JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

# $^{13}$C NMR-Based Metabolomics for the Classification of Green Coffee Beans According to Variety and Origin

Feifei Wei,[†,‡] Kazuo Furihata,[†] Masanori Koda,[†] Fangyu Hu,[†] Rieko Kato,[†] Takuya Miyakawa,[†] and Masaru Tanokura*[,†]

[†]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

[‡]Japan Society for the Promotion of Science, 8 Ichiban-cho, Chiyoda-ku, Tokyo 102-8472, Japan

**ABSTRACT:** $^{13}$C NMR-based metabolomics was demonstrated as a useful tool for distinguishing the species and origins of green coffee bean samples of arabica and robusta from six different geographic regions. By the application of information on $^{13}$C signal assignment, significantly different levels of 14 metabolites of green coffee beans were identified in the classifications, including sucrose, caffeine, chlorogenic acids, choline, amino acids, organic acids, and trigonelline, as captured by multivariate analytical models. These studies demonstrate that the species and geographical origin can be quickly discriminated by evaluating the major metabolites of green coffee beans quantitatively using $^{13}$C NMR-based metabolite profiling.

**KEYWORDS:** *$^{13}$C NMR, green coffee bean, PCA, OPLS-DA, species, origins*

Downloaded via UNIV NACIONAL AUTONOMA MEXICO on August 31, 2021 at 22:24:31 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

## INTRODUCTION

Coffee is one of the most important internationally traded products. Among all known coffee species, the most common are *Coffea arabica* L. (arabica) and *Coffea canephora* Pierre (robusta). Robusta has been characterized as a neutral coffee, weak-flavored, and occasionally with a strong and pronounced bitterness,[1] whereas arabica is higher-priced, milder, fruitier, and acidulous.[2] The quality of coffee depends strongly on species and geographic origin, with consequent wide variations in its commercial value. The importance of the coffee market and its globalization have increased concern about species and origin, and producers have responded by offering products with origin labeling to the consumer. Therefore, it is very important to guarantee the authenticity of species and geographic origins of coffee beans.

A variety of analytical techniques with various degrees of sensitivity and specificity, such as gas chromatography (GC),[3] high-performance liquid chromatography (HPLC),[4,5] isotope-ratio mass spectrometry (IRMS),[6] visible micro-Raman spectroscopy,[7] ultraviolet−visible absorption spectroscopy (UV−vis),[8] elemental analysis-like inductively coupled plasma−atomic emission spectrometry (ICP-AES),[9] and liquid chromatography−mass spectrometry (LC-MS),[10−12] have been conducted during the past decade to find chemical components that can be used to discriminate coffee species and origins. However, all of these techniques are compound-targeted; they can each assess differences in only one component or one class of components.

Recently, modern nuclear magnetic resonance (NMR) spectroscopy, with an informative, nondestructive, and non-targeted nature,[13,14] coupled with a multivariate analysis such as principal component analysis (PCA) or orthogonal partial least squares discriminated analysis (OPLS-DA), has been applied to obtain metabolite profiles of various kinds of biofluids[15,16] and foods, including honey,[17] meat,[19] mango juice,[19] tea,[20−22] wine,[23] and cheese.[24] In previous works we characterized green and roasted coffee bean extracts by NMR and monitored the roasting process by observing the time course of all the NMR-visible components.[25−27] NMR combined with PCA has been used in the discrimination of instant coffee among different producers by Charlton et al.,[28] as well as in the discrimination of roasted coffee beans from different geographical origins by Consonni et al.[29] However, in these studies, roasted coffee bean beverages rather than green coffees were analyzed, and the signal assignment was incomplete so that further assignments would be necessary to obtain more information from the NMR spectra. Thus, a global profiling of green coffee beans of different species and origins by NMR with detailed assignment information has not yet been reported. The detailed analysis of metabolites in various green coffee beans with the assignment information is needed to reveal and explain the differences between different species and origins.[25]

In the present study, $^{13}$C NMR spectroscopy with detailed assignment information, coupled with PCA and OPLS-DA models, was applied to distinguish the species and origins of green coffee beans and to identify significantly different metabolites between species and origins.

## MATERIALS AND METHODS

**Coffee Beans.** As shown in Table 1, we used arabica coffee beans from four origins (Brazil, Colombia, Guatemala, and Tanzania) and robusta coffee beans from two origins (Indonesia and Vietnam), which were kindly supplied by Ajinomoto General Foods, Inc. (Tokyo, Japan). The green coffee beans were frozen at −30 °C until analyzed.

**NMR Samples.** The green coffee beans were ground into grains about 1−2 mm in size using a Kalita C-120 coffee mill (Kalita Co., Ltd., Tokyo, Japan). The crushed beans (1.5 g) were incubated at 95

Received: April 25, 2012
Revised: September 17, 2012
Accepted: September 19, 2012
Published: September 19, 2012

 © 2012 American Chemical Society 

**Table 1. Species and Origins of Green Coffee Beans Used in the Present Study**

| species | origin | year of harvest | no. of batches |
|---|---|---|---|
| *arabica* | Brazil, South America | 2010, 2012 | 10 |
| | Colombia, South America | 2010, 2012 | 10 |
| | Guatemala, Central America | 2010, 2012 | 10 |
| | Tanzania, Africa | 2010, 2012 | 10 |
| *robusta* | Indonesia, Asia | 2010, 2012 | 10 |
| | Vietnam, Asia | 2010, 2012 | 10 |

°C in a closed plastic tube with $D_2O$ (3.50 mL, 99.7%; Shoko Co., Ltd., Tokyo, Japan) for 1 h. The extracts were cooled on ice for 15 min and then centrifuged at 5000*g* at 4 °C for 5 min. The supernatant (500 μL) was removed to a new tube and mixed with phosphate buffer (100 μL, 0.2 M sodium phosphate, pH 6.0). 4,4-Dimethyl-4-silapentane-1-sulfonate (DSS; Wako Pure Chemical Industries, Ltd., Osaka, Japan) was used as the internal reference, and its chemical shift was set to 0 ppm. The green coffee bean extracts (GCBE) were then transferred into 5 mm NMR tubes.[25]

**NMR Spectroscopic Analysis.** The one-dimensional (1D) $^1H$ and $^{13}C$ NMR spectra were measured at 500 and 125.65 MHz, respectively, on a Varian Unity INOVA-500 spectrometer. For the $^1H$ NMR spectra, the $H_2O$ signal was suppressed by the presaturation method, and the parameters for observation were as follows: number of data points, 64K; spectral width, 8000 Hz; acquisition time, 4.00 s; delay time, 2.0 s; number of scans, 128. The parameters for the $^{13}C$ NMR spectra were as follows: number of data points, 64K; spectral width, 31422 Hz; acquisition time, 1.04 s; delay time, 2.0 s; number of scans, 15000.

