Application News

GCMS-TQ™8040 NX Triple Quadrupole Mass Spectrometer
Smart Aroma Database™
Statistical Analysis Software eMSTAT Solution™

Discrimination of Coffee Bean Species Based on Aroma Compounds

Nozomi Maeshima, Saito Yoshihiro, Takehito Sagawa

User Benefits

- Qualitative analysis of aroma compounds can be easily performed using the Smart Aroma Database.
- Intuitive statistical analysis and exploration of characteristic compounds are possible using eMSTAT Solution.
- The AOC™-6000 multifunctional autosampler system enables the analysis of aroma compounds without sample pretreatment.

■ Introduction

Common coffee bean types are broadly classified into two types: *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta). Among them, Arabica is superior in flavor and is highly favored as a specialty beverage. However, in recent years, there have been multiple reports of products sold claiming to be 100% Arabica while being adulterated with Robusta[1].

A reported method for discriminating between the two is the analysis of lipophilic extracts of coffee beans using a benchtop NMR spectrometer[1]. This method uses 16-*O*-Methylcafestol, which is present in Robusta but almost undetectable in Arabica, as a marker. However, it requires complicated pretreatment such as extraction and concentration.

Therefore, in this Application News, we investigated a method to discriminate between Arabica and Robusta more easily and rapidly using GC-MS/MS and the statistical analysis software eMSTAT Solution.

■ Coffee Bean Species

Generally, three types of coffee are cultivated and enjoyed as beverages: Arabica, Robusta, and *Coffea liberica* (Liberica). Arabica has the highest production volume, accounting for over 60% of global coffee production. It is characterized by excellent aroma, acidity, and high quality, but it is vulnerable to high temperatures, high humidity, and pests, making it difficult to cultivate.

The coffee with the next highest production volume is Robusta. While the species is technically Canephora, it is usually called by this representative variety name. Compared to Arabica, Robusta is more resistant to pests and can adapt to hot and humid environments, giving it superior productivity. However, due to its characteristic strong aroma and bitterness, it is less popular on its own and is often used in blends with Arabica, or for instant and canned coffee.

Finally, there is Liberica, which is produced in very small quantities. Compared to Arabica, it lacks acidity and has a strong bitterness. For these reasons, it is rarely distributed and is consumed locally in producing regions such as West African countries. While there are over 100 species of coffee cultivated worldwide, approximately 99% of the beans are either Arabica or Robusta.

■ Sample and Analytical Methods

Commercially available coffee beans roasted at the same roastery were used as samples. Four brands of Arabica (Kilimanjaro, Brazil Bourbon, Mandheling G1, Brazil No.2 #18) and two brands of Robusta (Vietnam Robusta, Java Robusta) were prepared (Table 1). Each brand was individually crushed with a mill, and 1 g was weighed and sealed in a screw vial.

Table 1 Sample list

Brand	Types
Kilimanjaro	Arabica
Brazil Bourbon	
Mandheling G1	
Brazil No.2 #18	
Vietnam Robusta	Robusta
Java Robusta	

Volatile compounds were concentrated and introduced into the GC-MS using the HS-SPME method and then analyzed under the conditions registered in the GC-MS(/MS) Smart Aroma Database (Table 2). The GCMS-TQ™8040 NX and the AOC-6000 multifunctional autosampler system were used as the analytical instruments (Fig. 1). To verify whether rapid analysis is possible, a temperature ramp of 10°C/min was adopted, and a low-polarity column was used to obtain stable retention times for each compound. Data acquisition was performed in Scan mode or MRM mode, and each sample was analyzed three times.

Table 2 Analytical conditions

Instrument Configuration	
GC-MS/MS	: GCMS-TQ8040 NX
Autosampler	: AOC-6000
Column	: SH-I-5Sil MS*1 (30 m, 0.25 mm I.D., df = 0.25 µm)
Database	: Smart Aroma Database
GC Conditions	
Carrier Gas	: Helium gas
Vaporizing Chamber Temp.	: 250°C
Injection Mode	: Splitless
Column Oven Temp.	: 50°C (0–5 min) → 10°C/min (5–25 min) → 250°C (25–35 min)
MS Conditions	
Interface Temp.	: 250°C
Ion Source Temp.	: 200°C
Ionization	: EI
Measurement Mode	: Scan (*m/z* 35–400) or MRM

*1 P/N: 221-75954-30

Fig. 1 GCMS-TQ™8040 NX and AOC™-6000 multifunctional autosampler system

■ Differential Analysis Using Scan Data

First, differential analysis of the Scan data was performed using the multivariate analysis software Signpost MS (Reifycs Inc.). Through alignment processing, 210 compounds were assigned. The results of hierarchical clustering using these compounds are shown in Fig. 2, where Arabica is clustered on the left side and Robusta on the right side, suggesting that discrimination between the two is possible through the analysis of aroma compounds.

