Differentiation of Chocolates According to the Cocoa's Geographical Origin Using Chemometrics

Amandine Cambrai, Christophe Marcic,* Stéphane Morville,
Pierre Sae Houer, Françoise Bindler, and Eric Marchioni

Laboratoire de Chimie Analytique et Sciences de l'Aliment (IPHC-UMR 7178), Faculté de Pharmacie,
74 route du Rhin, 67400 Illkirch, France

The determination of the geographical origin of cocoa used to produce chocolate has been assessed through the analysis of the volatile compounds of chocolate samples. The analysis of the volatile content and their statistical processing by multivariate analyses tended to form independent groups for both Africa and Madagascar, even if some of the chocolate samples analyzed appeared in a mixed zone together with those from America. This analysis also allowed a clear separation between Caribbean chocolates and those from other origins. Height compounds (such as linalool or (*E,E*)-2,4-decadienal) characteristic of chocolate's different geographical origins were also identified. The method described in this work (hydrodistillation, GC analysis, and statistic treatment) may improve the control of the geographical origin of chocolate during its long production process.

KEYWORDS: Chocolate; geographical origin; volatile compounds; statistical multivariate analysis

INTRODUCTION

Originally from the tropical rainforest of South America, chocolate is now a product of common consumption in industrialized countries, and its popularity resides mainly in its particular aromas and flavors.

The volatile compounds of chocolate responsible for its aromas have already been extensively studied. More than 500 compounds divided into 17 distinct chemical classes have been identified in roasted cocoa bean samples (*1*). However, most of these studies concern the volatile compounds present in the cocoa bean and not in the final product, chocolate.

Even if the compounds responsible for the characteristic aroma of cocoa have been widely studied over the past century, these compounds are still of interest in view of various papers and studies published more recently (*2−6*). The compounds that give cocoa and chocolate their flavors are molecules such as pyrazines, aldehydes (cocoa aroma, nut), esters (fruity aroma), and phenolic compounds (astringent property). Most of these molecules are mainly generated during the Maillard reaction (*7*). Indeed, volatile compounds present in the fresh beans are transformed during the process, whereas other volatile compounds more characteristic of chocolate appear during the fermentation and roasting (*8*). Besides, the aromatic quality of the cocoa beans is linked to its original variety, to its postharvest treatment (fermentation, drying, and storage), and to the parameters used during the industrial transformation. The type of soil, climate, area of culture, and degree of pod maturity also influence the aromatic profile (*9*).

Consumers have no way of tracing the origin of the cocoa used to produce their chocolate to a particular country, much less a particular site of agricultural production. To determine the quality of chocolate and the veracity of labeling, consumers need to be informed of the cocoa production site's country, even more with the rising market of "healthy chocolate" (*10*).

Some authors have studied the influence of cocoa's origin on the composition in volatile compounds. Most studies concern the fermented, dried, and, most often, roasted beans (*11, 12*). The profile comparison and multivariate analyses allowed beans (roasted or not) from various geographical origins to be distinguished. Cocoas of different geographical origins have different organoleptic characteristics and influence chocolate's quality, which was tested by sensory analysis (*9, 13*).

The compounds responsible for the characteristic aroma of chocolate are strongly correlated with volatile compounds. If differences appear at the level of sensory analysis, they will also appear at the level of volatile compounds, which include aromas. Counet et al. (*14*) have shown that for cocoa liquor, the total content of volatile compounds varied for certain geographic origins. However, no study of volatile compounds allowing chocolates to be differentiated according to their geographical origin has been done at the moment. In view of the many thermal treatments performed during the manufacture of chocolate, some of the most volatile compounds are removed from the final product (*15*). However, chocolate has a high fat content, which has the property to trap volatile compounds. Therefore, only the most volatile ones are likely to disappear. Consequently, volatile compounds are still present in chocolate and can be analyzed by chromatographic methods (*16*).

To fulfill the quality requirements and the traceability, the industries and fraud control need methods that enable the assessment of