**NMR Data Processing.** The free-induction decay (FID) NMR data were processed by the program MestRe Nova (version 5.3.0; MestReC, Santiago de Compostela, Spain).

**NMR Signal Assignments.** The signal assignments of the components in GCBE have been carried out on the basis of analysis of two-dimensional (2D) $^1H-^1H$ DQF-COSY, $^1H-^{13}C$ HSQC, and $^1H-^{13}C$ CT-HMBC NMR spectra, and the results have been previously published.[25]

**Multivariate Data Analysis.** The $^{13}C$ NMR spectral data were reduced into 1 ppm spectral buckets, and all spectra were aligned and normalized by MestRe Nova. The resulting data sets were then imported into SIMCA-P version 12.0 (Umetrics, Umeå, Sweden) for further multivariate statistical analysis.

Prior to PCA, data were mean-centered and then scaled using Pareto or UV scaling. Hotelling's T2 region, shown as an ellipse in the score plots, defined the 99% confidence interval of the modeled variation. The quality of the model was described by $Rx^2$ and $Q^2$ values. $Rx^2$ was defined as the proportion of variance in the data explained by the model and indicates goodness of fit. $Q^2$ was defined as the proportion of variance in the data predictable by the model and indicates predictability.

To maximize the separation among samples, OPLS-DA was applied, which is described as the regression extension of PCA, an unsupervised pattern recognition method, giving the maximum covariance between the measured data ($X$) and the response variable ($Y$). For the OPLS-DA model, the confidence level for membership probability was considered to be 95%; observations at <5% are considered to be outliers. The overall predictive ability of the model is assessed by cumulative $Q^2$ representing the fraction of the variation of $Y$ that can be predicted by the model, which was extracted according to the internal cross-validation default method of SIMCA-P software. The method we used appears parsimonious and robust, with general applicability to data mining from metabolomic and similar data.

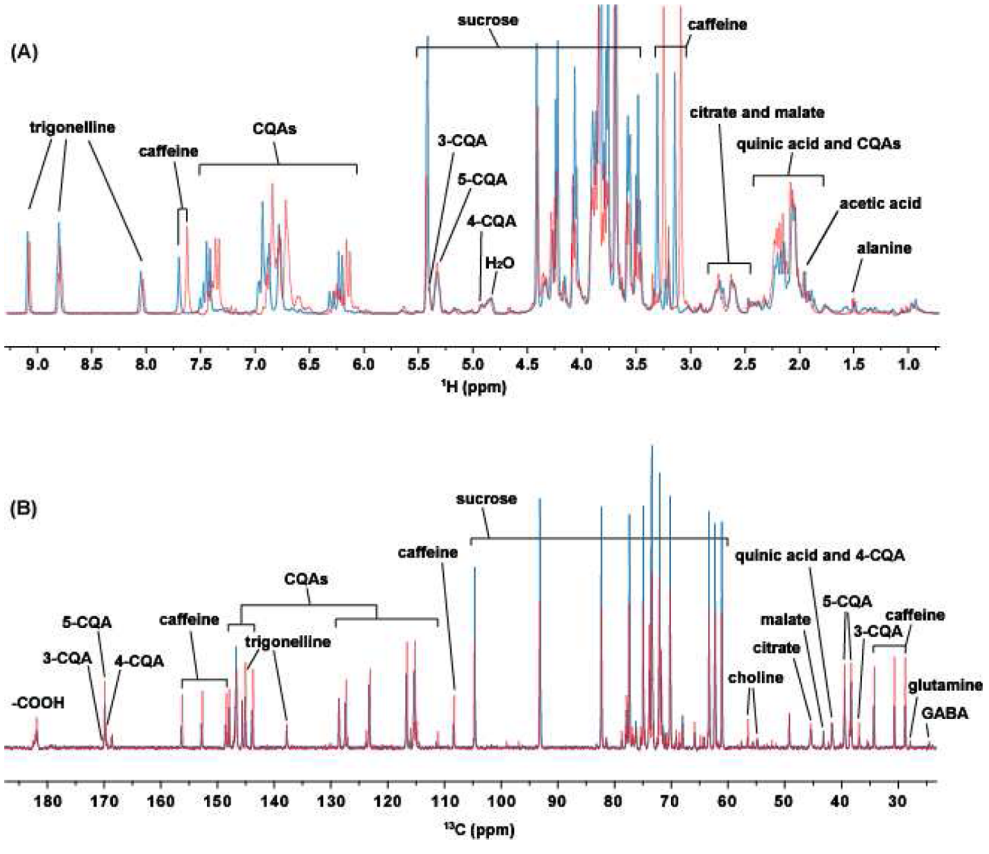


**Figure 1.** Assigned (A) $^1H$ and (B) $^{13}C$ NMR spectra of two typical green coffee bean extracts of the arabica from Colombia (blue line) and the robusta from Indonesia (red line).

## RESULTS AND DISCUSSION

**$^{1}$H and $^{13}$C NMR Spectra of Green Coffee Bean Extracts of Arabica and Robusta.** Figure 1 shows representative 1D $^{1}$H and $^{13}$C NMR spectra of aqueous GCBE samples of arabica from Colombia and robusta from Indonesia. The 1D NMR spectra of arabica and robusta give similar signal patterns, which contain the resonances from sucrose, caffeine, caffeoylquinic acids (CQAs), trigonelline, choline, *myo*-inositol, quinic acid, citrate, malate, acetic acid, and amino acids.

As shown in Figure 1A, the overall $^{1}$H NMR fingerprints between the arabica and the robusta samples are similar. However, close inspection of the spectra revealed that the chemical shifts due to caffeine and CQAs, the most abundant chlorogenic acids in coffee beans,[30] were shifted remarkably between species. According to a previous study, caffeine interacts with chlorogenic acid molecules in aqueous solution to form the caffeine−chlorogenate complex, the chemical shifts of which change with the relative concentrations of the two components.[31] Although the phosphate buffer was used to remove the pH variation among samples, the concentration-related changes in chemical shifts due to the caffeine−chlorogenate complex could not be removed by sample preparation. The concentration-related changes in chemical shifts induced new overlapping signal patterns, especially in the regions of 3.0−3.4 and 6.5−7.0 ppm, which led to large errors in signal recognition, alignment, and binning of data processing for PCA. Furthermore, owing to the strong signal overlapping of $^{1}$H resonances, which increased the difficulty of and frequency of errors in the identification of variables in PCA and OPLS-DA models, the informative $^{13}$C NMR spectra, the signals of which are well separated, were used in the classification of green coffee beans according to their species and origins.