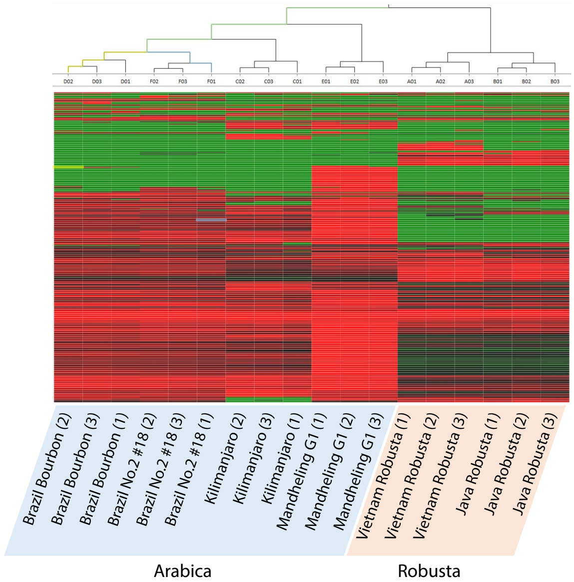

Fig. 2 Hierarchical clustering using scan data

Next, Principal Component Analysis (PCA) was performed on the same Scan data using Signpost MS. The results of the (a) score plot and (b) loading plot are shown in Fig. 3. From the score plot, Arabica is clustered on the left side and Robusta on the right side, suggesting that discrimination between the two is also possible in this analysis.

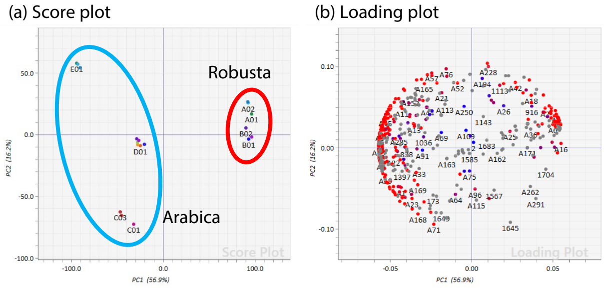

Fig. 3 PCA using scan data

■ Differential Analysis Using MRM Data

The Scan data suggested that discrimination between the two groups using aroma compounds is possible. Therefore, to obtain data with higher sensitivity and reproducibility and to improve reliability, data acquisition was performed in MRM mode. The Smart Aroma Database, which contains MRM conditions for approximately 500 aroma compounds, was used.

When the samples were analyzed in MRM mode, 178 compounds out of the aroma compounds registered in the Smart Aroma Database could be identified. In the alignment by Signpost MS, 210 compounds were assigned, indicating that over 80% of the compounds detected by GC-MS were successfully identified.

Next, statistical analysis of the identified compounds was performed using eMSTAT Solution. eMSTAT Solution is capable of both unsupervised PCA and supervised Partial Least Squares Discriminant Analysis (PLS-DA). For the analysis, 175 compounds without missing values were used. The results of PCA and PLS-DA are shown in Fig. 4 and Fig. 5, respectively. Although the left-right positioning is reversed between the two plots, both analyses clearly show that the samples are clustered by types.

These results suggest that discrimination between the two groups is possible not only with supervised PLS-DA, as expected, but also with unsupervised PCA.

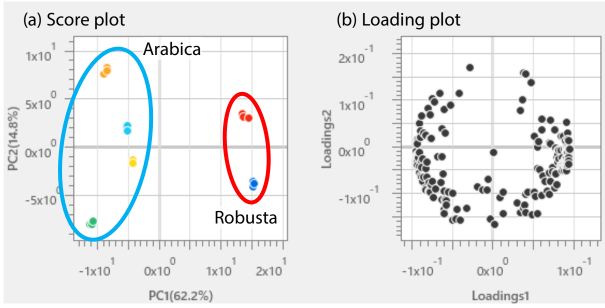

Fig. 4 PCA using MRM data

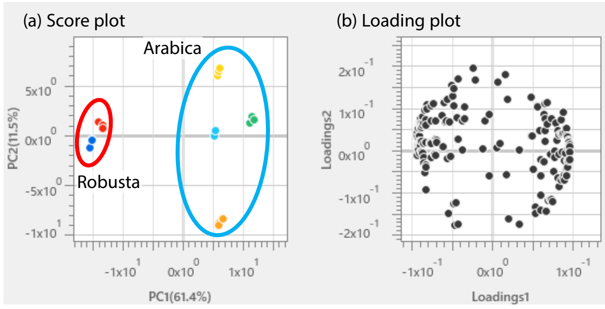

Fig. 5 PLS-DA using MRM data

Next, cluster analysis was performed using eMSTAT Solution based on the acquired MRM data, and a dendrogram was created (Fig. 6). As a result, Robusta was classified in the upper cluster and Arabica in the lower cluster, confirming clear grouping by type.