---

*Author to whom correspondence should be addressed (e-mail christophe.marcic@unistra.fr; telephone +33 3 68 85 41 60; fax +33 3 68 85 43 25).
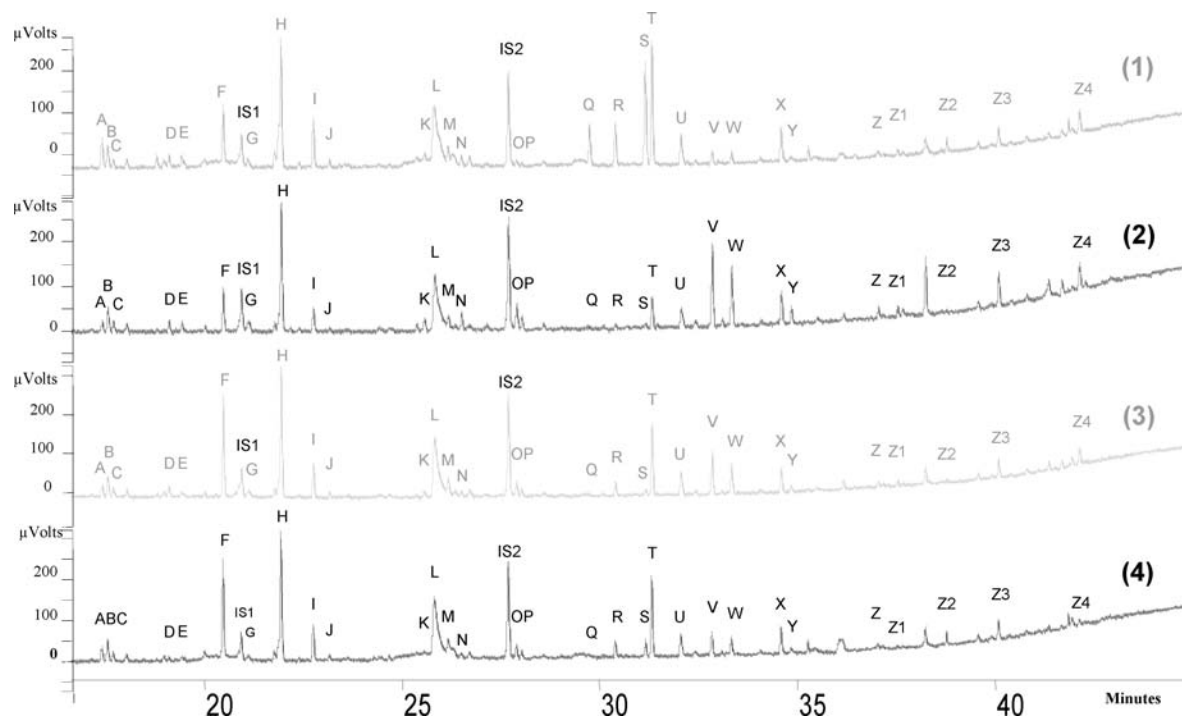
Figure 1. Selection of the 30 significant peaks, annotated from A to Z4, for 4 samples of different origins: (1) Trinidad; (2) Venezuela; (3) Ghana; (4) Madagascar (IS, internal standards). The chromatograms were obtained by GC-FID.

the origin of cocoa as a raw material of the final product, the chocolate. Thus, the aim of this work was to develop an analytical method to differentiate chocolate samples, according to their geographical origin, by analysis of their volatile compounds. Hydrodistillation was chosen as a simple extraction technique of these compounds, coupled with a separation by gas chromatography and mass spectrometry (GC-MS) detection or flame ionization detection (GC-FID). Two statistical analyses and a blind test were performed on the chemical composition to assess the country of origin of chocolate samples. Moreover, the chemical structures of seven compounds found in the chocolate samples were identified in order to have chemical tracers of the cocoa's origin.

MATERIALS AND METHODS

Materials. Fifty-one commercial dark chocolates of various origins (eight different countries) were supplied by Valrhona (Tain-l'Hermittage, France). Prior to analysis, they were hermetically stored sheltered from light at 18 °C in clean plastic bags.

Chemicals. Isooctane of HPLC grade was obtained from Acros Organics (Geel, Belgium). Ultrapure water was produced by a Milli-Q water system (Millipore, SAS, Molsheim, France). Standards of benzaldehyde, linalool, phenylacetaldehyde, (*E,E*)-2,4-nonadienal, ethyl phenylacetate, (*E,E*)-2,4-decadienal, 2-phenyl-2-butenal, and 4-methyl-2-phenyl-2-pentenal were purchased from Sigma-Aldrich (Steinheim, Germany). Nonanoic acid methyl ester and undecanoic acid methyl ester, used as internal standartds, were from Polyscience Corp. (Niles, IL).

Extraction of Volatile Compounds. Chocolate (30 g) was reduced to a powder with a mixer and then introduced into a round-bottom flask with 500 mL of ultrapure water. Two hundred microliters of a 100 ppm solution of nonanoic acid methyl ester (methyl nonanoate) in isooctane was added as a first internal standard (IS1). The volatile flavor compounds were then extracted and isolated using a hydrodistillation system composed of a heating mantle, a Vigreux column, and a thermostated condenser with circulating water. The distillate flask was placed in an ice bath to avoid volatile compounds losses. One hundred milliliters of the hydrodistillate was collected in about 40 min.

Ten milliliters of the solution obtained by distillation was placed in a 15 mL Pyrex tube and was extracted with 1 mL of isooctane. The tube was

hermetically sealed and vigorously stirred on vortex for 2 min. After liquid−liquid extraction, organic extracts were isolated and stored at 4 °C and analyzed no later than 2 days afterward. Fifty microliters of a 100 ppm solution of undecanoic acid methyl ester was added in 950 μL of extract in isooctane just before the GC injection as a second internal standard (IS2) to normalize the injection volume.