Figure 1B shows the $^{13}$C NMR spectra of arabica and robusta GCBE samples. The signals were narrow and less overlapped than those in the $^{1}$H NMR spectrum. Although no clear apparent differences were observed in the overall spectral patterns, the chemical compositions of metabolites obtained from the different species were distinctly different. The $^{1}$H resonances due to the caffeine−chlorogenate complex were shifted remarkably when the relative concentrations varied between species, but no such effect was found in the $^{13}$C NMR spectra. The reason for this may be that the noncovalent correlations between caffeine and chlorogenic acids are not strong and thus affect the chemical environments in $^{1}$H nuclei but not those in $^{13}$C nuclei.[26] Therefore, the present $^{13}$C NMR spectroscopy was considered to be useful for metabolomics of GCBE. As described in a previous study, $^{13}$C NMR spectroscopy would be useful in metabolomics by providing complementary component information while potentially reducing the problems of overlap that occur in $^{1}$H NMR spectroscopy.[32] The problem of long $T_1$ relaxation times reduces signal intensities, and hence the quantification can be problematic. Nevertheless, for metabolomic studies in which all samples are measured under identical conditions, such quantification is less necessary if the $T_1$ times remain constant, because it is the overall pattern of response that can be interpreted.[27]

**Multivariate Statistical Analysis of Green Coffee Bean Extracts by $^{13}$C NMR Spectroscopy.** To see the differences in chemical components in GCBE with respect to different species and origins, six different GCBE samples, as listed in Table 1, were analyzed. PCA is an unsupervised classification method requiring no a priori knowledge of the data set and acts to reduce the dimensionality of multivariate data while preserving most of the variance within it.[33] The PCA score plot with high statistical values of $Rx^2$ (0.913) and $Q^2$ (0.853) is given in Figure 2A, which is derived from the $^{13}$C NMR spectra of GCBE of two species and from the six countries. $Rx^2$ represents the goodness of fit, and $Q^2$ reveals the predictability of the PCA model.[18] The PCA models using projections into two dimensions (PC1 of 0.62 and PC2 of 0.11) show statistically significant separation among the GCBE samples,

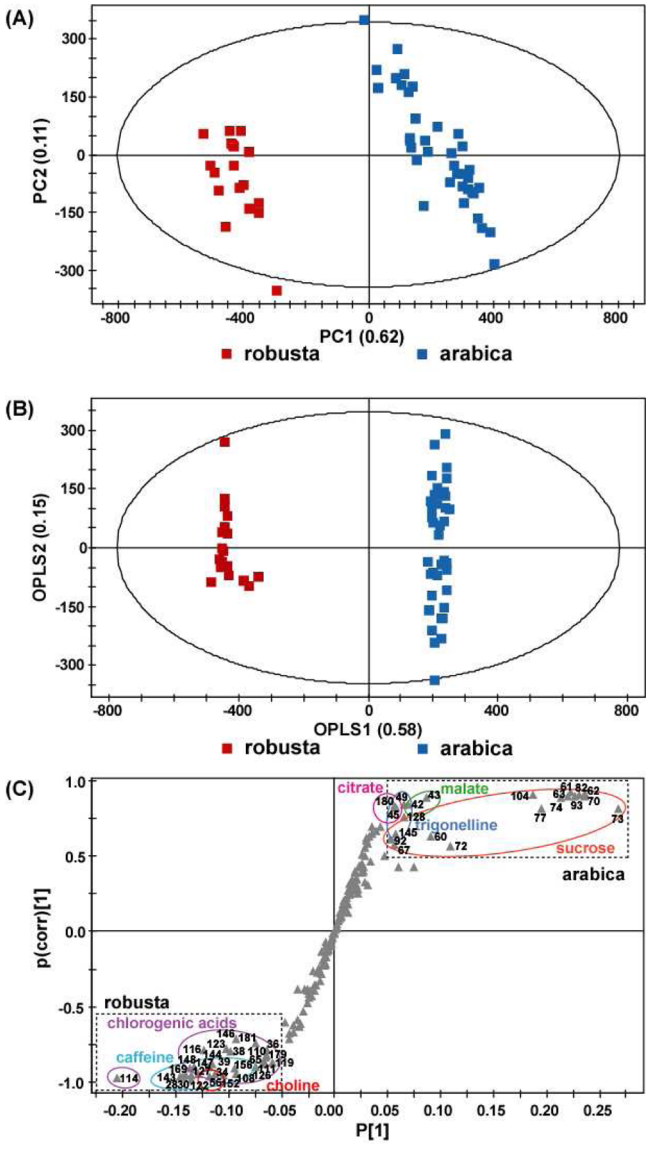


**Figure 2.** (A) PCA and (B) OPLS-DA score plots derived from the $^{13}$C NMR spectra of all GCBEs of arabica and robusta; (C) S-plot generated from the OPLS-DA model. The range of the variables selected is highlighted with a dotted rectangle. Cutoff values for the covariance of $|p| \geq 0.05$ and for the correlation of $|p(\mathrm{corr})| \geq 0.5$ were used. The variables in dotted rectangles represent the metabolites responsible for differentiation in OPLS-DA score plots; the names of metabolites corresponding to the variable are sucrose, trigonelline, malate, and citrate for arabica and chlorogenic acids, caffeine, and choline for robusta.

