

Application of Correlation Analysis to Classify Ground Coffee Bean Extracts Using UPLC-HRMS in MS^E Mode

Wang Han, Mark Ritchie

Waters Corporation



Abstract

In this application note, samples of coffee brews rapidly profiled using UPLC-MS/MS in order to elucidate the origin and treatment of the beans.

Benefits

- Rapidly identify patterns of difference between multiple ground coffee bean extracts using Correlation Analysis in Progenesis QI
- Provide meaningful interpretations from Data Independent Analysis (DIA) of food products using UPLC-MS^E

Introduction

Coffee is a chemically complex beverage, brewed from the roasted seeds of the berries (referred to as beans) most commonly from two plants of the *Coffea* genus: *Coffea arabica* (Arabica or Arabian coffee) and *Coffea canephora* (Robusta coffee). The two different species give rise to distinct variation in the taste and aroma of the drink. Further variations can be introduced by the fermentation, processing, roasting, grinding, brewing, and storage, which can make the distinction between the species more difficult to the consumer, and there are concerns that the cheaper Robusta may be fraudulently substituted for Arabica.

Much of the chemistry of the flavor and aroma compounds is understood, and attributed to compounds such as sucrose, furans, pyrazine, alkyl-pyridines, pyrroles, and of course caffeine. Related precursors/metabolites of these compounds that are seen as desirable include trigonelline and chlorogenic acid.

Providing chemical information on coffee brews can be used to assess the quality of a grind, process or species, or to further explore the taste and aroma chemistry. In this application note, samples of coffee brews were provided blind, in order to elucidate the origin and treatment of the beans.

Experimental

Sample preparation

Coffee beans were ground and extracted with boiling water (brewed), and the soluble supernatants provided. The four samples (A, B, C and D) were provided by a major food company without identification of origin or processing history. Each sample was centrifuged at 10,000 rpm for 10 minutes to remove insolubles and 1 μ L of undiluted supernatant was injected on to the UPLC-QToF MS system. QC samples were made by mixing equal amounts of the four samples, and 12 injections from this mixture were used to condition the system at the start of the analysis. To ensure good experimental design, triplicates of the sample were then injected in a randomized run order, and the whole set were bracketed by a run order QC. Finally, the analytical run was concluded by two blank injections. This process was carried out in both positive and negative electrospray ionization conditions, and APCI ionization in both positive and negative modes.

UPLC conditions

UPLC system:	ACQUITY UPLC I-Class PLUS
Column:	ACQUITY UPLC HSS T3 1.8 μ m, 2.1 $\times$ 150 mm
Column temp.:	40 $^{\circ}$ C
Sample temp.:	10 $^{\circ}$ C
Flow rate:	0.6 mL/min
Injection volume:	1 μ L
Mobile phase A:	Water with 0.1% formic acid
Mobile phase B:	Acetonitrile with 0.1% formic acid
Total run time:	16 min

Gradient table:

Time (min)	%A	%B	Curve
0	99	1	Initial
0.5	99	1	6
12	60	40	6
14	0	100	1
16	99	1	1

MS conditions

MS system:	Xevo G2-XS QToF
Ionization mode:	ESI+, ESI-, APCI+, APCICollision energy (LE): 4 eV
Collision energy (HE ramp):	10–45 eV
Scan time:	0.1 s
Acquisition range:	50–1550 <i>m/z</i>
Capillary voltage:	1 kV
Sampling cone:	40 V
Source temp.:	120 °C
Desolvation temp.:	500 °C
Cone gas flow:	30 L/hr

Desolvation gas flow:	1000 L/hr
Lock mass:	Leucine enkephalin (556.2771 <i>m/z</i>)
Acquisition mode:	MS ^E
Instrument control software:	MassLynx (v4.2)

Data management and processing

Raw data was processed using Waters Progenesis Q1 Software (v2.3), exporting data to EZInfo and querying online databases to help with unknown elucidation. Correlation Analysis within Progenesis Q1 was used to group together patterns of change across the samples.

Results and Discussion

UPLC-MS chromatograms of the four samples showed good separation with narrow chromatographic peak widths typically 3–5 s wide. Each replicate set overlays well across the replicates, showing good reproducibility of retention time and intensity. Although some of the peaks varied in intensity between samples, the retention times were well conserved across the entire data set, as shown in Figure 1.