GC-FID. Measurements were performed with a Varian (Palo Alto, CA) 3300 equipped with a split−splitless injector and a flame ionization detector (FID). For the gas chromatographic separation, a polar DB-Wax column (J&W Scientific, Folsom, CA) was used (60 m × 0.32 mm, film thickness = 0.25 μm). The carrier gas was helium at a flow rate of 1 mL min^{-1}. Injections were done in splitless mode (split closed for 0.75 min).

The injector and the detector temperatures were kept at 200 and 220 °C, respectively. The oven temperature program was held at 50 °C for 1 min, varied between 50 and 165 °C at 3.8 °C min^{-1}, and raised to 220 °C at 5 °C min^{-1}. The temperature was kept at 220 °C for 20 min. One microliter of sample was injected by a Varian 3800 autosampler.

GC-MS. The analysis of volatiles was carried out using a Varian 3400 CX gas chromatograph with a split−splitless injector and coupled to a Varian Saturn 3 MS detector with an ion trap. The same GC conditions (column, temperature program, and carrier gas) as for GC-FID analysis were used. Injections (1 μL volume samples) were done manually, in the splitless mode (split closed for 0.75 min.).

The injector and the transfer line temperatures were kept at 200 and 220 °C, respectively. The samples were analyzed by electron impact ionization (EI) at 70 eV, and the mass detection was done in full-scan mode from m/z 30 to 500. The compounds were identified by a combination of the NIST and Saturn libraries of mass spectra and gas chromatographic retention time and mass spectra of pure compounds used as reference measurement standards.

Analytical blanks of solvent (isooctane) and blank method were carried out before every extract analysis.

Chemometrics. The variables used in the statistical analysis symbolize the area of the chromatographic peaks obtained by GC-FID normalized with the two internal standards. The data were processed with JMP software (version 5, SAS Institute, Cary, NC) (*17*) and StatBox software (version 6.5, Grimmer Logiciels, Paris, France).

Principal component analysis (PCA) was applied to a set of variables obtained from chocolates made from cocoa of different origins. PCA is based on a vector space transform that reduces the dimensionality in a data
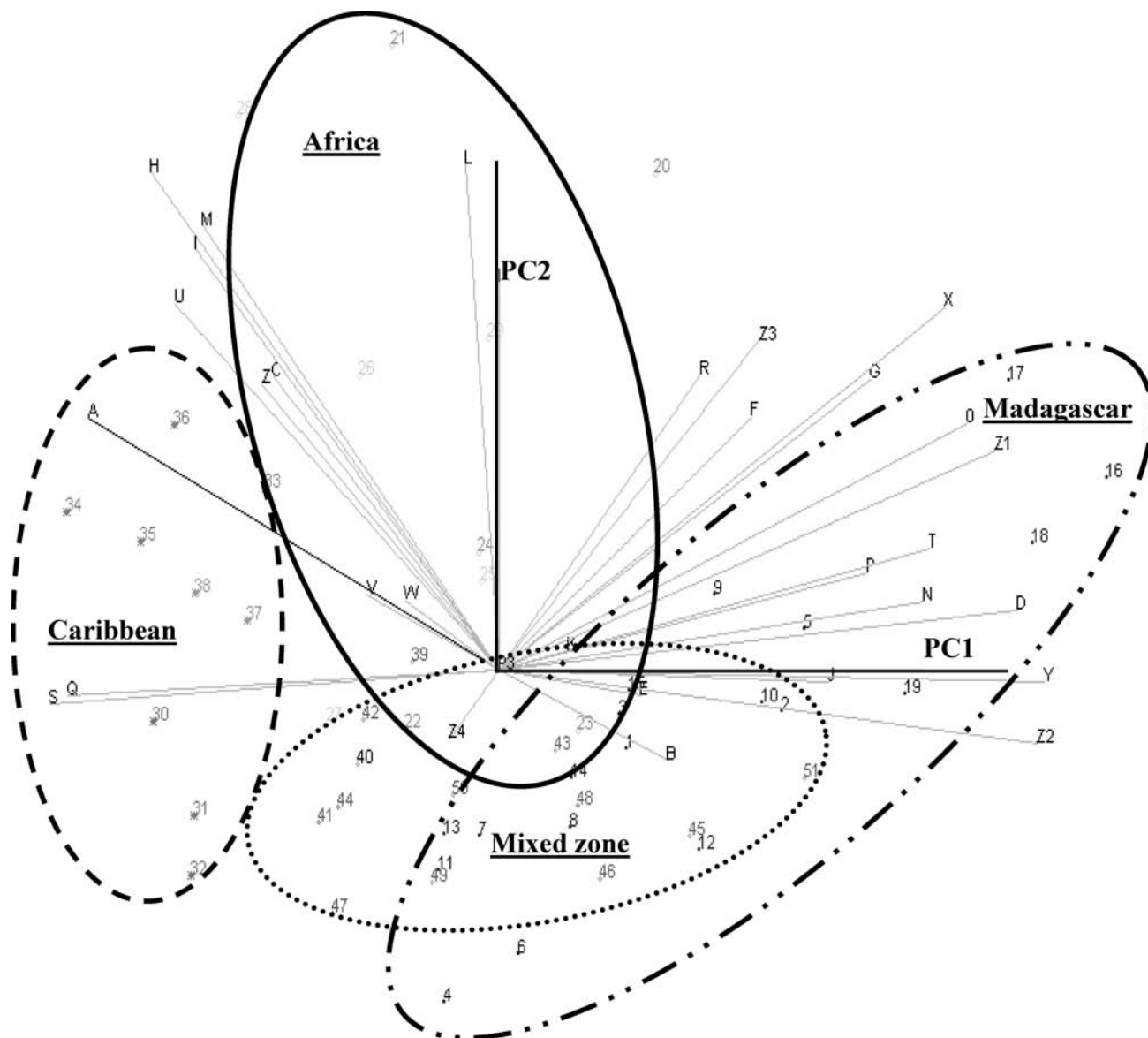
Figure 2. Principal component analysis, PC2 = *f*(PC1). Numbers represent chocolate samples: 1–19, Madagascar; 20–24, Ivory Coast; 25–29, Ghana; 30–32, Jamaica; 33–36, Trinidad; 37–38, Dominican Republic; 39–40, Ecuador; 41–51, Venezuela. Vectors A–Z4 represent the projections of each variable (peak).