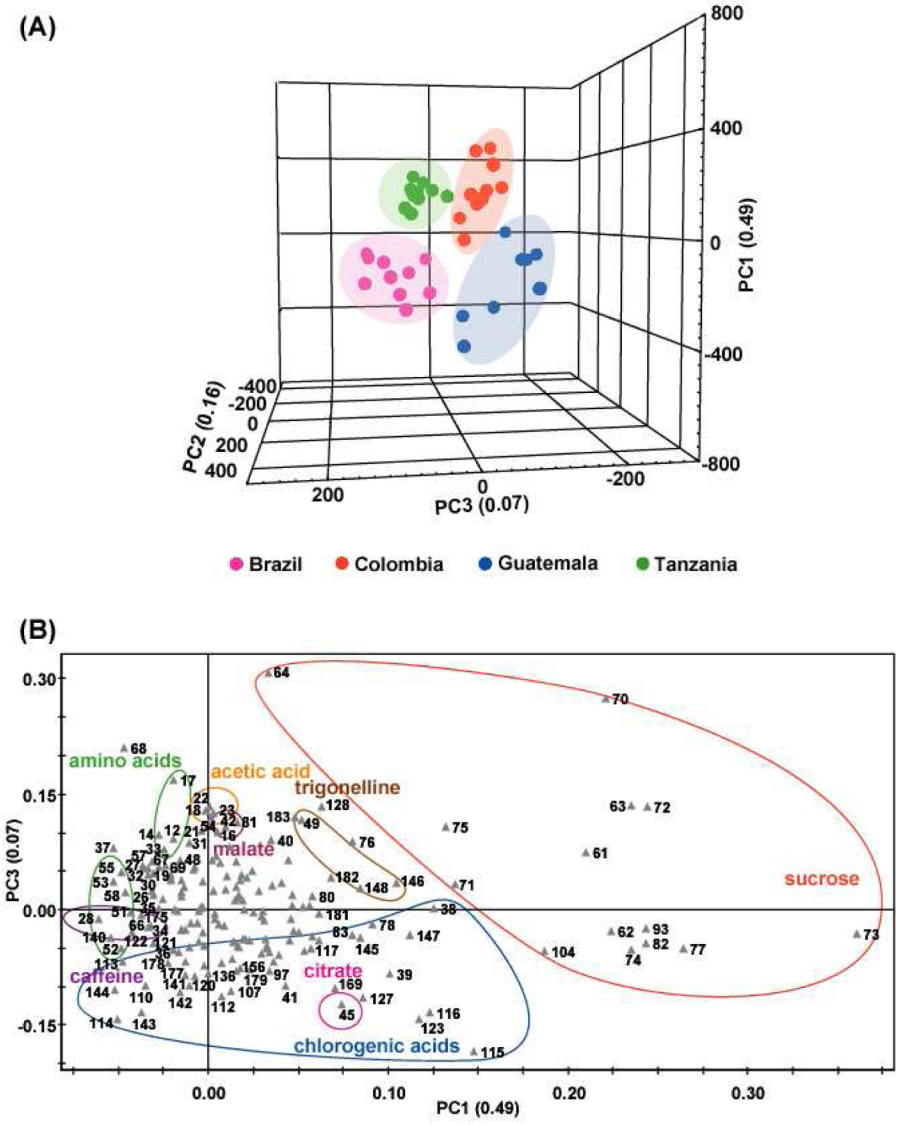


**Figure 3.** PCA (A) 3D score plot and (B) loading scatter plot (PC1 and PC3) derived from the $^{13}$C NMR spectra of GCBEs of the arabica from four different origins: Brazil, Colombia, Guatemala, and Tanzania.

indicating that the most significant differences in metabolite composition are between species. To clearly identify the significantly correlated metabolites between species, OPLS-DA modeling was applied to the total data set. The OPLS model (see Figure 2B), established the use of two predictive components and one orthogonal component and revealed very high statistical values: OPLS1 of 0.58, OPLS2 of 0.15, $Rx^2$ of 0.792, $Ry^2$ of 0.992, and $Q^2$ of 0.988. Figure 2B shows clear separation between arabica and robusta GCBE. To further understand the underlying variables contributing to the differentiation, we constructed the S-plot from the OPLS-DA model as shown in Figure 2C. The S-plot visualizes both the covariance $p$ and correlation $p(\mathrm{corr})$ between the metabolites and the modeled class designation.[34] The S-plot helps identify statistically significant and potentially biochemically significant metabolites, on the basis of both contributions to the model and their reliability. The variables selected in the S-plot are highlighted with a dotted rectangle. Cutoff values for the covariance of $|p| \geq 0.05$ and for the correlation of $|p(\mathrm{corr})| \geq 0.5$ were used.[18,35] Therefore, variables in the dotted rectangles of Figure 2C contributed to the group separation and were considered to be statistically significant metabolites in arabica and robusta GCBEs. The significant variables were identified according to the $^{13}$C assignment information[25] as shown in Figure 2C. Compared with the robusta, the arabica contained significantly higher levels of sucrose, citrate, trigonelline, and malate, whereas the robusta was characterized by higher levels of CQAs, caffeine, and choline.

To investigate the metabolic differences among different origins, PCA modeling was applied to the arabica data set. As shown in Figure 3A, GCBEs of arabica from Colombia, Brazil, Guatemala, and Tanzania were clearly distinguished from one another using a three-dimensional (3D, PC1 of 0.49, PC2 of 0.16, and PC3 of 0.07) PCA score plot with high statistical values of $Rx^2$ (0.810) and $Q^2$ (0.705). The loading scatter plot of PC1 and PC3 is shown in Figure 3B, because PC1 coupled with PC3 provided a clear classification according to origin. According to the assignment of $^{13}$C NMR signals due to GCBE components, some significant variables were captured and highlighted, and these are responsible for the differentiation in the PCA score plot. As shown in Figure 3B, the top right section of the loading plot, characterized by PC1 > 0 and PC3


dx.doi.org/10.1021/jf3033057 | J. Agric. Food Chem. 2012, 60, 10118−10125

> 0, indicates the relatively high levels of sucrose, acetic acid, and trigonelline in GCBEs from Tanzania. The bottom right section, characterized by PC1 > 0 and PC3 < 0, reveals the relatively high levels of CQAs, citrate, and sucrose in GCBEs from Colombia. The bottom left section, characterized by PC1 < 0 and PC3 < 0, represents the variables for caffeine, showing that GCBEs from Guatemala contain more caffeine than other arabica GCBEs. The top left section, characterized by PC1 < 0 and PC3 > 0, reveals the relatively high levels of amino acids in GCBEs from Brazil.

As to GCBEs of the robusta, PCA modeling was first applied to confirm if the differences due to different origins could be detected. As shown in Figure 4A, a PCA model constructed with two principal components (PC1, 0.32; and PC2, 0.24) showed separation between the GCBEs from Indonesia and those from Vietnam by PC2. The statistical values of $Rx^2$ and $Q^2$ were 0.917 and 0.698, respectively. To clearly identify the significantly correlated metabolites between GCBEs from Indonesia and Vietnam, OPLS-DA modeling was applied. The OPLS model (see Figure 4B), which is established by two predictive components and one orthogonal component, gave very high statistical values: OPLS1 of 0.20, OPLS2 of 0.23, $Rx^2$ of 0.821, $Ry^2$ of 0.998, and $Q^2$ of 0.991. Figure 4B shows clear separation between GCBEs from Indonesia and Vietnam. The S-plot from the OPLS-DA model was constructed to further clarify the underlying variables contributing to the differentiation (see Figure 2C). The variables chosen in the S-plot are highlighted with a dotted rectangle. Variables representing trigonelline, citrate, malate, 5-CQA, and 4-CQA were captured as the characteristic components in GCBEs from Indonesia. The highlighted variable on the left corresponds to caffeine and 3-CQA according to the assignment information, indicating relatively high levels in GCBEs from Vietnam.