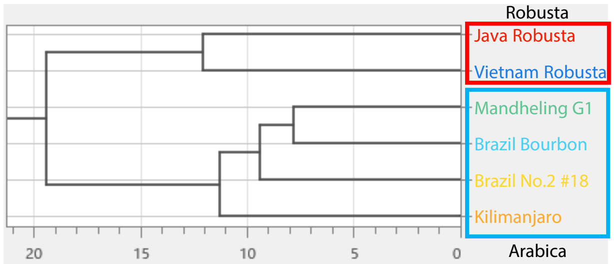

Fig. 6 Dendrogram analysis using MRM data

■ Exploration of Characteristic Aroma Compounds for Arabica and Robusta

Next, we investigated which compounds contribute to the discrimination between Arabica and Robusta. The PCA score plot (Fig. 4) reflects the differences in the tendencies of each group, with Arabica distributed on the left side and Robusta on the right side, clearly separated along the first principal component (PC1) axis. Referring to the corresponding loading plot, compounds located on the left side of the origin tend to have higher area values in Arabica, which is also positioned on the left in the score plot. On the other hand, compounds on the right side tend to show higher values in Robusta, which is located on the right in the score plot. In other words, the difference in the balance of these compounds is the main factor separating the two groups.

First, Arabica was commonly rich in compounds such as 5-Methyl furfural, Acetoin, and Furaneol acetate. MS chromatograms analyzed using LabSolutions Insight GCMS are shown in Fig. 7. In eMSTAT Solution, not only score plots and loading plots but also box plots can be easily drawn, making it possible to easily visualize and compare the distribution tendencies of each compound. From the box plots of the above compounds (Fig. 8), it was visually confirmed that all of them are contained in abundance in Arabica. These compounds are sometimes described as having a roasted or creamy quality, and it can be inferred that this result supports the characteristics of Arabica, which is superior in aroma and acidity.

On the other hand, a compound found in abundance in Robusta was *p*-Vinylguaiacol. This compound is known to cause an off-flavor similar to a medicinal odor in sake, and its content was low in Arabica. From this, it was suggested that *p*-Vinylguaiacol may be one of the compounds contributing to the bitterness and off-tastes characteristic of Robusta.