set. It transforms the original set of variables to a new set of uncorrelated variables of greatest variance called principal components (PC). It allows relevant information to be extracted from confusing data sets.

In complement of PCA, a factorial discriminant analysis (FDA) was used. FDA is a multivariate method allowing the determination of which variables discriminate between two or more naturally occurring groups. This mathematical procedure maximizes the variance between groups and minimizes the variance within each group in such a way that outsiders can be detected more easily than by PCA.

RESULTS AND DISCUSSION

Extraction of Volatile Compounds. As shown in **Figure 1**, 30 significant peaks were selected in the chromatogram of the chocolate extracts of various origins. Peaks for which the area remains under the detection limit (determined as 3 times the background level) were not taken into account. The intensities of the peaks of interest vary according to geographical origin. The peaks located at the end of the chromatogram (from Q to Z2) show the main variation of intensity according to origin.

Principal Component Analysis. PCA was performed with a set of variables obtained from 51 samples of chocolate from

8 different countries. The goal was to correlate the molecules analyzed by GC-FID with the geographical origin of the samples. In the graphic representations of PCA results (**Figure 2**), dots (from 1 to 51) represent the projections of the individuals (51 samples of chocolate according to their geographical origin) in the chosen plane. Vectors, represented by segments of a straight line, correspond to the projections of variables (chromatographic peaks from A to Z4). The outcome of such an analysis is presented as a two-dimensional scatter plot (**Figure 2**) with the PC of higher representativity. In our case, the projection of results according to PC1 and PC2 is the most significant as it represents 40% of the variance. Circles gather all of the points corresponding to a given geographical origin. Collinear vectors highlight respective correlated variables, meaning their proportions in chocolates are linked.

The areas marked by circles contain only chocolates of pure origin (for example, the Caribbean area contains only Caribbean chocolates) except for the mixed zone.

PCA allows the determination graphically of the formation of groups according to geographical origins. Indeed, as shown in
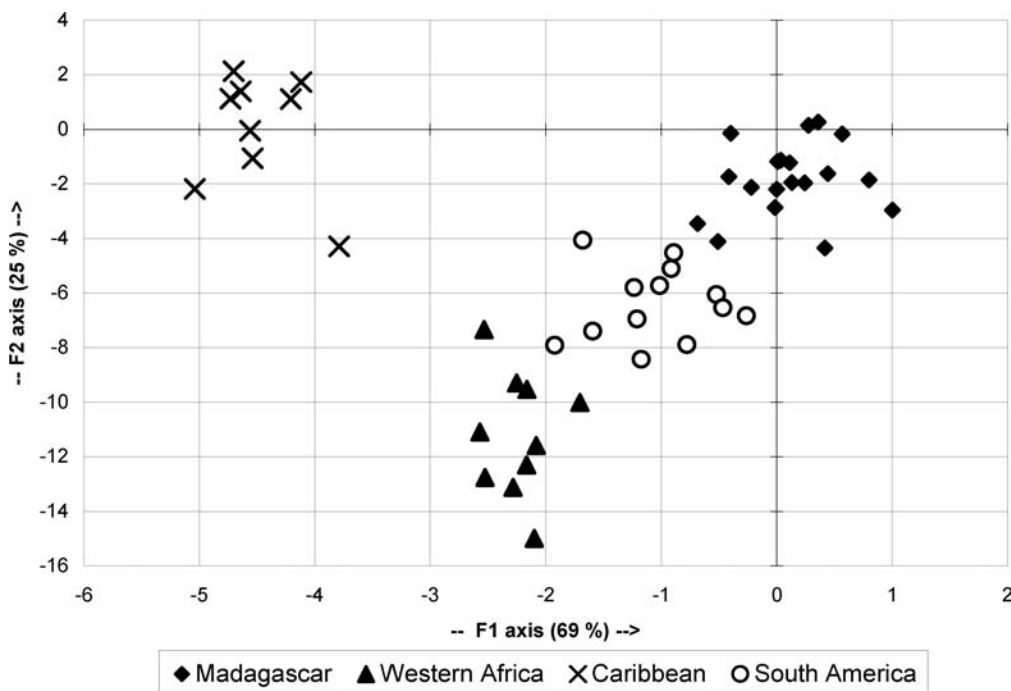
Article