In these analyses of the separation of GCBEs, it is possible that, besides species and origins, other parameters such as the altitude of the plantation and processing techniques may also affect the separation results of GCBEs by affecting other PCs in the PCA models. However, only the differentiation due to species and origins was picked up and identified in this study.

Figure 5 summarizes the significantly different metabolites captured by PCA and OPLS-DA models in the discrimination of GCBEs according to their species and origins. The levels of the identified metabolites differed dramatically among GCBE samples of different species and from different origins. As shown in Figure 5, GCBEs of the arabica showed higher levels of sucrose than those of the robusta by the present $^{13}C$ NMR-based metabolomics. Sucrose is a compound in raw coffee beans that has been implicated as an important precursor of coffee flavor and aroma.[36] In terms of cup quality, arabica is appreciated to a greater extent by consumers because it is less bitter and tastes better than the robusta. The reason the robusta accumulates less sucrose than the arabica is that the former has higher sucrose synthase and acid invertase activities early in grain development but less capacity for sucrose resynthesis at the final stages of grain development.[37,38] Among the arabica samples, GCBEs from Colombia and Tanzania showed higher levels of sucrose; this was captured by the PCA model in the present study. The present study showed higher levels of citrate, malate, and trigonelline in the arabica than in the robusta, which is consistent with previous studies.[39−42] The arabicas are valued for the impact that their higher acidity has on taste, whereas tasting notes for the robustas usually do not include this parameter. The level of citrate was higher in

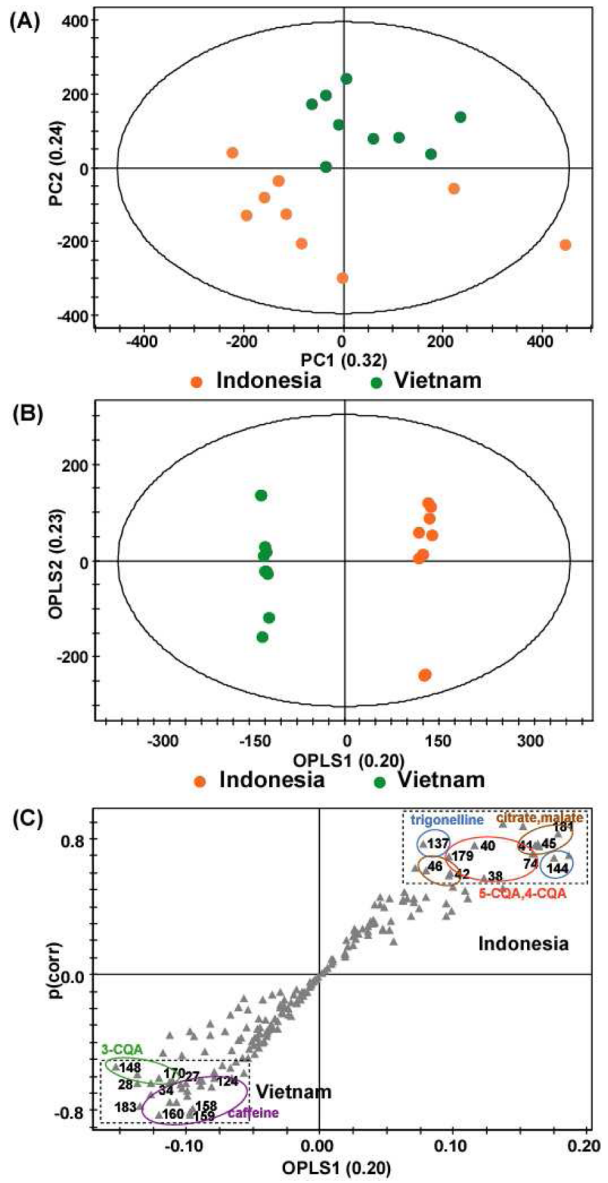


**Figure 4.** (A) PCA and (B) OPLS-DA score plots derived from the $^{13}C$ NMR spectra of GCBEs of the robusta from two different origins, Indonesia and Vietnam; (C) S-plot generated from the OPLS-DA model. The range of the variables selected is highlighted with a dotted rectangle. Cutoff values for the covariance and correlation are $|p| \geq 0.05$ and $|p(\mathrm{corr})| \geq 0.5$, respectively. The variables in the dotted rectangles represent the metabolites responsible for differentiation in the OPLS-DA score plots; the names of metabolites corresponding to the variables are trigonelline, citrate, malate, 5-CQA, and 4-CQA for beans from Indonesia and caffeine and 3-CQA for beans from Vietnam.

GCBEs from Indonesia than in those from Vietnam. This can be considered a marker for distinguishing green coffee beans from Indonesia or Vietnam. Trigonelline levels were higher in GCBEs of the arabica than in those of the robusta. This finding is consistent with the results of a previous study.[43] When only the arabicas were under consideration, the lowest level of trigonelline was in the GCBEs from Guatemala.

Caffeine, choline, and chlorogenic acids showed higher levels in the robusta than in the arabica, consistent with previous observations.[44,45] It has been reported that the patterns of fluctuations of the caffeine biosynthetic activity in *Coffea* plants

