysis (B.07.00 from Agilent Technologies) where the migration time and abundance of the molecular features were annotated. The MFE parameters were as follows: “small molecules (chromatographic)” mode; peaks with height ≥ 500 counts; peak spacing tolerance for isotope grouping was 0.0025 m/z plus 7.0 ppm; isotope model = common organic molecules; and the charge states were limited to 2. Moreover, to identify different ion species coming from the same molecular feature, H^+, Na^+, K^+, and NH_4^+ adducts were considered.

Migration time correction and alignment of molecular features were conducted using the Mass Profiler Professional (MPP) software (B.02.00 from Agilent). To carry out the alignment, a migration time window of 0.3 min, with a mass window of 30 ppm + 0.02 Da, were employed. Then, data were normalized using the intensity of the IS. To remove non-reproducible signals before performing statistical analysis, molecular features were filtered by retaining masses present in at least 80 % of all injected QC samples and with a coefficient of variation below 35%.

Multivariate statistical analysis of the data matrix containing the filtered molecular features was performed using SIMCA 14.0 software (Umetrics, Umeå, Sweden). After log-transformation (to approximate to a normal distribution) and Pareto scaling (for reducing the relative importance of larger values) [44] both principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA) models were used to investigate clustering existing in the analyzed samples and to find differences between samples according to their roasted degree. Quality of the models was assessed by the R^2 (R^2X, R^2Y) and Q^2 values. Potential biomarkers of roasting degree of coffee were found by two-class comparisons: GCB vs LRC, GCB vs MRC, and GCB vs DRC. Only features with variable importance in the projection (VIP) values of the first component of the PLS-DA models higher than 1.0 were considered as significant. Moreover, univariate statistical analysis using the Mann-Whitney U test was performed in R (http://www.R-project.org). Benjamini-Hochberg false discovery rate was employed for multiple testing correction.

2.6 Metabolite identification

Potential markers of coffee roasting process were identified by matching the experimental accurate mass values with theoretical mass values available in the CEU Mass Mediator database [45] considering an error of 30 ppm. This database, available online, allows to obtain information simultaneously from different databases such as KEGG, METLIN, HMDB and LipidMaps. In addition, the FooDB database (http://foodb.ca/) was also employed. Those metabolites that were commercially available as standards compounds were also analyzed by the developed CE-MS methodology to perform their unequivocal identification according to their migration time and MS/MS fragmentation pattern. When the standards could not be acquired, a tentative identification was carried out comparing the experimental MS/MS spectra

obtained for each molecular feature with those predicted in HMDB database, CFM-ID (cfmid.wishartlab.com) and/or literature.

3. Results and discussion

3.1 Development of a CE-MS method for metabolic fingerprinting

In order to develop a CE-MS methodology enabling the fingerprinting of green and roasted coffee and to study the changes occurring in the metabolic profiles of coffee samples submitted to different roasting degrees, those parameters related to the CE-MS coupling (composition and flow rate of the sheath liquid), the electrophoretic separation (BGE, temperature, voltage, and injection time) and the MS detection were optimized using as model a QC sample (injected in duplicate). The selection of each parameter was based on the number of molecular features extracted by the MFE tool (see section 2.5). It should be mentioned that only ESI in positive mode was evaluated as ionization mode in this work. This was due to the fact that up to date, most of the metabolomics studies based on the use of CE-MS as analytical platform have employed this ionization mode [46, 47], since using a fused silica capillary and normal polarity, the electroosmotic flow (EOF) moves from the anode to the cathode (where the MS detector is located), and the CE electrical current is stabilized. In addition, both EOF and the electrophoretic mobility of cationic metabolites are toward the MS detector what enabled to achieve rapid analysis with good resolution [40].

The effect of the composition and flow rate of the sheath liquid was investigated because they are critical variables not only to establish an adequate electric contact between the CE system and the mass spectrometer but also to obtain a good metabolite ionization. Thus, using 1 M formic acid (pH 1.8) as BGE and a methanol:water (50:50 v/v) sheath liquid as initial conditions, different additives, such as 1 M formic acid, 1 M

245 acetic acid, 5 mM ammonium formate and 5 mM ammonium acetate were added to the
246 sheath liquid to study their influence on the ionization. Among these additives, 1 M
247 acetic acid was chosen for further experiments since its addition to the sheath liquid
248 enabled to obtain the highest number of molecular features. Contrary to what it would
249 be expected, the use of acetic acid in the sheath liquid and formic acid in the running
250 buffer allowed to detect the maximum number of molecular features. The increment in
251 molecular features was truly due to the appearance of new compounds and not to
252 artifacts derived from the combination of both buffers.
253 The next step was to evaluate the influence of the organic solvent present in the sheath
254 liquid. Mixtures of methanol or isopropanol with water at 50:50 (v/v) containing 1 M
255 acetic acid were tested. Since the use of methanol allowed to obtain a greater number of
256 molecular features and a higher current stability, it was selected as organic solvent.
257 Also, different sheath liquid compositions (methanol:water 50:50, 70:30 and 80:20
258 (v/v)) were compared, selecting methanol:water 50:50 (v/v) due to the higher number of
259 molecular features observed with this mixture. Finally, the flow rate was optimized
260 using 4, 6, and 8 μL/min. It was necessary to employ the highest flow (i.e. 8 μL/min) in
261 order to enhance the ionization which also enabled obtaining the maximum number of
262 molecular features.
263 Once the best conditions for the sheath liquid were selected, the next step was to
264 optimize the variables affecting the electrophoretic profile. The nature of the BGE (1 M
265 formic acid or 1 M acetic acid), the working temperature (15, 20 or 25 ºC), the applied
266 voltage (20, 25 or 30 kV) and the injection time (10, 20, 80 and 120 s) were evaluated in
267 terms of electrophoretic separation, peak efficiency, and sensitivity to ensure the
268 detection of the largest number of molecular features in the coffee samples. The

optimized conditions were 1 M formic acid (pH 1.8), a working temperature of 20 ºC, a separation voltage of 30 kV, and an injection time of 80 s.

Finally, the effect of different ESI parameters, such as capillary voltage (2000-4000 V), fragmentator voltage (110-175 V), nozzle voltage (0-100 V), skimmer (50-60 V), octapole (160 or 750 V), drying gas temperature (150-300 °C), drying gas flow (5-10 L/min) and sheath gas of jet stream temperature (150-300 °C), was evaluated. The most adequate values to achieve the largest number of molecular features with the highest sensitivity were 3000 V for capillary voltage, 175 and 0 V for fragmentator and nozzle voltages, respectively, 60 V for skimmer, 750 V for octapole, 180 ºC and 5 L/min for drying gas, and 150 ºC for sheath gas temperature.

Figure 1 shows the metabolic profiles of GCB, LRC, MRC and DRC samples analyzed by the developed CE-MS methodology.

3.2 Metabolic fingerprinting of coffee samples by CE-MS and potential markers of roasting process

The CE-MS method developed was applied to the metabolic fingerprinting of coffee samples submitted to different roasting levels to evaluate the potential of this technique for differentiating molecules related to the roasting process. For this purpose, a total of 60 samples (five replicates of each group (GCB, LRC, MRC, and DRC) injected in triplicate) and different injections of a QC sample distributed across the metabolomics sequence (see section 2.4) were analyzed. 1275 different molecular features were found which demonstrated the potential of CE-MS for high-throughput metabolomics analysis.