J. Agric. Food Chem., Vol. 58, No. 3, 2010 **1481**

Figure 3. Sorting of chocolate samples according to their geographical origins within the frame of reference F1−F2 created by FDA.

Figure 2, Caribbean chocolates (Jamaica, Dominican Republic, and Trinidad) form a homogeneous group, distinctively separated from other origins. This group is characterized by variables (chromatographic peaks) such as A, Q, and S.

On the other hand, the majority of Madagascar chocolates and the majority of those from western Africa (Ivory Coast and Ghana) also form two distinct groups. Chocolates from Madagascar are more particularly characterized by variables such as X, O, Z1, D, Y, and Z2 as well as by high values according to PC1. In **Figure 2**, the Madagascar group is diametrically opposed to the Caribbean chocolate samples. Variables L, H, and M seem to be characteristic of western Africa. This group is mainly characterized by low values for the main component PC1 and positive high values for component PC2.

A pool of chocolates made from cocoa beans from various origins can be observed at the bottom center of the chart in the mixed zone. This mainly represents chocolates from South America (Venezuela, Ecuador) and from Madagascar, but also a chocolate from Ghana and two from the Ivory Coast. No characteristic variable clearly appears for this group.

In conclusion, by the reduction of the number of dimensions, PCA allows the extraction of three well-defined geographical areas from the analysis of volatile compounds of chocolate. Moreover, one or two characteristic compounds (e.g., H for Africa, Q for Caribbean) represented by chromatographic peaks can be identified and attributed to each geographical origin.

Factorial Discriminant Analysis. The algorithm of FDA shows chocolates in a two-dimensional landmark, as for PCA. The F1 and F2 axes have the same meaning as PC1 and PC2 in the case of PCA. In this case, only the first two areas were chosen because F1 represents 69% of the total variance, F2 25%, and F3 6%. F3's contribution was therefore directly considered to be insignificant, given that the landmark F2 = f(F1) already accounts for 94% of the variance. As FDA takes into account the composition of groups to determine the best factorial combination allowing them to be categorized, the represented percentages of variance are always much higher than for PCA.

Chocolates of different groups are therefore represented according to the F1 and F2 axes (**Figure 3**). The graph highlights the sorting into four groups of the 51 chocolates analyzed. Chocolates from Madagascar always formed a specific group (19 individuals). Ghana and Ivory Coast were grouped within the Africa group (10 individuals). Trinidad, Jamaica, and Dominican Republic were united in the Caribbean group (9 individuals), and, finally, Venezuela and Ecuador formed the South America group (13 individuals).

As within the PCA, the Caribbean group can easily be identified in **Figure 3**.

Figure 4 depicts the relationship between the variables analyzed (the volatile compounds of the chromatogram) and the geographical origin. It shows that the variables S, Q, A, and Z are characteristic of the Caribbean group. Variables H, M, U, and I allow the separation of the group Africa. For these two origins, the characteristic variables are similar to those determined by PCA.

South American chocolates appear to have intermediate characteristics between chocolates of West Africa and Madagascar and are thus located, in **Figure 4**, between those two.

This pattern differs slightly from PCA. Chocolates from the "Madagascar" group are the only ones influenced by the variables of the upper right quadrant (as, for example, B, J, N, D, T, Z2), but none of these variables is typically characteristic of this origin, as is clearly the case for the Caribbean and Africa groups. South America has intermediate characteristics between these two sources, L, C, E, K, and Z3, that is, the variables previously characteristic of Ecuador (E and K) and Venezuela. The results obtained by FDA are in agreement with those obtained by PCA.

To verify the results obtained previously, a blind analysis was carried out on 15 new chocolate samples from a different production year (referenced from 101 to 115). These samples were then analyzed under the same conditions as the previous 51 chocolates.