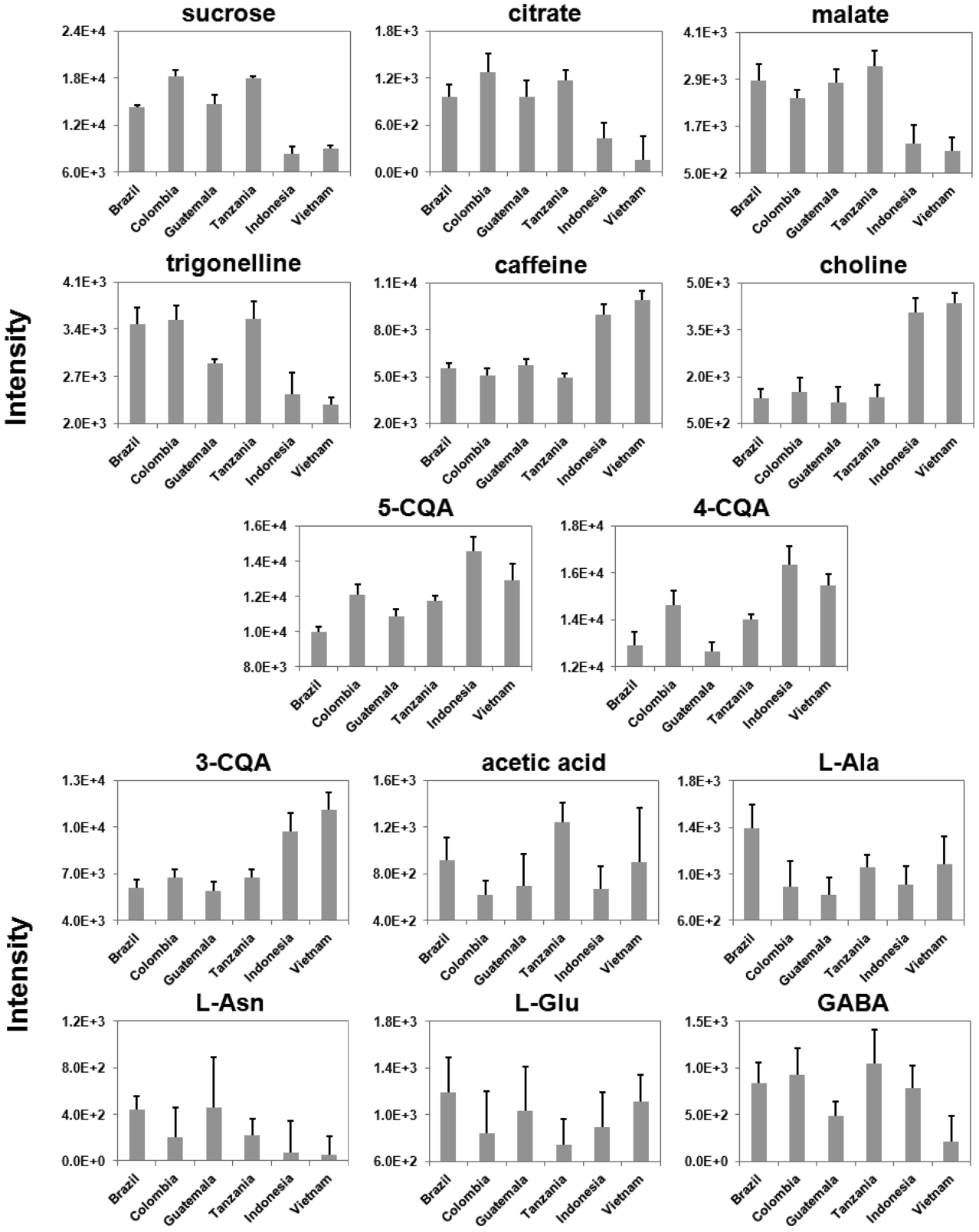


**Figure 5.** Significantly different metabolites captured by PCA and OPLS-DA models derived from the $^{13}C$ NMR spectra of all GCBEs: sucrose, Var ID 104; citrate, Var ID 45; malate, Var ID 43; trigonelline, Var ID 49; caffeine, Var ID 28; choline, Var ID 56; 5-CQA, Var ID 39; 4-CQA, Var ID 38; 3-CQA, Var ID 170; acetic acid, Var ID 23; L-Ala, Var ID 17; L-Asn, Var ID 52; L-Glu, Var ID 55; and GABA, Var ID 40.

are all similar, although the final concentration of caffeine varies among different species of coffee beans.[46] Caffeine showed higher levels in the robusta GCBEs from Vietnam than in those from Indonesia, which is consistent with the previous profiling by HPLC.[4] It has also been reported that the amounts of caffeine in green coffee beans from Central America (e.g., Guatemala) are higher than in beans from South America (e.g., Brazil and Colombia) and Africa (e.g, Tanzania).[4] In the


dx.doi.org/10.1021/jf3033057 | J. Agric. Food Chem. 2012, 60, 10118−10125

present study, the same tendency was observed. Furthermore, for beans from South America, beans from Brazil were observed as containing more caffeine than those from Colombia. Choline showed higher levels in the robusta than in the arabica in the present study. Coffee is the top contributor of free choline, which is an essential nutrient required for methyl group metabolism, and it has been related to the reduction of breast cancer risk.[47] Green coffee beans of the robusta contained higher amounts of CQAs. As for the arabica, the highest 5-CQA levels were detected in GCBEs from Colombia and the lowest levels from Brazil. According to the previous study that used HPLC, the highest 5-CQA levels were observed in beans from Africa and the lowest levels from South America.[4] It could be that although beans from Colombia contain the greatest amount of 5-CQA, the average amounts of beans from the two countries, Columbia and Brazil, of South America show the lowest levels of 5-CQA when compared to beans from Tanzania. The lowest levels of 4-CQA were detected in beans from Guatemala, and the highest levels of 3-CQA were observed in beans from Colombia, which are both consistent with a previous study.[4] The lower levels of CQAs in the arabica make these beans more vulnerable to phytopathogens as well as to biological and mechanical stress than the robusta beans.[4,48]

Acetic acid and amino acids also showed an important role in the origin discrimination of GCBEs. The GCBE with the highest level of acetic acid was from Tanzania in this study. GABA exists in higher levels in GCBEs from Tanzania than in those from any other origin, whereas L-glutamine was higher in GCBEs from Brazil and Guatemala. The arabica from Brazil also showed higher levels in L-alanine and L-asparagine. A previous study presented a discrimination methodology between the arabica and robusta coffee species on the basis of their amino acid enantiomers.[3] In the present study, the variation of amino acids contributed to the assessment of geographical origins of the arabica GCBEs. Currently, the reasons for the geographical differences in metabolomic profiles are not fully understood. Therefore, reliable analytical techniques and studies on the physiology of coffee beans are still required to confirm and guarantee the authenticity of the species and geographic origins of coffee beans.

In summary, the present study demonstrated that $^{13}C$ NMR-based metabolomics is a useful tool for distinguishing species and origins of GCBE samples and that its combination with chemometric analysis largely improves sample classification. By applying signal assignment information, significantly different metabolites captured by multivariate analysis were identified in the classification. Our study demonstrates that the species and geographical origin of green coffee beans can be rapidly discriminated by evaluating the major metabolites quantitatively using metabolite profiling.