It should be highlighted that data processing must be carefully performed since the algorithms used for peak fitting of some tools have been developed for LC purposes and

294 the scores should be adjusted because of the slight different shape of the CE peaks
295 compared to LC signals and their width [48]. In fact, the molecular features alignment
296 in CE-MS is normally more problematic than in LC-MS due to the higher migration
297 time shift.
298 After migration time correction, alignment, normalization, and filtering (see section
299 2.5), the resulting dataset comprised 39 time-aligned metabolic features. The reason
300 behind the reduction of molecular features after filtering (from 1275 to 39) is because in
301 CE the migration time variability is much larger than in any chromatographic technique
302 and that artificially increases the number of features which are actually the same
303 features repeated across samples. For this reason, the preprocessing steps such as
304 alignment and normalization are crucial in CE. It should be mentioned that **Figure 1**
305 shows a metabolomic profile more in line with the number of molecular features
306 obtained (around 39).
307 Then, logarithmic transformation and Pareto scaling were used to approximate a normal
308 distribution and for reducing the relative importance of larger values, respectively [44].
309 Regarding data analysis, the data matrix was first subjected to PCA not only to evaluate
310 the consistency of the metabolomics sequence but also to observe the variability
311 existing in the dataset. As it can be seen in **Figure 2**, PCA clearly showed not only
312 differences among the four groups of samples analyzed but also it demonstrated the
313 consistence of the analytical sequence since QC samples were clustered in the center of
314 the plot showing the low analytical variability existing between runs. Note that, high
315 percentages of variability were explained for the first two components (66 and 15 % for
316 the first and second component, respectively). The first principal component is
317 responsible for the separation of the gradual roasting process (from light to dark coffee)
318 whereas the second one differentiates mainly GCB from LRC. This fact could be

explained because the differences between raw and light roasting coffee beans are milder; however, the differences between GCB, MRC and DRC appear not only in the first component but also in the second one, highlighting a more different phenotype than with the LRC. Loading plots of the PCA model are shown in **Figure S1.**

To carry out the sample classification and to simplify the search of those variables that are potential markers of coffee roasting level, PLS-DA models were performed. **Figure 3** shows the PLS-DA models for different two-class comparison (GCB vs LRC, GCB vs MRC, and GCB vs DRC). In addition, **Table 1** shows R^2X, R^2Y and Q^2 quality parameters as well as F and p-values of the cross validated ANOVA for all PLS-DA models. In all cases, high Q^2 values (> 0.919) were achieved and the values obtained in the cross validated ANOVA (F values higher than 58.2 and p-values lower than 5.2×10^{-12}) demonstrated the robustness of the proposed models and a good classification existing between groups. Moreover, the results obtained in the permutation tests (based on 200 permutations), employed to validate all PLS-DA models, indicated that differences in the PLS-DA were indeed due to differences in the metabolic profile of coffee samples and not due to data overfitting (see **Figure 3**) [49].

Next, once demonstrated the differences in the metabolic profiles of the coffee samples submitted to different roasting process, the variable importance in projection (VIP) value was selected to point out potential markers since it summarizes the contribution of each variable to the PLS model (loading plots of the three PLS-DA models are shown in **Figure S2**). Thus, 13 variables with VIP values higher than 1.0 were chosen as potential relevant molecular features in coffee roasting process. The extracted ion electropherograms (EIEs) of the 13 significant variables are represented in **Figure 4**.

3.3 Metabolite identification

The identification of the 13 molecular features highlighted as markers of the coffee roasting process by the VIP values was performed using their *m/z* value, isotopic pattern, and MS/MS fragmentation pattern as it has been described in section 2.7. Using different databases, a list of possible metabolites was proposed. To filter them, the first step was to remove compounds such as drugs or compounds of biological relevance that are not present in vegetables matrices, keeping only these compounds whose presence could be probable in coffee, for further interpretation. Then, the isotopic profile and the fragmentation pattern (MS/MS spectra) were carefully evaluated. **Table 2** summarizes the migration time, the molecular formula, the experimental *m/z* value, the mass error (ppm), the main fragments obtained in MS/MS spectra, and statistical parameters, i.e. VIP values from PLS-DA models and the p-values (in brackets) from non-parametric univariate Mann-Whitney U test for each of the three pairwise comparisons. Also, the trend observed for all significant metabolites along the roasting process of coffee is included in this table. As **Table 2** shows, 7 metabolites were identified using this approach (2 of them were unequivocally identified). Interestingly, the levels of most compounds decreased with the roasting process except the compounds 3, 5, 6, 8 and 12 whose levels increased along this process.

Compound 2 (t_m = 8.4 min) exhibited a $[M+H]^+$ ion at *m/z* 143 with an intense MS/MS fragment ion at *m/z* 82. Looking to the isotopic profile of this compound hints the presence of a sulphur atom. Moreover, by comparing the MS/MS spectrum obtained to the predicted one reported in the HMDB database, this compound was tentatively identified as S-(2-furanylmethyl)methanethioate. Even though the presence of this compound in roasted coffee has been previously reported in the literature [50], as far as we know, its experimental MS/MS spectrum has not been yet reported.

From the different molecular features selected as markers of the coffee roasting process in some cases, identification could be corroborated by co-injection with commercial standards. For instance, **compound 3** (t_m = 9.9 min) with a $[M+H]^+$ ion at *m/z* 108 and **compound 4** (t_m = 10.1 min) with a $[M]^+$ ion at *m/z* 104 were unequivocally identified as 3-ethylpyridine and choline, respectively, based on the comparison of their migration times and MS/MS fragmentation patterns to those obtained for the commercial standards. The presence of both compounds in roasted coffee has been widely reported in some works [51-55]. As can be seen in the box plot (see **Figure 5),** 3-ethylpyridine (**compound 3)** is one of the compounds whose levels increased along roasting process which is in agreement with the results previously reported by Dorfner et al. [56]. This provides a way to validate our metabolomic approach herein developed. Regarding choline, Wei et al. observed that this compound slightly decreased during the roasting process [57].

Compounds 5 (t_m = 10.4 min) and **8** (t_m = 11.1 min) displayed $[M+H]^+$ ions at *m/z* 110 and *m/z* 124, respectively. Both compounds were tentatively identified as pyrrole derivatives. Standard solution of 1-methyl-2-pyrrolecarboxaldehyde and 2-acetylpyrrole were injected to try to identify **compound 5**. Although, the migration time of these standards did not match the one obtained for **compound 5**, the fragment ions obtained from the MS/MS spectrum of 1-methyl-2-pyrrolecarboxaldehyde matched those obtained for **compound 5** (ions *m/z* 67 and 80), suggesting this compound could be a derivative of methyl-pyrrolecarboxaldehyde. On the other hand, **compound 8** was tentatively identified as N-acetyl-2-methylpyrrole since it showed fragment ions at *m/z* 80, 81 and 53, which is in agreement with previous results described in the literature [58]. Several works have reported the role of pyrrole compounds in the aroma and flavour of coffee [58-60]. An increment of levels of these two pyrrole derivatives during