The new results were integrated into the previous data pool and added to the FDA as "unknown" samples. The algorithm, having already the benchmark established by the 51 previous chocolates,
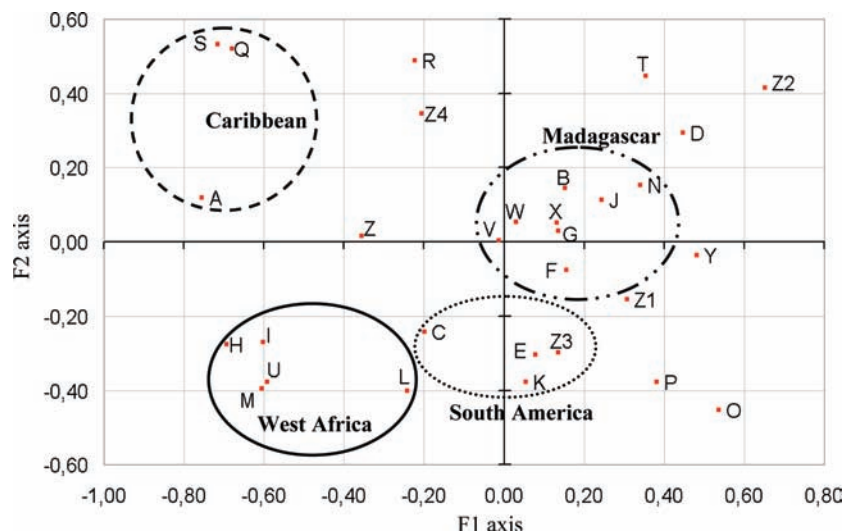
Figure 4. Coordinates of variables (volatile compounds) within the frame of reference F1−F2 created by FDA (classification by geographical zone).
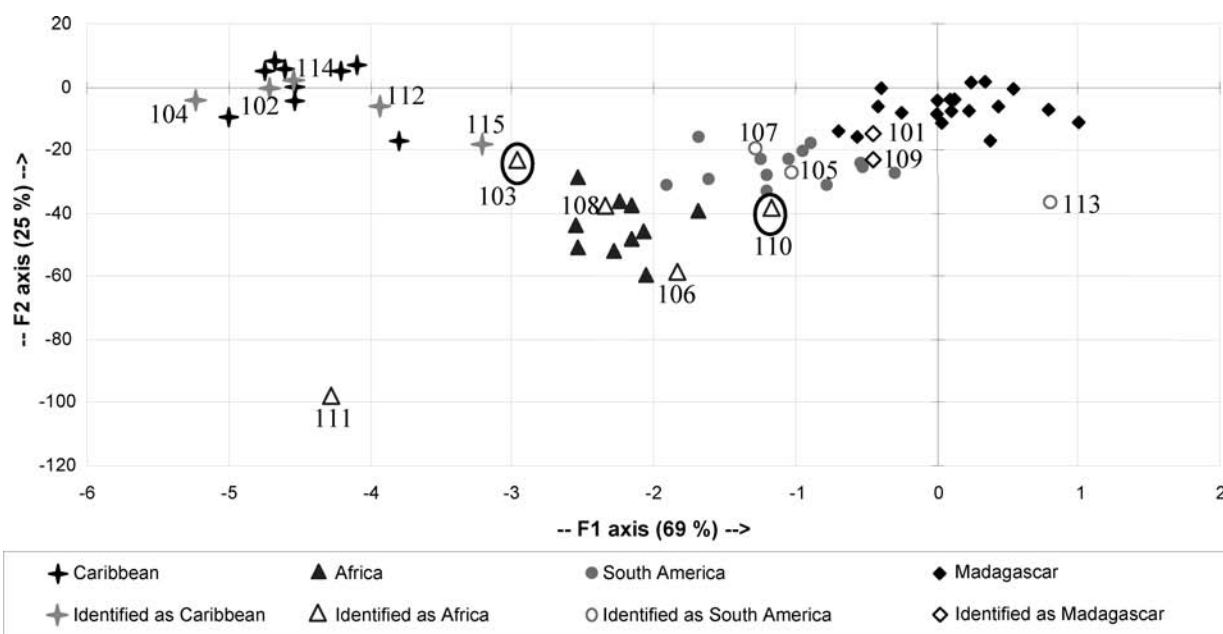

Figure 5. Classification of unknown chocolate samples after a blind analysis. The circled points are the samples that have not been correctly identified.

calculates the coordinates corresponding to these new samples in the plane defined by F1 and F2 and classifies them within the group to which they are closest. The results provided by the software are presented in **Figure 5**.

Among the 15 blind samples analyzed, 2 were classified in a group that did not correspond to their geographical origin. It is interesting to note that both incorrectly classified chocolates come from Venezuela, chocolate samples attributed to the mixed zone (**Figure 2**).

