SHIMADZU

Investigation of Varietal Classification Method for Coffee Beans Using Aroma Compounds

Tetsuo Iida[1], Nozomi Maeshima[1], Yoshihiro Saito[1], Takehito Sagawa[1], Manami Kobayashi[1]
1) SHIMADZU Corporation, Japan

1. Introduction

Common coffee bean species 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.

2. Methods

2.1. Sample

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.

2.2. Analytical Methods

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 1 Sample list

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

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

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
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

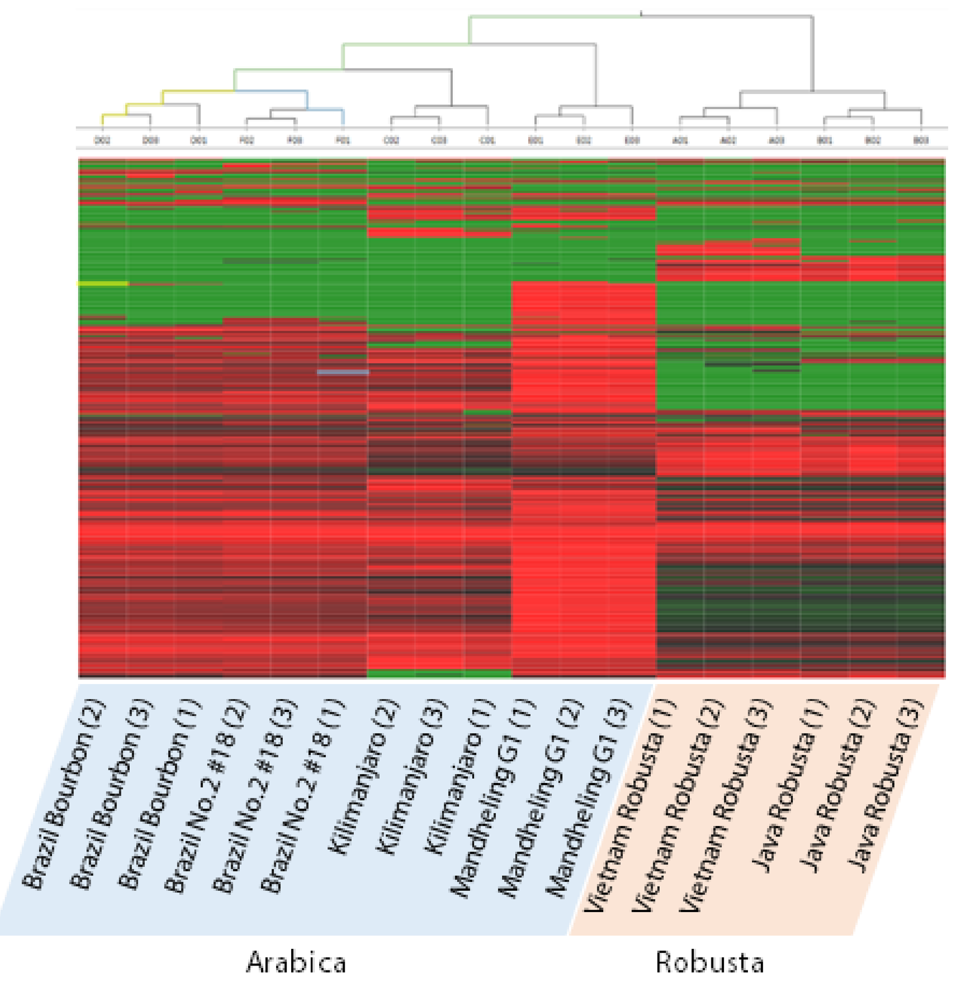

Fig. 2 Hierarchical clustering using scan data

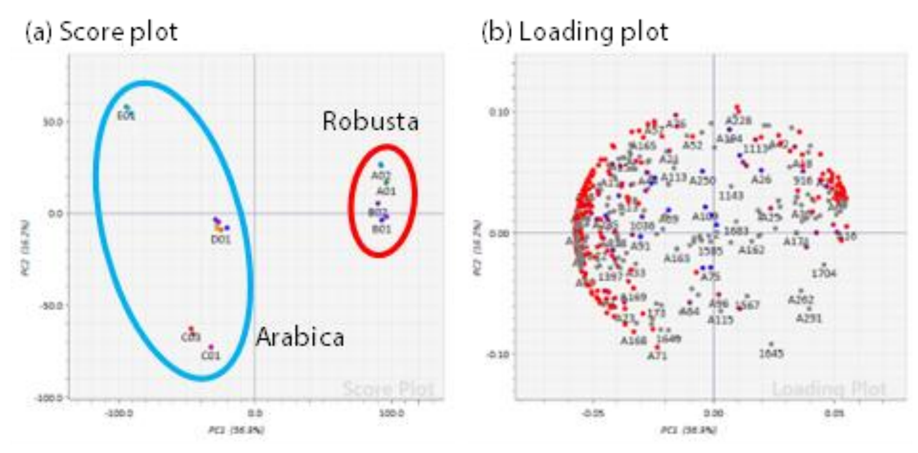

Fig. 3 PCA using scan data

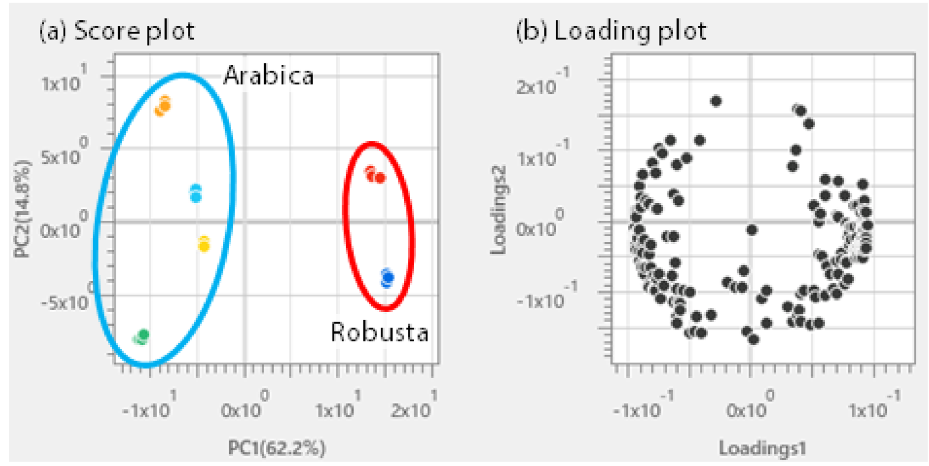

Fig. 4 PCA using MRM data

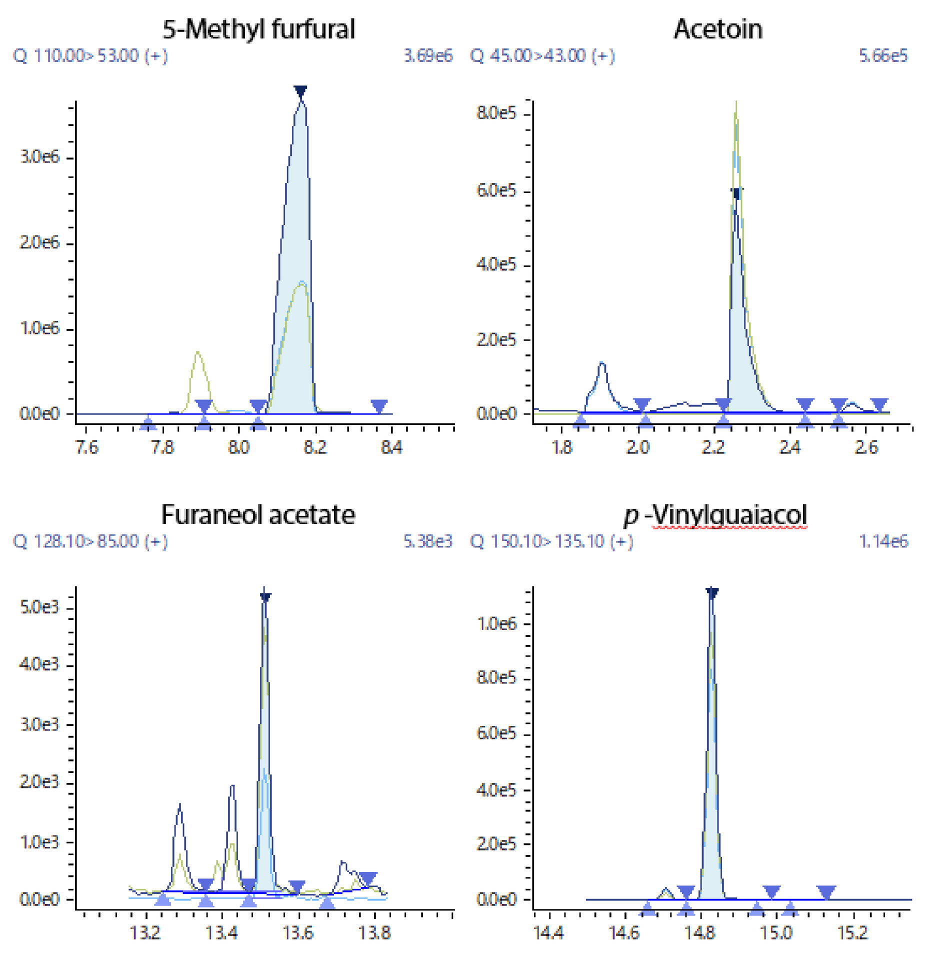

Fig. 5 MS chromatograms of representative compounds highly contributing to PC1