393 roasting process was observed in the box plots (**Figure 5**). Another metabolite
394 tentatively identified was 3-dehydroxycarnitine (**compound 9**, t_m = 11.2 min), having a
395 $[M+H]^+$ ion at *m/z* 146. The two product ions from MS/MS spectrum (*m/z* 43 and 87)
396 matched the predicted one obtained in the HMDB database.
397 The last compound tentatively identified was **compound 12** (t_m = 12.7 min) that
398 displayed a $[M+H]^+$ ion at *m/z* 204. A standard solution of indole-3-butyric acid (t_m =
399 22.5 min) was analysed in order to obtain its migration time and MS/MS pattern to
400 compare with the ones obtained for **compound 12**. Although the migration times did
401 not match, the MS/MS pattern shows similar fragments for both compounds at *m/z* 158,
402 144, and 186 suggesting that **compound 12** could be an indole derivative. In fact,
403 despite being very scarce the available information about the formation pathway and
404 kinetics of indole during coffee roasting, the presence of this family compounds has
405 been reported in coffee. Silwar and Lüllman [61] measured the formation of indole
406 during coffee roasting and reported that it appears in considerable amounts at 170 °C
407 and increases with roasting time. This corroborates our findings once again.
408 Other standard compounds such as calcium acetate (t_m = 9.1 min), methyl anthranilate
409 (t_m = 13.4 min) and L-(+)-arabinose were analysed to know their migration times and
410 MS/MS spectra with the aim of comparing them with those obtained for **compounds 1**
411 and **7** (**see Table 2)**. However, the possibility of matching calcium acetate or methyl
412 anthranilate with one of our metabolites was discarded due to differences in the
413 migration time and MS/MS fragmentation pattern. On the other hand, the ionization of
414 L-(+)-arabinose was not possible under the given CE-MS conditions.
415 Unfortunately, none of the **compounds 1, 6, 7, 10, 11** and **13** matched the experimental
416 MS/MS fragmentation pattern nor the migration times of the available standards.
417 Comparison with theoretical MS/MS fragmentation pattern found in the literature also

418 did not result in a match. This all made their identification unfruitful and, thus, their
419 identity as "unknowns" in **Table 2**.
420 Thus, in the present study, several markers of the coffee roasting process have been
421 identified. Some of these metabolites have already been proven to be related to the
422 roasting process. This validates our approach and help us demonstrating the potential of
423 CE-MS in the metabolomic analysis of coffee samples.
424 Bearing in mind the results obtained in this work and those previously described by our
425 research group in which a RPLC-MS non-targeted metabolomics strategy was applied
426 to analyze coffee beans extracts submitted to different roasting degrees [35], it is clear
427 the importance of combining orthogonal analytical platforms to achieve a broad
428 metabolite coverage to provide a better knowledge of the changes occurring during the
429 roasting process of green coffee beans. Several metabolites showed statistically
430 significant differences among the coffee bean samples analyzed by each analytical
431 platforms. Amino acids, betaines and products formed during Maillard reaction were
432 tentatively identified by CE-MS, whereas RPLC-MS mainly enabled the tentatively
433 identification of compounds belonging to the hydroxycinnamic acids family. None of
434 these metabolites were common in both analytical platforms what demonstrates the
435 complementarity of these techniques and the relevance of their combined use to study
436 the full metabolome. In fact, RPLC enables to separate from medium-polar to non-polar
437 compounds whereas CE is a powerful tool for the analysis of highly polar and ionized
438 compounds (which cannot be determined by RPLC).
439
440 **4. Conclusions**
441 In this work, a new CE-MS method has been developed enabling the metabolic
442 fingerprinting of coffee samples submitted to different roasting degrees. This advanced

analytical methodology enables a reliable comparison of metabolic profiles in which peak alignment was successfully carried out using vendor software. The feasibility of the methodology was demonstrated by the analysis of coffee samples submitted to three different roasting degrees (light, medium, and dark coffee) in order to investigate changes occurring during this process. This approach allowed to propose 13 compounds as potential markers of the coffee roasting process. 7 of these compounds could be identified, being 2 of them unequivocally identified. Different families of compounds such as derivatives from pyridine, pyrrole, betaine, or indole have been pointed out as markers of the coffee roasting process. In summary, the developed CE-MS methodology is presented as a useful and powerful strategy to obtain information on the polar metabolome, being highly complementary to other previously used in metabolomics.

Acknowledgements

Authors thank financial support from the Comunidad of Madrid (Spain) and FSE and FEDER programs (project S2018/BAA-4393, AVANSECAL-II-CM), and the University of Alcalá (project CCG2015/EXP-032). M.C.P. thanks the Spanish Ministry of Economy and Competitiveness (MINECO) for their "Ramón y Cajal" contract (RYC-2013-12688). M.P. thanks the same Ministry (MINECO) and the University of Alcalá for her "Juan de la Cierva" (IJCI-2014-22143) and postdoctoral research contracts, respectively. R.P.M. thanks the University of Alcalá for her pre-doctoral contract. Authors gratefully acknowledge "Café Fortaleza" for providing coffee samples.

References

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

[2] I. S. Mussatto, E.M.S. Machado, S. Martins, J.A. Teixeira, Production, composition, and application of coffee and its industrial residues, Food Bioprocess Technol. 4 (2011) 661-672.

[3] M.R. Ruosi, C. Cordero, C. Cagliero, P. Rubiolo, C. Bicchi, B. Sgorbini, E. Liberto, A further tool to monitor the coffee roasting process: aroma composition and chemical indices, J. Agric. Food Chem. 60 (2012) 11283-11291.

[4] E.Liberto, M.R. Ruosi, C. Cordero, P. Rubiolo, C. Bicchi, B. Sgorbini, Non-separative headspace solid phase microextraction-mass spectrometry profile as a marker to monitor coffee roasting degree, J. Agric. Food Chem. 61 (2013) 1652-1660.

[5] I. Flament, Coffee, cacao and tea, Food Rev. Int. 5 (1989) 317-414.

[6] I. Flament, I. Coffee, cacao and tea, in: H. Maarse (Ed), Volatile Compounds in Foods and BeVerages, Dekker, New York, 1991, pp. 617-669.

[7] R. Tressl, Formation of flavor components in roasted coffee, in: T.H. Parliament, R. J. McGorrin, C.T. Ho (Eds), Thermal generation of aromas, American Chemical Society, Washington, DC, 1989, pp 285-301.

[8] W. Holscher, H. Steinhart, Formation pathways for primary roasted coffee aroma compounds, in: T.H. Parliament, M. Morello, R. J. McGorrin (Eds), Thermally generated flavors, American Chemical Society, Washington, DC, 1992, pp 207-217.

[9] C.T. Ho, H.I. Hwang, T.H. Yu, J. Zhang, An overview of the Maillard reactions related to aroma generation in coffee, in: Proceedings of the 15th ASIC Colloquium (Montpellier), ASIC, Paris, France, 1993, pp 519-527.

[10] A.N. Gloess, C. Yeretziana, R. Knochenmuss, M. Groessl, On-line analysis of coffee roasting with ion mobility spectrometry-mass spectrometry (IMS–MS), Int. J. Mass Spec. 424 (2018) 49-57.

[11] C.A.B. De Maria, L.C. Trugo, F.R. Aquino Neto, R.F.A. Moreira, C.S. Alviano, Composition of green coffee water-soluble fractions and identification of volatiles formed during roasting. Food Chem. 55 (1996) 203-207.

[12] C. Crews, L. Castle, A review of the occurence, formation and analysis of furan in heat-processes foods, Trends Food Sci. Technol. 18 (2007) 365-372.