For the majority of the chocolate samples, the probability of belonging to a group is >90%, but there are chocolates for which the classification is less easy. Thus, two "unknown" chocolates of Madagascar, 101 and 109 (**Figure 5**), present coordinates that seem to be intermediate between the Madagascar group and the South America group. However, the probability of belonging to the original group is >75%, which remains a relevant value. This faltering reflects the difficulty of separating these two groups, quite close regarding the varieties, already visible in the results of PCA.

Therefore, the control of the reliability of this method by the addition of "unknown" additional samples showed that

FDA recognizes satisfactorily chocolates according to their geographical origin with a data set of volatile compounds analysis.

Identification of Volatile Compounds. To establish the volatile profile of the geographical origin of the cocoa used in the chocolate samples, identification of the molecule has been achieved. The identification has been made on the peaks that seem to be relevant candidates to label chemical tracers of continental origin of the cocoa beans used to produce chocolate.

Two or three characteristic compounds of each origin were identified by GC-MS analysis (**Table 1**). The identity of the impact compounds was confirmed by comparison of retention time and mass spectra of reference measurement standards.

Variables H, I, and L, characteristic of the Africa group, were identified as benzaldehyde, linalool, and phenylacetaldehyde, respectively. These three compounds have previously been identified in cocoa, like linalool, which was identified by Brainbridge and Davies (*18*). Phenylacetaldehyde formed by Strecker degradation of phenylalanine was earlier suggested as an important contributor to the cocoa aroma (*19, 20*). This compound exhibits

Table 1. Identification of Volatile Compounds According to Chocolates Geographical Origin

variable	identified as	group
H	benzaldehyde	Africa
I	linalool	Africa
L	phenylacetaldehyde	Africa
Q	(*E,E*)-2,4-nonadienal	Caribbean
S	(*E,E*)-2,4-decadienal	Caribbean
X	2-phenyl-2-butenal	Madagascar
Y	4-methyl-2-phenyl-2-pentenal	Madagascar

a honey note (*21*), whereas benzaldehyde added a hazelnut/almond note to the aroma profile of chocolate.

Furthermore, linalool, characteristic of the Africa group, presents the main sensorial attributes of roasted cocoa, which is responsible for the typical floral smell (*22*).

Two aldehydes were identified as characteristic compounds from the Madagascar group. Indeed, variables X and Y correspond to 2-phenyl-2-butenal and 4-methyl-2-phenyl-2-pentenal, respectively. 2-Phenyl-2-butenal is generated via aldol condensation of Strecker aldehydes and can work as an indicator of the roasting intensity of cocoa beans (*23*). Furthermore, it seems to be of great value for the richness of cocoa flavor and is used as a main component of artificial chocolate flavors (*24*). 4-Methyl-2-phenyl-2-pentenal, less reported in the literature, seems to give an intense bitter taste to cocoa (*2*).

(*E,E*)-2,4-Nonadienal (variable Q) and (*E,E*)-2,4-decadienal (variable S) were identified as characteristic for the Caribbean group. Although (*E,E*)-2,4-decadienal (fatty, waxy odor) has been reported as a primary odorant of milk products (*25*), it might be speculated that it is generated at the higher temperatures applied during conching of the chocolate (*26*).

The method described in this paper integrates the whole production process of chocolate. The analyses of chocolate samples, as a final product of this process, give results that take into account the origin of the cocoa beans, the fabrication process made by Valrhona, and the final product, the chocolate. The two statistical analyses allow the differentiation of chocolate samples according to the cocoa's geographical origin: Caribbean, Madagascar, Africa, and South America. Moreover, seven compounds have been identified and can be used as chemical tracers of the cocoa's continental origin.

LITERATURE CITED

(1) Flament, I. Coffee, cocoa, and tea. *Food Rev. Int.* **1989**, *5*, 317–414.

(2) Serra Bonvehi, J. S. Investigation of aromatic compounds in roasted cocoa powder. *Eur. Food Res. Technol.* **2005**, *221*, 19–29.

(3) Ramli, N.; Hassan, O.; Said, M.; Samsudin, W.; Idris, N. A. Influence of roasting conditions on volatile flavor of roasted Malaysian cocoa beans. *J. Food Process. Preserv.* **2006**, *30* (3), 280–298.

(4) Moulay, L.; Manzanares, P.; Valles, S.; Ramon, D.; Gil, J. V. Effect of enzyme treatments and drying temperatures on methylpyrazine content in cocoa (*Theobroma cacao* L.) powder extract. *J. Food Sci.* **2006**, *71*, S621–S625.

(5) Jonfia-Essien, W. A.; Alderson, P. G.; Tucker, G.; Linforth, R.; West, G. The growth of *Tribolium castaneum* (Herbst) and *Lasioderma serricorne* (Fabricius) on feed media dosed with flavour volatiles found in dry cocoa beans. *Pakistan J. Biol. Sci.* **2007**, *10*, 1301–1304.