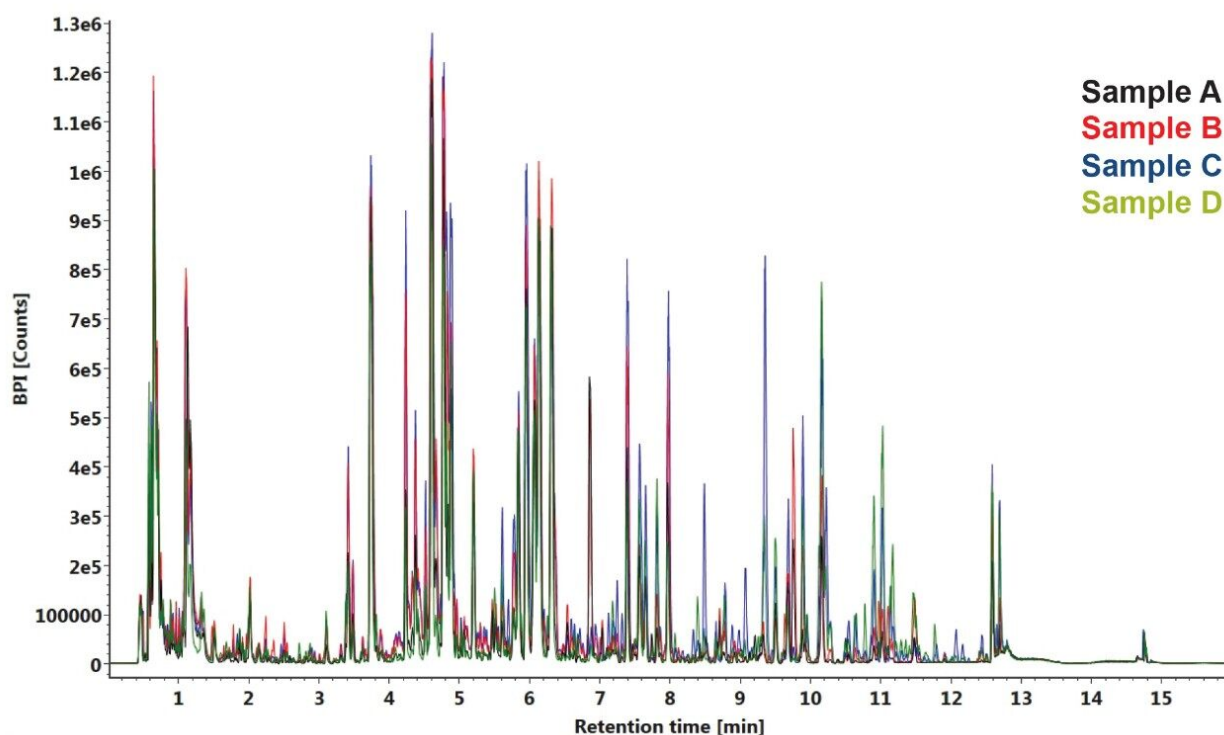


Figure 1. Overlaid UPLC-MS chromatograms for three replicates of four samples of coffee bean extract.

Employing the Data Independent Analysis (DIA) method described here, MS^E , users can gather reproducible relative quantitative data, while simultaneously acquiring fragment ions for compound identification and elucidation. The aggregate dataset was searched against several locally stored databases (HMDB, Metlin MS/MS spectral library, NIST, lipid maps) in Progenesis Q1 Software to identify the compounds present utilizing both the accurate mass precursor and the fragment ion information. More than 300 compounds were assigned from this search with a mass error of <5 ppm, an isotopic distribution similarity >85%, and a fragment score >10. Identified molecules included acids, esters, polysaccharides, glycosides, ketones, salts, and amino acids. Further compounds were also identified with high quality but were unresolved due to ambiguity caused by isomerization in the compound classes, especially in the chlorogenic acids (CGA) and associated lactones.

Long compound lists, although informative, provide little insight on their own when comparing complex samples. Further analysis was conducted utilizing the relative quantitative aspect of the data, using multivariate statistics by exporting the processed data to EZInfo. The Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) model allows the determination of which compounds within a given sample contribute the most to the difference between that sample and the others. The compounds

representing the largest difference in each of the samples after pair wise comparison of samples were selected and tagged, and those assignments returned to Progenesis.