[13] Ivon Flament, Coffee, cocoa, and tea, Food Rev. Int. 5 (1989) 317-414

[14] R. Tressl, Formation of flavor components in roasted coffee, in: T Parliament et al., Thermal Generation of Aromas, ACS Symposium Series, American Chemical Society: Washington, DC, 1989.

[15] A. Ciampa, G. Renzi, A. Taglienti, P. Sequi, M. Valentini, Studies on coffee roasting process by means of nuclear magnetic resonance spectroscopy, Journal of Food Quality 33 (2010) 199-211.

[16] J.R. Santos, O. Viegas, R.N.M.J. Páscoa, I.M.P.L.V.O. Ferreira, A.O.S.S. Rangel, J. A. Lopes, In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and color, Food Chem. 208 (2016) 103-110.

[17] S. Casal, M.B.P.P. Oliveira, M.R. Alves, M.A. Ferreira, Discriminate analysis of roasted coffee varieties for trigonelline, nicotinic acid, and caffeine Content, J. Agric. Food Chem. 48 (2008) 3420-24.

512 [18] B. Sanchez-Bridge, M. Renouf, J. Sauser, M. Beaumont, L. Actis-Goretta, The
513 roasting process does not influence the extent of conjugation of coffee chlorogenic and
514 phenolic acids, BioFactors, https://doi.org/10.1002/biof.1268.
515 [19] R. Garrett, B.G. Vaz, A.M.C. Hovell, M.N. Eberlin, C.M. Rezende, Arabica and
516 robusta coffees: Identification of major polar compounds and quantification of blends
517 by direct-infusion electrospray ionization–mass spectrometry, J. Agric. Food Chem. 60
518 (2012) 4253-58.
519 [20] B.R. Toledo, L.W. Hantao, T.D. Ho, F. Augusto, J.L. Anderson, A chemometric
520 approach toward the detection and quantification of coffee adulteration by solid-phase
521 microextraction using polymeric ionic liquid sorbent coatings, J. Chromatogr. A. 1346
522 (2014) 1-7.
523 [21] A.L. Dawidowicz, R. Typek, Transformation of chlorogenic acids during the coffee
524 beans roasting process, Eur. Food Res. Technol. 243 (2017) 379–390.
525 [22] U.P. Nerhing, H.G. Maier, Indirect determination of the degree of roast in coffee,
526 Z. Lebensm.-Unters. Forsch. 195 (1992) 39−42.
527 [23] L. Hashim, H. Chaveron, Use of methyl pyrazine ratios to monitor the coffee
528 roasting. Food Res. Int. 28 (1995) 619-623.
529 [24] M.P. Purdon, D.A. McCamey, Use of a 5-caffeoylquinic acid/ caffeine ratio to
530 monitor the coffee roasting process, J. Food Sci. 52 (1987) 1680-1683.
531 [25] E. Trujillo, C. Davis, J. Milner, Nutrigenomics, proteomics, metabolomics, and the
532 practice of dietetics, J. Am. Diet. Assoc. 106 (2006) 403–413.
533 [26] M. Castro-Puyana, M. Herrero, Metabolomics approaches based on mass
534 spectrometry for food safety, quality and traceability, TrAC 52 (2013) 74–87.

535 [27] M. Castro-Puyana, R. Pérez-Míguez, L. Montero, M. Herrero, Application of mass
536 spectrometry-based metabolomics approaches for food safety, quality and traceability,
537 TrAC 93 (2017) 102e118.
538 [28] R. Garrett, E.M. Schmidt, L.F.P. Pereira, C.S.G. Kitzberger, M.B.S. Scholz,
539 M.N.E. Berlin, C.M. Rezende, Discrimination of Arabica Coffee cultivars by
540 electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry
541 and chemometrics, Food Sci. Technol. 50 (2013) 496-502.
542 [29] D. Setoyama, K. Iwasa, H. Seta, H. Shimizu, Y. Fujimura, D. Miura, H. Wariishi,
543 C. Nagai, K. Nakahara, High-throughput metabolic profiling of diverse green coffea
544 arabica beans identified tryptophan as a universal discrimination factor for immature
545 beans. PLOS ONE, 9 (2013) e70098.
546 [30] M.Y. Choi, W. Choi, J.H. Park, J. Lim, S.W. Kwon, Determination of coffee
547 origins by integrated metabolomic approach of combining multiple analytical data,
548 Food Chem. 121 (2010) 1260-68.
549 [31] J.H. Da Silva Taveira, F. Meira Borém, L. Pereira Figueiredo, N. Reis, A.S.
550 Franca, S.A. Harding, C.J. Tsai, Potential markers of coffee genotypes grown in
551 different Brazilian regions: A metabolomics approach, Food Res. Int. 61 (2014) 75-82.
552 [32] K.L. Chang, P.C. Ho PC, Gas chromatography time-of-flight mass spectrometry
553 (GC-TOF-MS)-based metabolomics for comparison of caffeinated and decaffeinated
554 coffee and its implications for Alzheimer's disease, PLOS ONE, 9 (2014) e104621.
555 [33] F. Wei, K. Furihata, F. Hu, T. Miyakawa, M. Tanokura, Two-dimensional 1H–13C
556 nuclear magnetic resonance (NMR)-based comprehensive analysis of roasted coffee
557 bean extract, J. Agric. Food Chem. 59 (2011) 9065-73.
558 [34] J. Santos da Rosa, O. Freitas-Silva, J.R. Costa Rouws, I. Gonçalves da Silva
559 Moreira, F.J. Moreira Novaes, D. De Almeida Azevedo, N. Schwab, R. Luiz de Oliveira

560 Godoy, M. Nogueira Eberlin, C. Moraes de Rezende, Mass spectrometry screening of
561 Arabica coffee roasting: A non-target and non-volatile approach by EASI-MS and ESI-
562 MS, Food Res Int. 89 (2016) 967-75.
563 [35] R. Pérez-Míguez, E. Sánchez-López, M. Plaza, M. Castro-Puyana, M.L. Marina, A
564 non-targeted metabolomic approach based on reversed-phase liquid chromatography–
565 mass spectrometry to evaluate coffee roasting process, Anal. Bioanal. Chem. 410 (2018)
566 7859-7870.
567 [36] C. Ibáñez, C. Simó, V. García-Cañas, A. Cifuentes, M. Castro-Puyana,
568 Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-
569 mass spectrometry in Foodomics: A review, Anal. Chim. Acta 802 (2013) 1–13.
570 [37] C. Barbas, E.P. Moraes, A.Villaseñor, Capillary electrophoresis as a metabolomics
571 tool for non-targeted fingerprinting of biological samples, J. Pharmaceut. Biomed.
572 Analysis, 55 (2011) 823-831.
573 [38] R. Ramautar, G.W. Somsen, G.J. de Jong, Inside front cover: CE- MS for
574 metabolomics: Developments and applications in the period 2016–2018,
575 Electrophoresis, 40 (2019) 1-216.
576 [39] V. Rodríguez Robledo, W. Franklin Smyth, Review of the CE- MS platform as a
577 powerful alternative to conventional couplings in bio- omics and target- based
578 applications, Electrophoresis, 35 (2014) 2292–2308.
579 [40] T. Acunha, C. Simóa, C. Ibáñeza, A. Gallardo, A. Cifuentes, Anionic metabolite
580 profiling by capillary electrophoresis–mass spectrometry using a noncovalent polymeric
581 coating. Orange juice and wine as case studies, J. Chromatogr. A, 1428 (2016) 326–335
582 [41] R. García-Villalba, C. León, G. Dinelli, A. Segura-Carretero, A. Fernández-
583 Gutiérreza, V. Garcia-Cañas, A. Cifuentes, Comparative metabolomic study of