(6) Luna, F.; Crouzillat, D.; Cirou, L.; Bucheli, P. Chemical composition and flavor of Ecuadorian cocoa liquor. *J. Agric. Food Chem.* **2002**, *50*, 3527–3532.

(7) Schwan, R.; Wheals, A. The microbiology of cocoa fermentation and its role in chocolate quality. *Crit. Rev. Food Sci. Nutr.* **2004**, *44* (4), 205–221.

(8) Cros, E.; Jeanjean, N. Formation de l'arôme cacao. In *Cacao et Chocolat: Production, Utilisation, Caractéristiques*; Tec & Doc − Lavoisier: Paris, France, 1989; pp 187−206.

(9) Kattenberg, H. R.; Kemmink, A. The flavor of cocoa in relation to the origin and processing of the cocoa beans. *Dev. Food Sci. Food Flavors, Ingred. Compos.* **1993**, *32*, 1–22.

(10) Cidell, J. L.; Alberts, H. C. Constructing quality: the multinational histories of chocolate. *Geoforum* **2006**, *37*, 999–1007.

(11) Hernandez, C. V.; Rutledge, D. N. Multivariate statistical analysis of gas chromatograms to differentiate cocoa masses by geographical origin and roasting conditions. *Analyst (Cambridge, U.K.)* **1994**, *119*, 1171–1176.

(12) Müggler-Chavan, F.; Reymond, D. Aromatic components of cocoa from different origins. *Mitt. Gebiete Lebensmittelunters. Hyg.* **1967**, *58*, 466–473.

(13) Jinap, S.; Dimick, P. S.; Hollender, R. Flavour evaluation of chocolate formulated from cocoa beans from different countries. *Food Control* **1995**, *6*, 105–110.

(14) Counet, C.; Ouwerx, C.; Rosoux, D.; Collin, S. Relationship between procyanidin and flavor contents of cocoa liquors from different origins. *J. Agric. Food Chem.* **2004**, *52*, 6243–6249.

(15) Counet, C.; Callemien, D.; Ouwerx, C.; Collin, S. Use of gas chromatography−olfactometry to identify key odorant compounds in dark chocolate. Comparison of samples before and after conching. *J. Agric. Food Chem.* **2002**, *50*, 2385–2391.

(16) Cambrai, A. *Authenticity of chocolate: design of analytical method for the determination of geographical origin of chocolate*. Thesis, University Louis Pasteur of Strasbourg, **2007**; 195 pp.

(17) Sall, J.; Creighton, L.; Lehman, A. *JMP ® Start Statistics, A Guide to Statistics and Data Analysis Using JMP® and JMP IN® Software*, 3rd ed.; SAS Institute Inc./Thomson Brooks/Cole: Belmont, Canada, 2005; 560 pp.

(18) Bainbridge, J. S.; Davies, S. H. The essential oil of cocoa. *J. Chem. Soc.* **1912**, *101*, 2209–2221.

(19) van Praag, M.; Stein, H. S.; Tibbetts, M. S. Steam volatile aroma constituents of roasted cocoa beans. *J. Agric. Food Chem.* **1968**, *16*, 1005–1008.

(20) Frauendorfer, F.; Schieberle, P. Identification of the key aroma compounds in cocoa powder based on molecular sensory correlations. *J. Agric. Food Chem.* **2006**, *54*, 5521–5529.

(21) Holscher, W.; Vitzthum, O. G.; Steinhart, H. Identification and sensorial evaluation of aroma-impact-compounds in roasted Columbian coffee. *Cafe, Cacao, The* **1990**, *34*, 205–212.

(22) Ziegleder, G. Linalool contents as characteristic of some flavour grade cocoas. *Lebensm. Unters Forsch.* **1990**, *191*, 306–309.

(23) Fadel, H. H. M.; Abdel Mageed, M. A.; Abdel Samad, A. K. M. E.; Lotfy, S. N. Cocoa substitute: Evaluation of sensory qualities and flavour stability. *Eur. Food Res. Technol.* **2006**, *223*, 125–131.

(24) Ziegleder, G. In *Modern Methods of Plant Analysis*; Linskens, H., Jackson, J. F., Eds.; Springer: Berlin, Germany, 1988; pp 289−299.

(25) Schieberle, P.; Gassenmeier, K.; Sen, A.; Guth, H.; Grosch, W. Character impact odour compounds of different kinds of butter. *Lebensm. Wiss. Technol.* **1993**, *26*, 347–356.

(26) Schnermann, P.; Schieberle, P. Evaluation of key odorants in milk chocolate and cocoa mass by aroma extract dilution analyses. *J. Agric. Food Chem.* **1997**, *45*, 867–872.