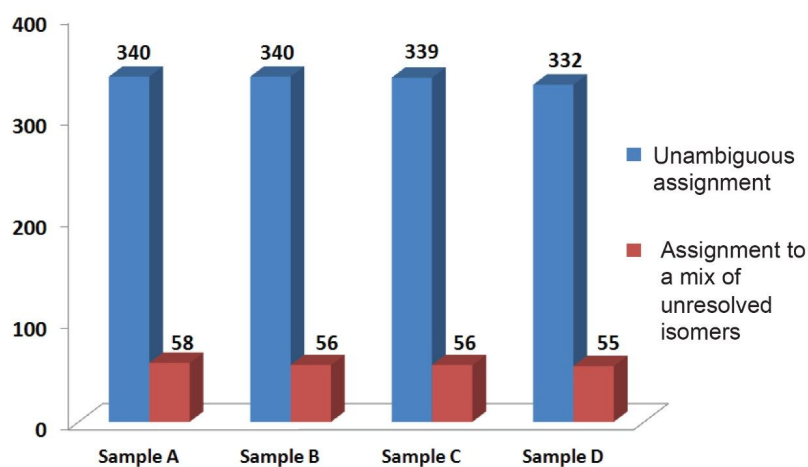


Figure 2. Compounds identified through database searching of UPLC-MS^E data for four coffee bean extracts. The majority of compounds could be clearly identified, however some compounds could only be elucidated to a group of structural isomers.

Correlation Analysis within Progenesis QI Software was used to group the compounds that shared a particular pattern of change across the dataset. 39 compounds were found to have significant changes across the four datasets, which were then used to elucidate the species and roasting of the coffee beans. These were clustered into nine groups of distinct profiles of abundance across the samples.

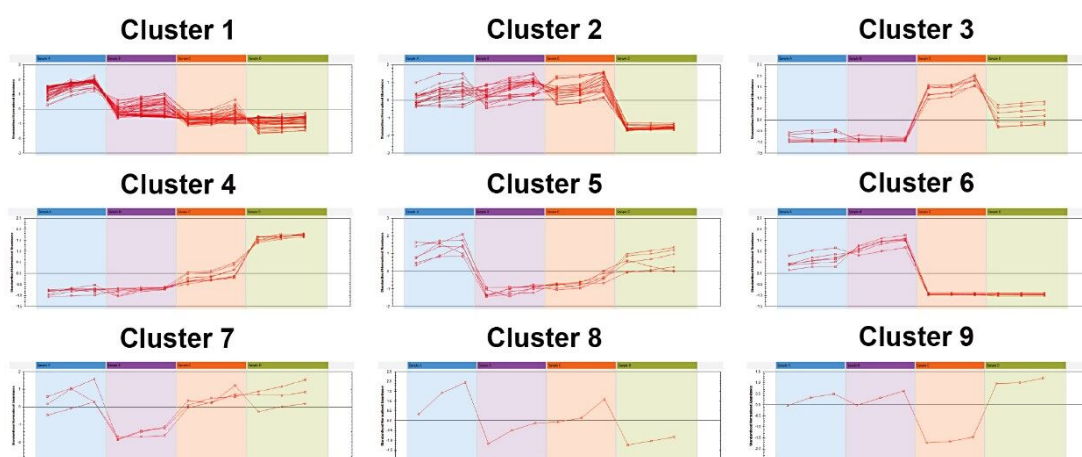


Figure 3. Progenesis Q1 Correlation Analysis of loadings plot differentiated compounds from four extracts of coffee beans, shown as standardized normalized abundance. This is used to cluster the compounds that share a particular pattern of change across the dataset. Samples are left to right in each cluster; A (blue), B (purple), C (orange), and D (green). Compounds in clusters 2 and 6 are discussed in the Results and Discussion section.

Cluster 6 showed the greatest differentiators between samples A+B, and samples C+D. One of the compounds identified was Mozambioside, which is a bitter tasting furokaurane glycoside that is specific to Arabica beans¹ and thought to contribute to Arabica's distinctive taste. This was identified at high levels in samples A and B (Figure 4), indicating that these were extracts of Arabica beans. The lack of Mozambioside detected in samples C and D implied that these were likely to be extracts from Robusta beans. Samples C and D also had a higher relative level of caffeine than in samples A and B, which also supports this (data not shown).