584 transgenic versus conventional soybean using capillary electrophoresis-time-of-flight
585 mass spectrometry, J. Chromatogr. A, 1195 (2008) 164–173.
586 [42] T. Levandi, C. Leon, M. Kaljurand, V. Garcia-Cañas, A. Cifuentes, Capillary
587 electrophoresis time-of-flight mass spectrometry for comparative metabolomics of
588 transgenic versus conventional maize, Anal. Chem. 80 (2008) 6329–63.
589 [43] T. Sangster, H. Major, R. Plumb, A.J. Wilson, I.D. Wilson, A pragmatic and
590 readily implemented quality control strategy for HPLC-MS and GC-MS-based
591 metabonomic analysis, Analyst, 131 (2006) 1075–1078.
592 [44] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der
593 Werf, Centering, scaling, and transformations: improving the biological information
594 content of metabolomics data, BMC Genomics, 7 (2006) 142
595 [45] A. Gil de la Fuente, J. Godzien Alberto, M. Fernández López, F.J. Rupérez, C.
596 Barbas, A. Otero, Knowledge-based metabolite annotation tool: CEU Mass Mediator, J.
597 Pharmaceu. Biomed. Anal. 154 (2018) 138-149
598 [46] R. Ramautar, G. W. Somsen, G. J. de Jong, CE–MS for metabolomics:
599 Developments and applications in the period 2014–2016, Electrophoresis 38 (2017)
600 190–202.
601 [47] R. Ramautar, G. W. Somsen, G. J. de Jong, CE–MS for metabolomics:
602 Developments and applications in the period 2016–2018, Electrophoresis 40 (2019)
603 165–179.
604 [48] A. García, J. Godzien, Á. López- Gonzálvez, C. Barbas, Capillary electrophoresis
605 mass spectrometry as a tool for untargeted metabolomics, Bioanalysis 9 (2017) 99–130.
606 [49] E. Saccenti, H.C.J. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M.W.B. Hendriks,
607 Reflections on univariate and multivariate analysis of metabolomics data, Metabolomics
608 10 (2014) 361–374.

609 [50] G.A. Burdock, Fenaroli's Handbook of Flavor Ingredients (6th ed.), in: CRC Press,
610 Boca Raton, FL, USA, 2016
611 [51] D. Ryana, R. Shelliea, P. Tranchida, A. Casillib, L. Mondello, P. Marriott, Analysis
612 of roasted coffee bean volatiles by using comprehensive two-dimensional gas
613 chromatography–time-of-flight mass spectrometry, J. Chromatogr. A 1054 (2004) 57–
614 65.
615 [52] L. Mondell, A. Casilli1, P. Quinto, T. Paola Dugo, R. Costa, S. Festa, G.Dugo,
616 Comprehensive multidimensional GC for the characterization of roasted coffee beans, J.
617 Sep. Sci. 27 (2004) 442-450.
618 [53] L. Mondello, R. Costa, P. Quinto, T.Paola Dugo, M. Lo Presti, S. Festa, A. Fazio,
619 G. Dugo, Reliable characterization of coffee bean aroma profiles by automated
620 headspace solid phase microextraction-gas chromatography-mass spectrometry with the
621 support of a dual-filter mass spectra library, J. Sep. Sci. 28 (2005) 1101-109.
622 [54] S.H. Zeisel, M.H. Mar, J.C. Howe, J.M. Holden, Concentrations of choline-
623 containing compounds and betaine in common foods, Journal of Nutrition, 133 (2003)
624 1302-1307.
625 [55] D.J. Kwon, H.J. Jeong, H. Moon, H.N. Kim, J.H. Cho, J.E. Lee, K.S. Hong, Y.S.
626 Hong, Assessment of green coffee bean metabolites dependent on coffee quality using a
627 1H NMR-based metabolomics approach, Food Res. Int. 67 (2015) 175-182.
628 [56] R. Dorfner, T. Ferge, C. Yeretzian, A. Kettrup, R. Zimmermann, Laser mass
629 spectrometry as on-line sensor for industrial process analysis: process control of coffee
630 roasting, Anal. Chem. 76 (2004) 1386-1402.
631 [57] F. Wei, K. Furihata, M. Koda, F. Hu, T. Miyakawa, M. Tanokura, Roasting process
632 of coffee beans as studied by nuclear magnetic resonance: time course of changes in
633 composition, J Agric Food Chem. 60 (2012) 1005-12.

634 [58] O.G. Vitzthum, P. Werkhoff, Steam volatile aroma constituents of roasted coffee:
635 neutral fraction, Z. Lebensm. Unters.-Forsch. 160 (1976) 277-291.
636 [59] P. Thammarat, C. Kulsing, K. Wongravee, N. Leepipatpiboon, T. Nhujak,
637 Identification of volatile compounds and selection of discriminant markers for elephant
638 dung coffee using static headspace gas chromatography-mass spectrometry and
639 chemometrics, Molecules 23 (2018) 1910, doi:10.3390/molecules23081910
640 [60] T.E. Kinlin, R. Muralidhara, A. Pittet, A. Sanderson, J.P. Walradt, Volatile
641 components of roasted filberts, J. Agr. Food Chem. 20 (1972) 1021-1028.
642 [61] R. Silwar, C. Lüllmann, Investigation of aroma formation in Robusta coffee during
643 roasting, Café cacao thé 37 (1993) 145-152.

644 **Figure captions**

645 **Figure 1**. Base peak electropherograms obtained in positive ionization mode for green
646 coffee (GCB), light roasted coffee (LRC), medium roasted coffee (MRC) and dark
647 roasted coffee (DRC) under optimal separation conditions. CE-MS conditions are
648 summarized in Section 2.4.

649 **Figure 2**. Score plots of the PCA models from the CE-MS data obtained from the
650 analysis of GCB, LRC, MRC and DRC samples. (A) PCA including QC samples and
651 (B) PCA without QC samples.

652 **Figure 3.** Score plots for the PLS-DA models of two-class comparisons and their
653 corresponding permutation test.

654 **Figure 4.** Extracted ion electropherograms (EIEs) for the potential markers of coffee
655 roasting process obtained in positive ionization mode.

656 **Figure 5.** Box-plots for the different metabolites unequivocally and tentatively
657 identified.

Figure 1.