## AUTHOR INFORMATION


### Corresponding Author

*Phone: +81 3 5841 5165. Fax: +81 3 5841 8023. E-mail: amtanok@mail.ecc.u-tokyo.ac.jp.



### Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Foreign Researchers.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Dr. Imura of Ajinomoto General Foods, Inc., for providing the green coffee beans used in this study.

## REFERENCES


(1) Illy, A.; Viani, R. Chapter 3: The plant. In *Espresso Coffee: The Chemistry of Quality*; Academic Press: San Diego, CA, 1995; pp 19—20.

(2) Bertrand, B.; Guyot, B.; Anthony, F.; Lashermes, P. Impact of the *Coffea canephora* gene introgression on beverage quality of *C-arabica*. *Theor. Appl. Genet.* **2003**, *107*, 387−394.

(3) Casal, S.; Alves, A. R.; Mendes, E.; Oliveira, M. B. P. P.; Ferreira, M. A. Discrimination between arabica and robusta coffee species on the basis of their amino acid enantiomers. *J. Agric. Food Chem.* **2003**, *51*, 6495−6501.

(4) Alonso-Salces, R. M.; Serra, F.; Reniero, F.; Heberger, K. Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): chemometric evaluation of phenolic and methylxanthine contents. *J. Agric. Food Chem.* **2009**, *57*, 4224−4235.

(5) Alves, R. C.; Casal, S.; Alves, M. R.; Oliveira, M. B. Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. *Food Chem.* **2009**, *114*, 295−299.

(6) Rodrigues, C. I.; Maia, R.; Miranda, M.; Ribeirinho, M.; Nogueira, J. M. F.; Maguas, C. Stable isotope analysis for green coffee bean: a possible method for geographic origin discrimination. *J. Food Compos. Anal.* **2009**, *22*, 463−471.

(7) El-Abassy, R. M.; Donfack, P.; Materny, A. Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis. *Food Chem.* **2011**, *126*, 1443−1448.

(8) Silva, E. C.; Souto, U. T. C. P.; Pontes, M. J. C.; Galvao, R. K. H.; Araujo, M. C. U.; Sanches, F. A. C.; Cunha, F. A. S.; Oliveira, M. S. R. UV-Vis spectrometric classification of coffees by SPA-LDA. *Food Chem.* **2010**, *119*, 368−371.

(9) Anderson, K. A.; Smith, B. W. Chemical profiling to differentiate geographic growing origins of coffee. *J. Agric. Food Chem.* **2002**, *50*, 2068−2075.

(10) Choi, M. Y.; Choi, W.; Park, J. H.; Lim, J.; Kwon, S. W. Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. *Food Chem.* **2010**, *121*, 1260−1268.

(11) Kuhnert, N.; Jaiswal, R.; Eravuchira, P.; El-Abassy, R. M.; von der Kammer, B.; Materny, A. Scope and limitations of principal component analysis of high resolution LC-TOF-MS data: the analysis of the chlorogenic acid fraction in green coffee beans as a case study. *Anal. Methods* **2011**, *3*, 144−155.

(12) Jaiswal, R.; Patras, M. A.; Eravuchira, P. J.; Kuhnert, N. Profile and characterization of the chlorogenic acids in green robusta coffee beans by LC-MS$^n$: identification of seven new classes of compounds. *J. Agric. Food Chem.* **2010**, *58*, 8722−8737.

(13) Hu, F.; Furihata, K.; Ito-Ishida, M.; Kaminogawa, S.; Tanokura, M. Nondestructive observation of bovine milk by NMR spectroscopy: analysis of existing states of compounds and detection of new compounds. *J. Agric. Food Chem.* **2004**, *52*, 4969−4974.

(14) Hu, F.; Furihata, K.; Kato, Y.; Tanokura, M. Nondestructive quantification of organic compounds in whole milk without pretreatment by two-dimensional NMR spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 4307−4311.

(15) Zhou, A. G.; Liu, G. M.; Wang, Y.; Wang, Z. S.; Cai, J. Y.; Lv, X. Z. Nuclear magnetic resonance (NMR)-based metabolomic studies on urine and serum biochemical profiles after chronic cysteamine supplementation in rats. *J. Agric. Food Chem.* **2011**, *59*, 5572−5578.

(16) Qin, X. Y.; Wei, F.; Yoshinaga, J.; Yonemoto, J.; Tanokura, M.; Sone, H. siRNA-mediated knockdown of aryl hydrocarbon receptor nuclear translocator 2 affects hypoxia-inducible factor-1 regulatory signaling and metabolism in human breast cancer cells. *FEBS Lett.* **2011**, *585*, 3310−3315.