Figure 4. Standardized normalized abundance of Mozambioside across the four coffee bean extracts A–D (left to right). High levels of Mozambioside in samples A and B imply that these are extracts from Arabica beans, while the corresponding low levels in samples C and D imply that these extracts are from Robusta beans.

Cluster 2 highlights differences in sample D compared to the rest of the samples. Many of the compounds identified in this pattern were Dicafeoylquinic acid isomers (DiCQAs) (Figure 5).

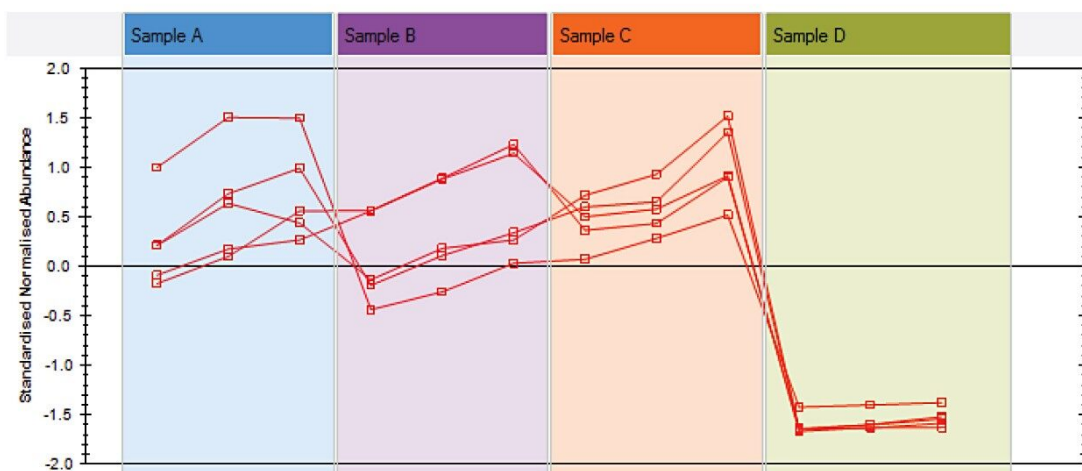


Figure 5. Standardized normalized abundance of Dicafeoylquinic acid isomers (DiCQAs) across the four coffee bean extracts samples A–D (left to right). DiCQAs are gradually transformed in chlorogenic acid lactones (CGLs) during the roasting process, so the low level in sample D implies a greater degree of roasting in that sample.

During the roasting process these compounds are gradually transformed into Chlorogenic Acid Lactones (CGLs).^{2,3} The very low DiCQA content in sample D implies that this has gone through the roasting to a

greater degree than the others. There was an earlier implication that the coffee beans used to make extracts A and B were different from extracts C and D, so further differences between samples C and D only were explored, to remove the species variable. Two compound profiles showed clear differences between these two samples: nicotinic acid and trigonelline, as seen in Figure 6. During roasting trigonelline is transformed into nicotinic acid,⁴ which correlates with the earlier observation that sample D went through a greater degree of roasting. Further to this, the quantity of trigonelline in sample C implies that these beans may not have been roasted at all. As levels of trigonelline vary between coffee species, with Arabica having slightly more than Robusta,⁵ this difference may not have been observed when looking across the full dataset.