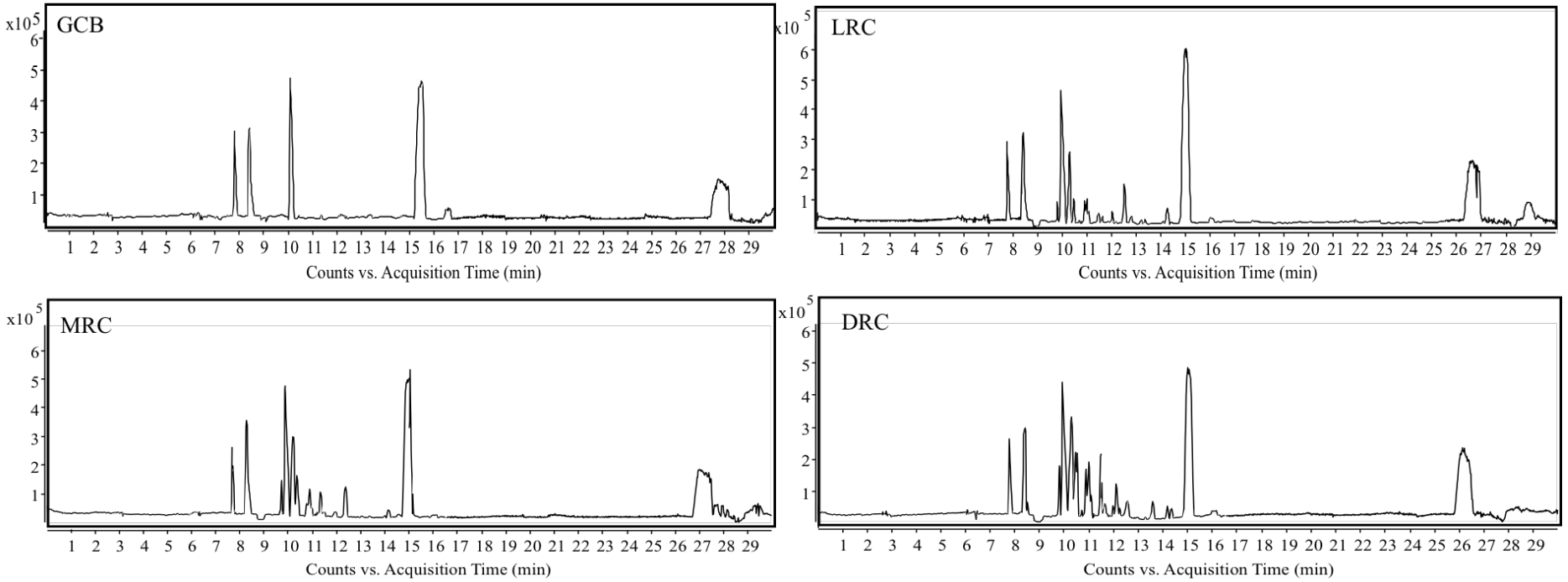

GCB
x10 5
Counts vs. Acquisition Time (min)
LRC
x10 5
Counts vs. Acquisition Time (min)
MRC
x10 5
Counts vs. Acquisition Time (min)
DRC
x10 5
Counts vs. Acquisition Time (min)

Figure 2.

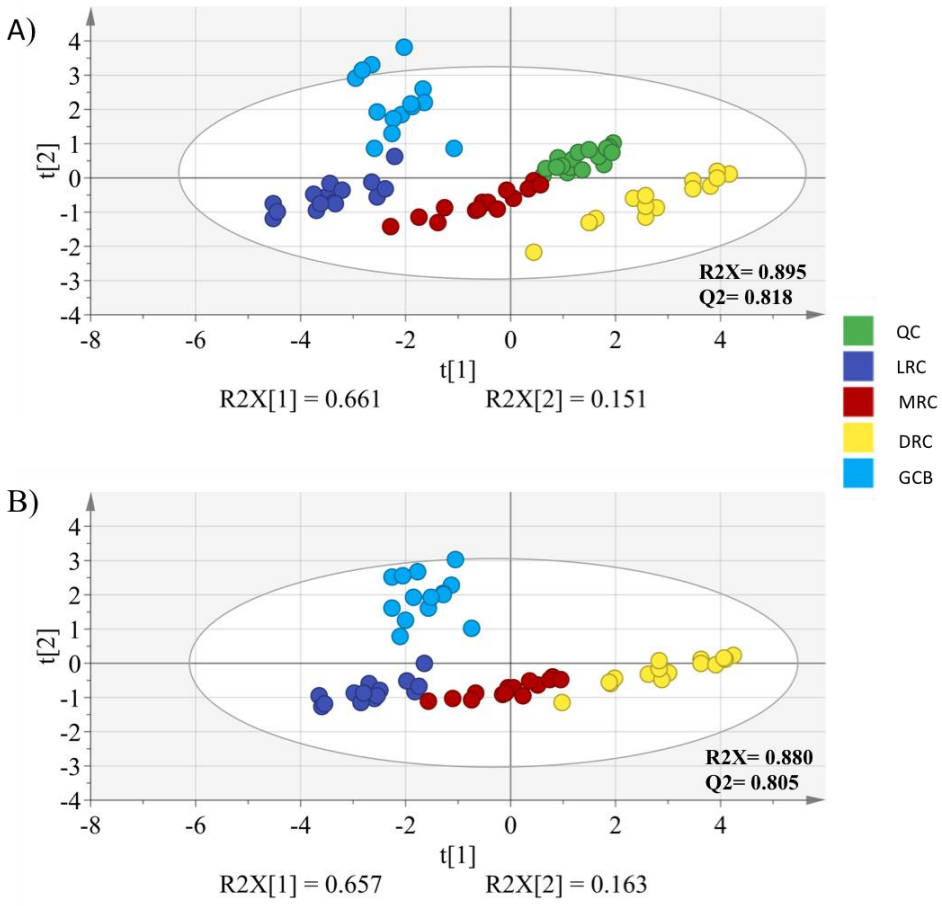

A)
t[2]
t[1]
R2X= 0.895
Q2= 0.818
R2X[1] = 0.661
R2X[2] = 0.151
QC
LRC
MRC
DRC
GCB
B)
t[2]
t[1]
R2X= 0.880
Q2= 0.805
R2X[1] = 0.657
R2X[2] = 0.163

Figure 3.