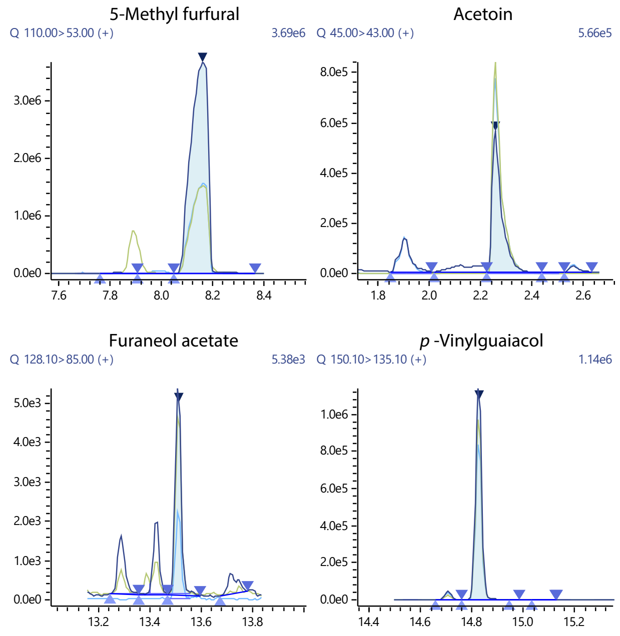

Fig. 7 MS chromatograms of representative compounds highly contributing to PC1

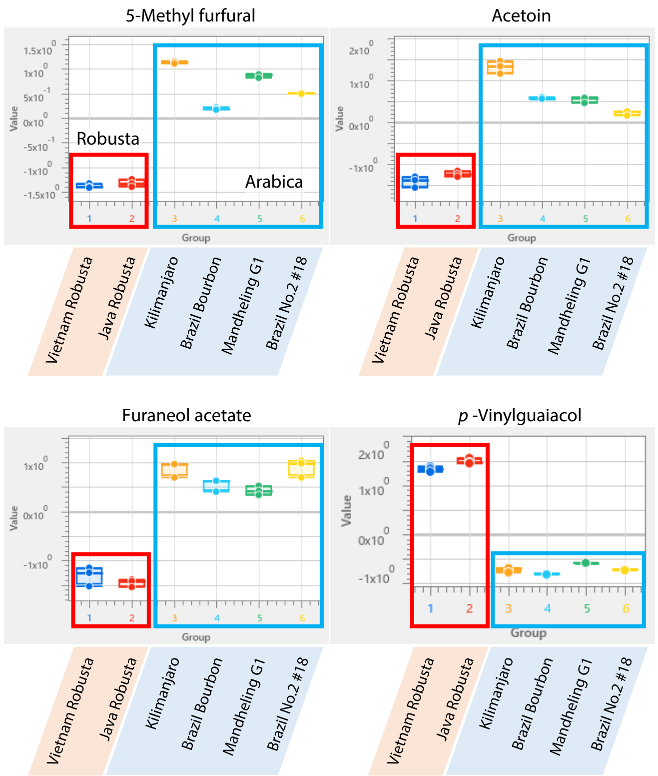

Fig. 8 Box plots of representative compounds highly contributing to PC1

The Smart Aroma Database includes not only MRM information but also comments on how the compound is evaluated in terms of aroma. Therefore, by referring to the comments, it is possible to infer what kind of aroma compound it is (Table 3).

Table 3. Representative aroma compounds contributing to PC1

Compound name	Sensory information registered in Smart Aroma Database
5-Methyl furfural	almond, caramel, burnt sugar
Acetoin	butter, cream
Furaneol acetate	caramel
p -Vinylguaiacol	clove, curry

■ Creation of a Discrimination Model Using eMSTAT Solution

eMSTAT Solution is capable of not only exploratory analysis such as PCA but also the creation of discrimination models. In this study, a Support Vector Machine (SVM) was used to verify the accuracy of a model that discriminates whether an unknown sample is Arabica or Robusta.

Specifically, targeting the four brands of Arabica (Table 1), a model was created using data from three of these brands plus the two Robusta brands as training data, leaving one brand out for verification. The excluded brand was then classified. As a result, all tested brands were correctly discriminated as "Arabica" with a score of 85 or higher (Table 4). These results suggest that highly accurate discrimination is possible using the model created by this method.

By using eMSTAT Solution, a discrimination model can be created without going through complex processes, enabling rapid and advanced analysis without requiring high-level specialized knowledge.

Table 4. Discrimination results by the model created using eMSTAT Solution

Brand	Score
Kilimanjaro	100
Brazil Bourbon	100
Mandheling G1	100
Brazil No.2 #18	85

■ Conclusion

In this Application News, we investigated the discrimination of coffee bean types based on aroma compounds using the triple quadrupole gas chromatograph mass spectrometer GCMS-TQ8040 NX and the GC-MS(/MS) Smart Aroma Database for aroma analysis.

Four types of Arabica and two types of Robusta were analyzed as samples. Analysis using Scan data confirmed a tendency for the two types to separate. Subsequently, data was acquired in MRM mode, which offers higher sensitivity and selectivity, and detailed analysis was performed using the statistical analysis software eMSTAT Solution. As a result, the two groups were clearly separated in PCA and cluster analysis, demonstrating the effectiveness of discrimination by this method. Additionally, a discrimination model was developed and validated, confirming that highly accurate discrimination is possible.

From the above results, we conclude that the discrimination of coffee beans by aroma compound analysis is highly feasible using this method. Furthermore, this method enables a comprehensive workflow, from discrimination to the identification of characteristic compounds for each type.

<Reference>

1) Y. Gunning et al., *Food Chem.*, 248, 52-60 (2018).

SHIMADZU
Shimadzu Corporation
www.shimadzu.com/an/

01-01123-EN First Edition: May 2026

> Please fill out the survey

Related Products Some products may be updated to newer models.

> GCMS-TQ8040 NX
Triple Quadrupole Gas Chromatograph Mass Spectrome…

> Smart Aroma Database™
Database for GC-MS and GC-MS/MS

> eMSTAT Solution
Statistical Analysis Software for Direct Ionizatio…

> AOC-6000 Plus
AOC-6000 Plus Multifunctional Autosampler

Related Solutions

> Food and Beverages

> Food Fraud

> Price Inquiry

> Product Inquiry

> Technical Service / Support Inquiry

> Other Inquiry