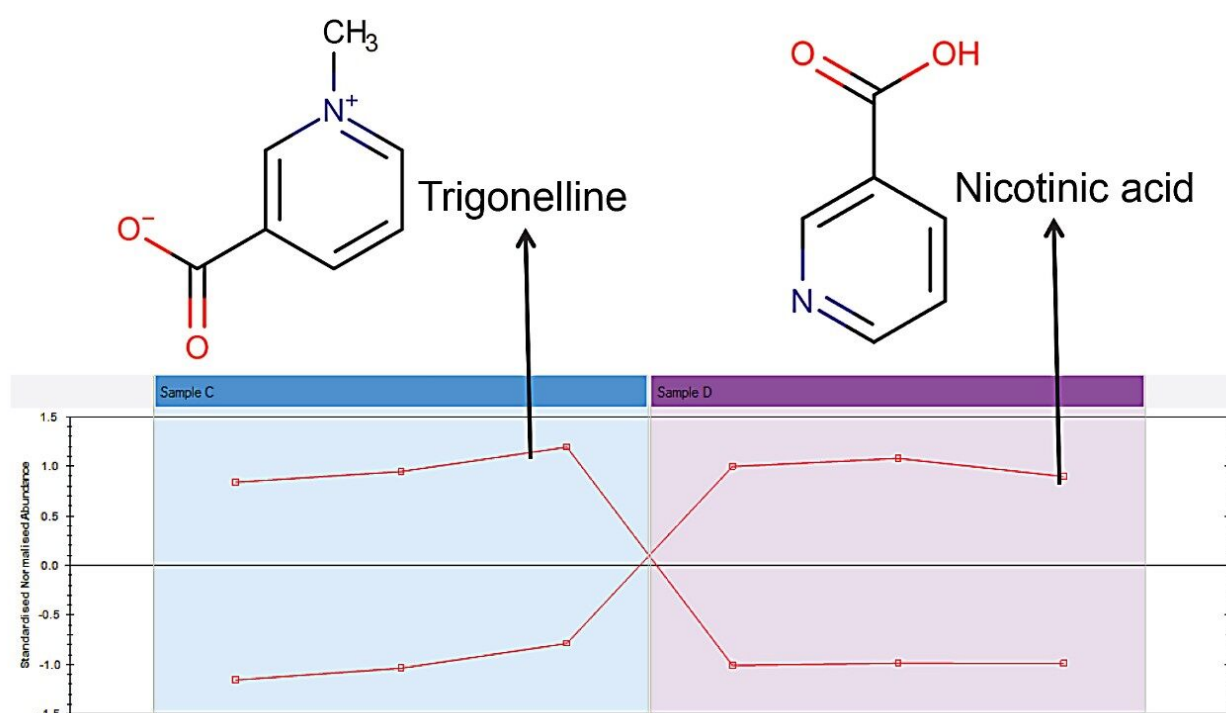


Figure 6. Standardized normalized abundance of trigonelline and nicotinic acid between two coffee bean extracts C and D (left to right), both tentatively assigned as Robusta in origin. Trigonelline is one of the major analytes in unprocessed coffee that is transformed during the roasting process to nicotinic acid. These data imply that sample C may not have been roasted.

Conclusion

- The ACQUITY UPLC I-Class PLUS System coupled with Xevo G2-XS QToF MS were used to rapidly profile coffee bean extracts using a Data Independent Acquisition (DIA) mode, MS^E.
- Correlation Analysis in Progenesis QI Software was used to cluster together patterns of difference between the samples.
- Detection of Mozambioside in samples A and B implied that these extracts were taken from Arabica coffee beans, and the lack of this metabolite in C and D implied that these were extracts of Robusta.
- Differences in trigonelline, nicotinic acid, and Dicafeoylquinic acid isomers implied that sample D underwent stronger roasting than sample C.
- Such approaches allow the classification and characterization of coffee extracts and could potentially be used to detect Robusta being presented fraudulently as Arabica.

References

1. Lang et al. *J. Agric. Food Chem.* 63, 48: 10492–10499.
2. Farah et al. *J. Agric. Food Chem.* (2005) 53: 1505–1513.
3. Lima et al. *J. Agric. Food Chem.* (2016) 64: 2361–2370.
4. Trugo L C. Analysis of Coffee Products. B. Caballero, L C Trugo and P M Finglas (Eds.), Oxford: Academic Press. (1st ed.) *Encyclopedia of Food Sciences and Nutrition*. (2003) 7: 498.
5. Stennert A, Maier H G. Trigonelline in Coffee. *Z Lebensm Unters Forch* (1994) 199: 198.
<https://doi.org/10.1007/BF01193443>.

Featured Products

Xevo G2-XS ToF Time-of-Flight Mass Spectrometry <https://www.waters.com/waters/en_US/Xevo-G2-XS-ToF-Time-of-Flight-Mass-Spectrometry/nav.htm?cid=134798183>

Progenesis QI Software <https://www.waters.com/waters/en_US/Progenesis-QI/nav.htm?cid=134790652>

ACQUITY UPLC I-Class PLUS System <https://www.waters.com/waters/en_US/UPLC-inlet-to-MS-with-the-best-dispersion/nav.htm?cid=134613317>

MassLynx MS Software <https://www.waters.com/waters/en_US/MassLynx-MS-Software/nav.htm?cid=513662>

720006504, February 2019

©2019 Waters Corporation. All Rights Reserved.