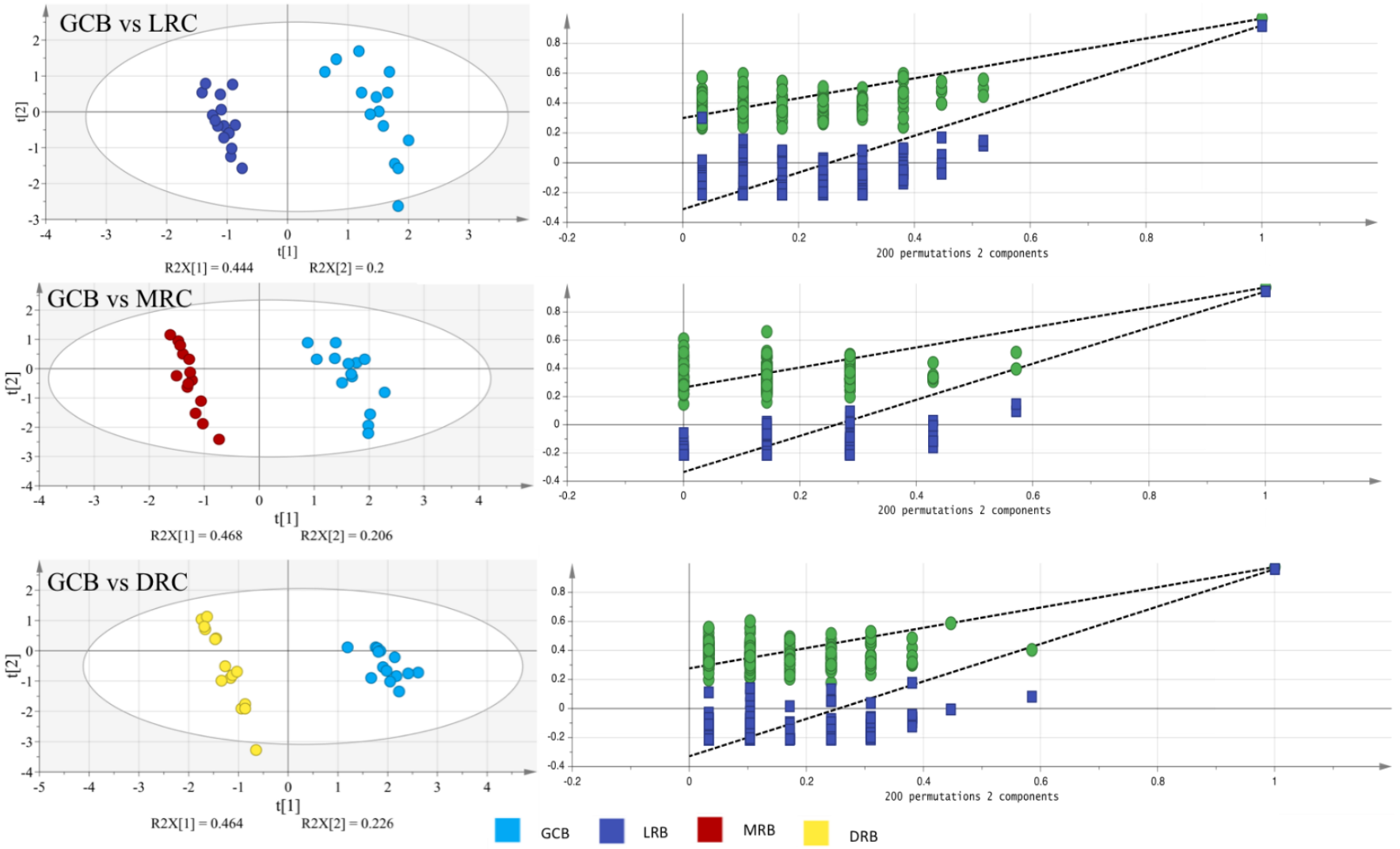

GCB vs LRC
t[2]
t[1]
R2X[1] = 0.444
R2X[2] = 0.2
200 permutations 2 components
GCB vs MRC
t[2]
t[1]
R2X[1] = 0.468
R2X[2] = 0.206
200 permutations 2 components
GCB vs DRC
t[2]
t[1]
R2X[1] = 0.464
R2X[2] = 0.226
200 permutations 2 components
GCB
LRB
MRB
DRB

Figure 4. (online version)

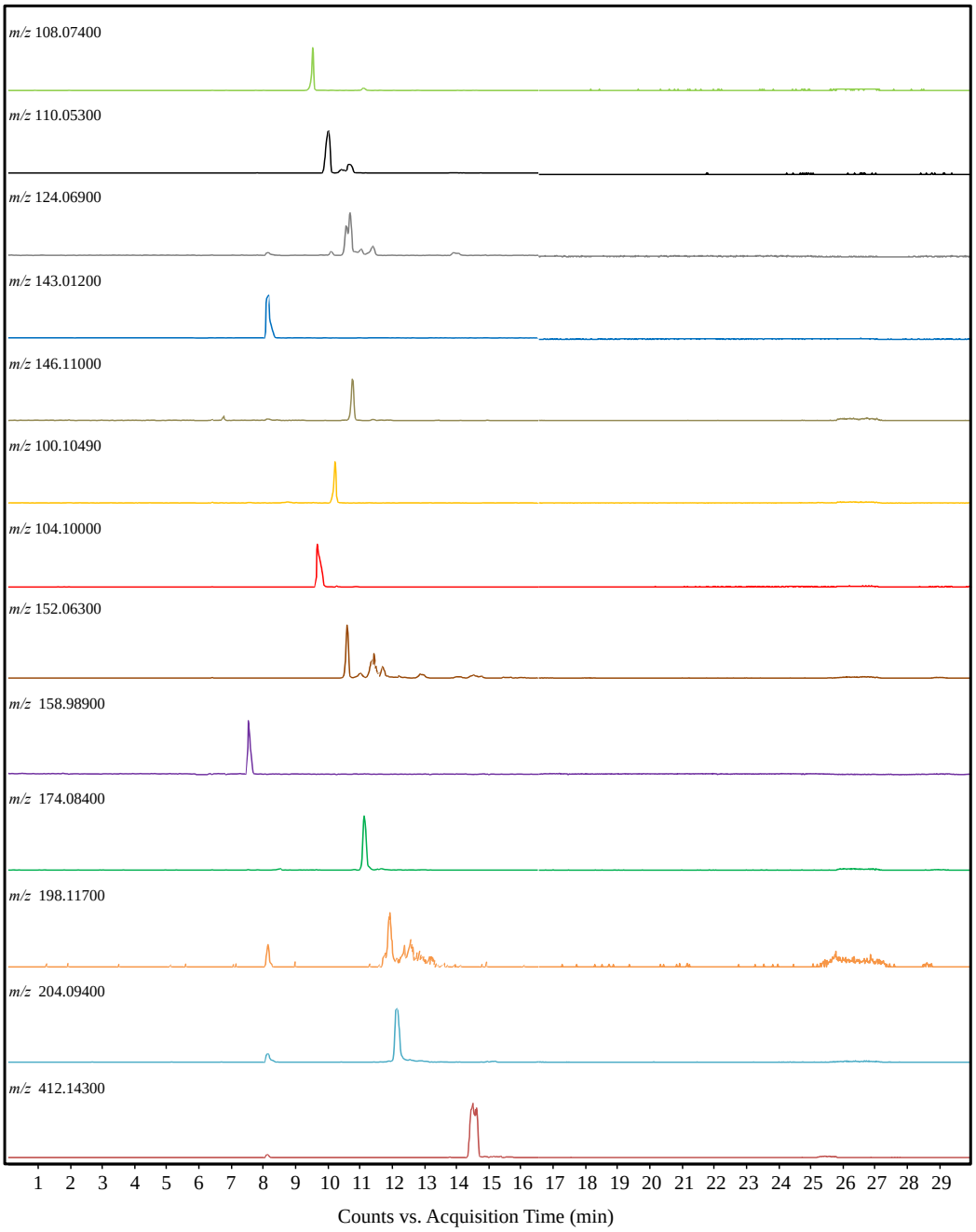
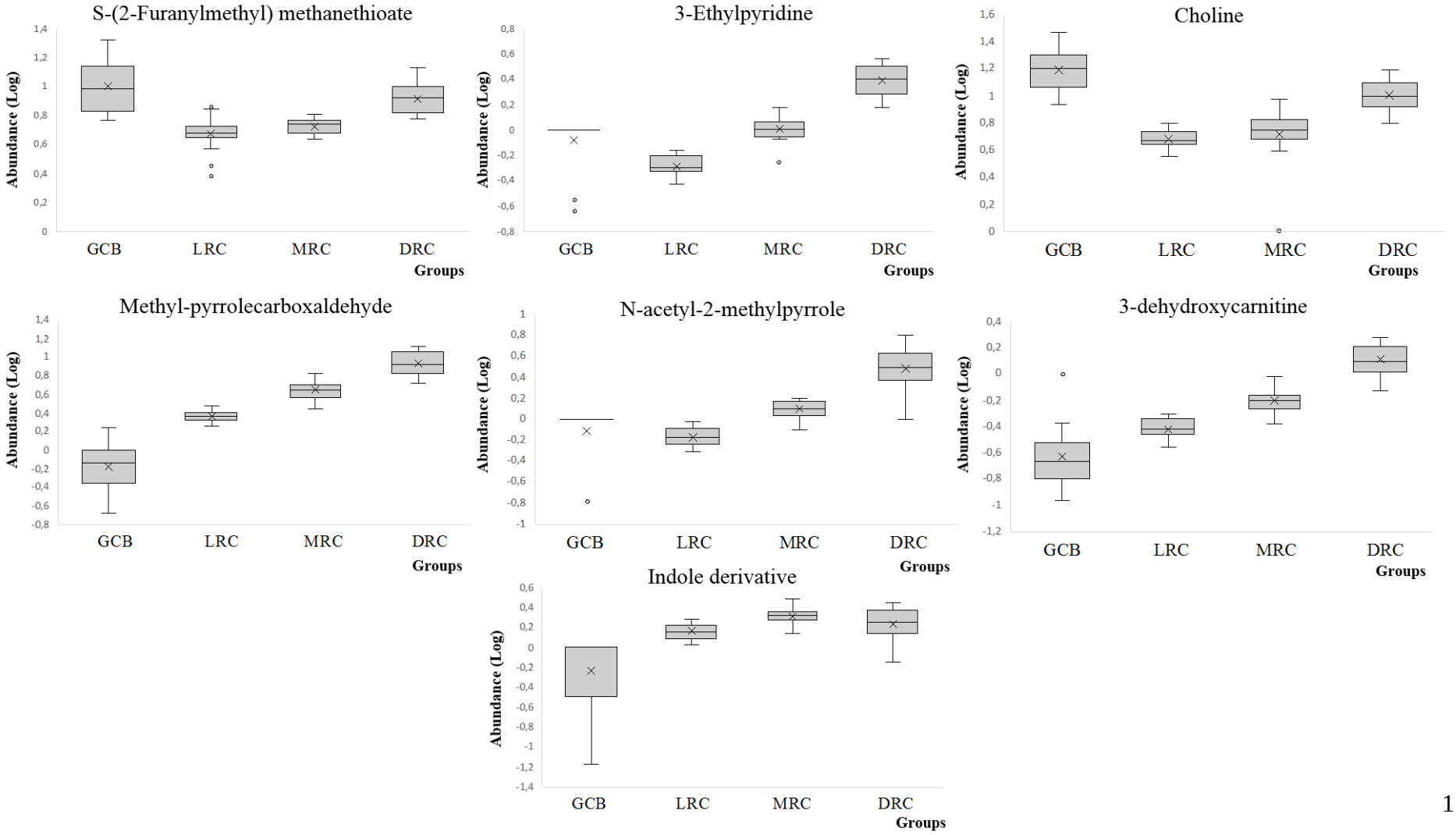