(17) Consonni, R.; Cagliani, L. R.; Cogliati, C. NMR characterization of saccharides in Italian honeys of different floral sources. *J. Agric. Food Chem.* **2012**, *60*, 4526−4534.
(18) Jung, Y.; Lee, J.; Kwon, J.; Lee, K.-S.; Ryu, D. H.; Hwang, G.-S. Discrimination of the geographical origin of beef by (1)H NMR-based metabolomics. *J. Agric. Food Chem.* **2010**, *58*, 10458−10466.
(19) Koda, M.; Furihata, K.; Wei, F. F.; Miyakawa, T.; Tanokura, M. Metabolic discrimination of mango juice from various cultivars by band-selective NMR spectroscopy. *J. Agric. Food Chem.* **2012**, *60*, 1158−1166.
(20) Van Dorsten, F. A.; Daykin, C. A.; Mulder, T. P. J.; Van Duynhoven, J. P. M. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J. Agric. Food Chem.* **2006**, *54*, 6929−6938.
(21) Ku, K. M.; Choi, J. N.; Kim, J.; Kim, J. K.; Yoo, L. G.; Lee, S. J.; Hong, Y. S.; Lee, C. H. Metabolomics analysis reveals the compositional differences of shade grown tea (*Camellia sinensis* L.). *J. Agric. Food Chem.* **2010**, *58*, 418−426.
(22) Ohno, A.; Oka, K.; Sakuma, C.; Okuda, H.; Fukuhara, K. Characterization of tea cultivated at four different altitudes using (1)H NMR analysis coupled with multivariate statistics. *J. Agric. Food Chem.* **2011**, *59*, 5181−5187.
(23) Mikros, E.; Anastasiadi, M.; Zira, A.; Magiatis, P.; Haroutounian, S. A.; Skaltsounis, A. L. (1)H NMR-based metabonomics for the classification of Greek wines according to variety, region, and vintage. comparison with HPLC data. *J. Agric. Food Chem.* **2009**, *57*, 11067−11074.
(24) Rodrigues, D.; Santos, C. H.; Rocha-Santos, T. A. P.; Gomes, A. M.; Goodfellow, B. J.; Freitas, A. C. Metabolic profiling of potential probiotic or synbiotic cheeses by nuclear magnetic resonance (NMR) spectroscopy. *J. Agric. Food Chem.* **2011**, *59*, 4955−4961.
(25) Wei, F. F.; Furihata, K.; Hu, F. Y.; Miyakawa, T.; Tanokura, M. Complex mixture analysis of organic compounds in green coffee bean extract by two-dimensional NMR spectroscopy. *Magn. Reson. Chem.* **2010**, *48*, 857−865.
(26) Wei, F.; Furihata, K.; Hu, F.; Miyakawa, T.; Tanokura, M. Two-dimensional (1)H-(13)C nuclear magnetic resonance (NMR)-based comprehensive analysis of roasted coffee bean extract. *J. Agric. Food Chem.* **2011**, *59*, 9065−9073.
(27) 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*, 1005−1012.
(28) Charlton, A. J.; Farrington, W. H. H.; Brereton, P. Application of H-1 NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee. *J. Agric. Food Chem.* **2002**, *50*, 3098−3103.
(29) Consonni, R.; Cagliani, L. R.; Cogliati, C. NMR based geographical characterization of roasted coffee. *Talanta* **2012**, *88*, 420−426.
(30) Clifford, M. N.; Johnston, K. L.; Knight, S.; Kuhnert, N. Hierarchical scheme for LC-MS*n* identification of chlorogenic acids. *J. Agric. Food Chem.* **2003**, *51*, 2900−2911.
(31) D'Amelio, N.; Fontanive, L.; Uggeri, F.; Suggi-Liverani, F.; Navarini, L. NMR Reinvestigation of the caffeine-chlorogenate complex in aqueous solution and in coffee brews. *Food Biophys.* **2009**, *4*, 321−330.
(32) Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskau, D.; Lindon, J. C.; Nicholson, J. K. Cryogenic probe C-13 NMR spectroscopy of urine for metabonomic studies. *Anal. Chem.* **2002**, *74*, 4588−4593.
(33) Choi, H. K.; Choi, Y. H.; Verberne, M.; Lefeber, A. W.; Erkelens, C.; Verpoorte, R. Metabolic fingerprinting of wild type and transgenic tobacco plants by $^{1}$H NMR and multivariate analysis technique. *Phytochemistry* **2004**, *65*, 857−864.
(34) Wiklund, S.; Johansson, E.; Sjostrom, L.; Mellerowicz, E. J.; Edlund, U.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* **2008**, *80*, 115−122.
(35) Sieber, M.; Wagner, S.; Rached, E.; Amberg, A.; Mally, A.; Dekant, W. Metabonomic study of ochratoxin a toxicity in rats after repeated administration: phenotypic anchoring enhances the ability for biomarker discovery. *Chem. Res. Toxicol.* **2009**, *22*, 1221−1231.
(36) Demaria, C. A. B.; Trugo, L. C.; Moreira, R. F. A.; Werneck, C. C. Composition of green coffee fractions and their contribution to the volatile profile formed during roasting. *Food Chem.* **1994**, *50*, 141−145.
(37) Geromel, C.; Ferreira, L. P.; Guerreiro, S. M. C.; Cavalari, A. A.; Pot, D.; Pereira, L. F. P.; Leroy, T.; Vieira, L. G. E.; Mazzafera, P.; Marraccini, P. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *J. Exp. Bot.* **2006**, *57*, 3243−3258.
(38) Privat, I.; Foucrier, S.; Prins, A.; Epalle, T.; Eychenne, M.; Kandalaft, L.; Caillet, V.; Lin, C. W.; Tanksley, S.; Foyer, C.; McCarthy, J. Differential regulation of grain sucrose accumulation and metabolism in *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) revealed through gene expression and enzyme activity analysis. *New Phytol.* **2008**, *178*, 781−797.
(39) Ky, C. L.; Louarn, J.; Dussert, S.; Guyot, B.; Hamon, S.; Noirot, M. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C-canephora* P. accessions. *Food Chem.* **2001**, *75*, 223−230.
(40) Alcazar, A.; Fernandez-Caceres, P. L.; Martin, M. J.; Pablos, F.; Gonzalez, A. G. Ion chromatographic determination of some organic acids, chloride and phosphate in coffee and tea. *Talanta* **2003**, *61*, 95−101.
(41) Rodrigues, C. I.; Marta, L.; Maia, R.; Miranda, M.; Ribeirinho, M.; Maguas, C. Application of solid-phase extraction to brewed coffee caffeine and organic acid determination by UV/HPLC. *J. Food Compos. Anal.* **2007**, *20*, 440−448.
(42) Rogers, W. J.; Michaux, S.; Bastin, M.; Bucheli, P. Changes to the content of sugars, sugar alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of Robusta (*Coffea canephora*) and Arabica (*C. arabica*) coffees. *Plant Sci.* **1999**, *149*, 115−123.
(43) Casal, S.; Oliveira, M. B. P. P.; Alves, M. R.; Ferreira, M. A. Discriminate analysis of roasted coffee varieties for trigonelline, nicotinic acid, and caffeine content. *J. Agric. Food Chem.* **2000**, *48*, 3420−3424.
(44) Martin, M. J.; Pablos, F.; Gonzalez, A. G. Discrimination between arabica and robusta green coffee varieties according to their chemical composition. *Talanta* **1998**, *46*, 1259−1264.
(45) Briandet, R.; Kemsley, E. K.; Wilson, R. H. Discrimination of Arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics. *J. Agric. Food Chem.* **1996**, *44*, 170−174.
(46) Koshiro, Y.; Zheng, X. Q.; Wang, M. L.; Nagai, C.; Ashihara, H. Changes in content and biosynthetic activity of caffeine and trigonelline during growth and ripening of *Coffea arabica* and *Coffea canephora* fruits. *Plant Sci.* **2006**, *171*, 242−250.
(47) Xu, X. R.; Gammon, M. D.; Zeisel, S. H.; Lee, Y. L.; Wetmur, J. G.; Teitelbaum, S. L.; Bradshaw, P. T.; Neugut, A. I.; Santella, R. M.; Chen, J. Choline metabolism and risk of breast cancer in a population-based study. *FASEB J.* **2008**, *22*, 2045−2052.
(48) Macheix, J. J.; Fleuriet, A.; Billot, J. *Fruit Phenolics*; CRC Press: Boca Raton, FL, 1990.