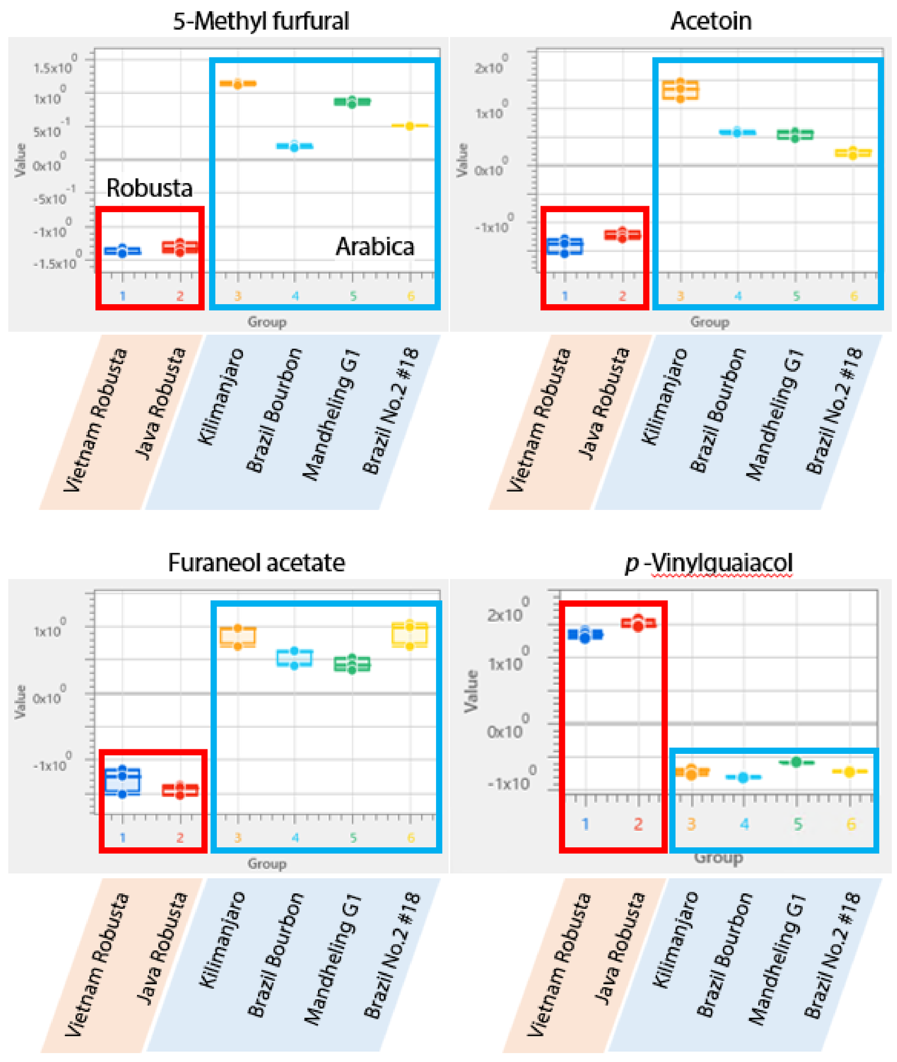

Fig. 6 Box plots of representative compounds highly contributing to PC1

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

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 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

3. Results

3.1. 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.

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.

3.2. 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. For the analysis, 175 compounds without missing values were used. The results of PCA are shown in Fig. 4. In the PCA score plot, Arabica is clustered on the left side and Robusta on the right side. These results suggest that discrimination between the two groups is also possible using MRM data.

3.3. Exploration of Characteristic Aroma Compounds for Arabica and Robusta

Next, we investigated which compounds contribute to the discrimination between Arabica and Robusta. First, Arabica was commonly rich in compounds such as 5-Methyl furfural, Acetoin, and Furaneol acetate (Table 3). MS chromatograms analyzed using LabSolutions Insight GCMS are shown in Fig. 5.

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. 6), 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.

3.4. 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.

4. Conclusions

- **Clear Group Separation**: Statistical analysis (PCA and cluster analysis) of data acquired in highly sensitive MRM mode clearly separated the two coffee bean types into distinct groups.
- **Successful Discrimination**: Our investigation suggested that the discrimination of Arabica and Robusta coffee beans based on aroma compounds is possible using the GCMS-TQ8040 NX, the Smart Aroma Database, and eMSTAT Solution.
- **Intuitive and Comprehensive Workflow**: Using eMSTAT Solution, statistical analysis and the creation of discrimination models can be easily performed without advanced expertise. Additionally, this method provides a comprehensive workflow that allows for the qualitative identification of key compounds contributing significantly to the discrimination.

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