Figure 5.


S-(2-Furanylmethyl) methanethioate
3-Ethylpyridine
Choline
Methyl-pyrrolecarboxaldehyde
N-acetyl-2-methylpyrrole
3-dehydroxycarnitine
Indole derivative
Abundance (Log)
GCB
LRC
MRC
DRC
Groups

Tables

Table 1. Quality parameters and statistical values for the PLS-DA models built for the three different pairwise groups comparisons.

	Quality parameters			Cross-validated ANOVA	
	R^2X	R^2Y	Q^2	F-value	p-value
GCB vs LRC	0.644	0.965	0.919	58.2	5.2×10^{-12}
GCB vs MRC	0.674	0.975	0.946	81.3	3.1×10^{-13}
GCB vs DRC	0.690	0.975	0.960	132.5	5.4×10^{-16}

Table 2. MS/MS fragmentation of potential markers of coffee roasting process found in this work

							VIP values (p-value)[a]			
#	Migration time (t_m)	Molecular Formula	Identification	$[M+H]^+$ (*m/z*)	Mass error (ppm)	Main MS/MS fragments	GCB vs LRC	GCB vs MRC	GCB vs DRC	Roasting trend
1	7.8	-	Unknown	158.9891	7	No fragmentation pattern	1.3 (1.7×10^{-4})	1.2 (1.8×10^{-4})	0.5 (0.0453)	↓
2	8.4	$C_6H_6O_2S$	S-(2-Furanylmethyl) methanethioate	143.0117	20	82.9977 39.0077	1.7 (5.2×10^{-6})	1.3 (1.9×10^{-6})	0.4 (0.2289)	↓
3	9.9	C_7H_9N	3-Ethylpyridine[b]	108.0737	2	65.0399 92.0575	0.9 (0.023)	1.2 (0.022)	1.4 (0.022)	↑
4	10.1	$C_5H_{14}NO$	Choline[b]	104.1061[c]	9	58.0655 60.0808 45.0564	2.3 (1.0×10^{-6})	1.8 (1.9×10^{-6})	0.9 (0.004)	↓
5	10.4	C_6H_7NO	Methyl-pyrrolecarboxaldehyde	110.0530	2	67.0417 95.0368 80.0486 41.0360	2.4 (2.0×10^{-6})	2.6 (1.9×10^{-6})	2.9 (1.7×10^{-6})	↑
6	10.6	-	Unknown	100.1049	-	44.0498 70.0649 41.0389	0.7 (0.021)	1.4 (0.022)	1.7 (0.020)	↑
7	11.0	-	Unknown	152.0631	-	53.0382 93.0546	1.1 (0.002)	1.1 (4.4×10^{-4})	1.5 (4.9×10^{-6})	↓

8	11.1	C_6H_7NO	N-acetyl-2-methylpyrrole	124.0685	2	80.0494 81.0536 109.0491 53.0401	1.4 (0.020)	1.5 (0.021)	1.7 (0.020)	↑
9	11.2	$C_7H_{15}NO_2$	3-dehydroxycarnitine	146.1101	1	43.0178 87.0428	1.4 (4.4×10^{-4})	1.9 (1.9×10^{-6})	2.3 (6.9×10^{-7})	↓
10	11.6	-	Unknown	174.0837	-	144.0793 92.0486 93.0568 65.0379 130.0620	0.9 (0.139)	0.8 (0.274)	1.2 (0.135)	↓
11	12.5	-	Unknown	198.1165	-	No fragmentation pattern	2.4 (5.0×10^{-7})	2.1 (3.6×10^{-6})	1.8 (1.7×10^{-6})	↓
12	12.7	$C_{12}H_{13}NO_2$	Indole derivative	204.0942	2	107.0731 158.0933 144.0834 186.0910	2.4 (3.6×10^{-5})	2.3 (3.7×10^{-4})	2.0 (0.612)	↑
13	15.4	-	Unknown	412.1427	-	138.0599 94.0646 192.0386 288.4826	1.5 (2.7×10^{-3})	1.6 (9.7×10^{-4})	1.6 (3.6×10^{-4})	↓

[a] p-value of Mann Whitney U (FDR) test. Cut-off value: 0.045.
[b] Identification corroborated using commercial standards.

[c] $[M]^+$

↑ The level of the compound increases with roasting; ↓ The level of the compound decreases with roasting.

Electronic Supplementary Material (online publication only)

Click here to download Electronic Supplementary Material (online publication only): Supplementary information_DEF_REV.docx