Differentiation of coffee varieties according to their sterolic profile

HELENA ČÍŽKOVÁ - VERONIKA SOUKUPOVÁ - MICHAL VOLDŘICH - RUDOLF ŠEVČÍK

Summary

The aim of the presented work was to differentiate between two important coffee varieties – *arabica* and *robusta* by means of chemical analyses of sterolic fraction of 13 model and 14 commercial roasted coffee samples. After the extraction of the coffee oil, the lipids have been saponified and sterols converted into trimethyl silyl derivatives and analysed by gas chromatography with mass spectrometric detection. For the evaluation of authenticity of coffee samples several statistical approaches, such as multilinear regression, principal component analyses and cluster analysis were used. Δ^5-avenasterol was found to be the chemical descriptor with the highest contribution to the model and the following correlation equation was achieved: percentage of *arabica* in coffee = 117 – 11.1 × percentage of Δ^5-avenasterol in total sterols, with a high correlation coefficient (R^2 = 0.98) and precision (*RSD* = 7.9 %). The method was applied to the determination of the *arabica*/*robusta* composition of commercial roasted coffee samples.

Keywords

coffee; *arabica*; *robusta*; authenticity; sterols, gas chromatography, multilinear regression, principal component analysis, cluster analysis

There are two main species of the coffee plant, *Coffea arabica* being the older one. While more susceptible to disease, its flavour is considered to be better than that of the other variety, *Coffea canephora (robusta)*. *Robusta*, which contains about 40–50% more caffeine, can be cultivated in environments where *arabica* will not thrive. This has led to its use as an inexpensive substitute for *arabica* in many commercial coffee blends. Compared to *arabica*, the *robusta* variety tends to be bitter and has little flavour, with a telltale "burnt rubber" or "wet cardboard" aroma and flavour. Good quality *robusta species* are used in some espresso blends to provide a better "cream" and to lower the ingredient cost. The large industrial roasters use a steam treatment process to remove undesirable flavours from *robusta* beans for use in the mass-marketed coffee blends [1].

Commonly, green coffee beans of the *arabica* and *robusta* varieties can be distinguished by their size and shape but the roasting process and milling eliminate these macroscopic criteria. This raises the possibility of fraudulence or mislabelling and chemical analysis is required to differentiate between these two varieties. Legislation of the Czech Republic does not regulate content and composition of coffee varieties on the label; however, in case the producer indicates the content and composition, this fact has to be fulfilled (Law No 110/1997 Food law – Law No 274/2003 as amended; Law No 634/1992 Consumer protection – law No 227/2003 as amended).

The approaches to the evaluation of the variety composition of coffee blends are summarized in Table 1. The majority of the studies are based on the chromatographic determination of the unsaponifiable lipid fraction.

The lipid contents and lipid composition of roasted coffee beans vary with the variety, maturation and drying procedure. *C. arabica* contains more lipids (15.5%) than *C. robusta* (9.8%), but the lipid profiles are similar. Triglycerides and diterpene alcohol esters are the major lipid classes in coffee brewed from the ground coffee beans, and ranged from 86.6 to 92.9% and 6.5 to 12.5% of total lipids, respectively. The main difference between the varieties can be found in sterolic fraction; both *arabica* and *robusta* contains 0.1% of sterols esters and 1.5 and 1.3% of sterols, respectively, however the sterol composition is different

Helena Čížková, Michal Voldřich, Rudolf Ševčík, Department of Food Preservation and Meat Technology, Institute of Chemical Technology, Technická 5, CZ-160 00 Prague 6, Czech Republic.
Veronika Soukupová, Czech Agriculture and Food Inspection Authority, Za Opravnou 300/6, CZ-150 06 Prague 5, Czech Republic.

Correspondence author:
Helena Čížková, e-mail: helena.cizkova@vscht.cz

Tab. 1. The approaches to the evaluation of the variety composition of coffee blends.

Analytical determination	Significant parameters for differentiation	Prediction error (expressed as in reference)	Chemometrics	References
Sterolic profile	Δ^5-avenasterol	prediction error: 1.1%	MLR, PCR, PCA	Valdenebro et al. [2]
Unsaponifiable lipid fraction	diterpenic alcohols profile	limit of detection (*robusta* in *arabica*) 5–10%	PCA	Frega et al. [3]
Diterpenic alcohols and sterols	kahweol, cafestol, 16-O-methyl-cafestol, Δ^5-avenasterol		none	Lercker et al. [4]
Sterolic profile	Δ^5-avenasterol, sitostanol	Cooman's weights: avenasterol - 6.62; sitostanol - 2.97	PCA	Carrera et al. [5]
Tocopherols and triglycerides	β-tocopherol and γ-tocopherol		PCA and LDA	Gonzales et al. [6]
Fatty acids	fatty acids profile	*Arabica*/*robusta* recognition ability: 100%	PCA and LDA	Martin et al. [7]
Aroma	volatile compounds profile		CA and PCA	Freitas et al. [8]
Metal content	P, Mn and Cu	prediction error: 7%	PCA	Martin et al. [9]
Trace heavy metals	Mn and Zn		none	Merletta et al. [10]

MLR - multiple linear regression, PCR - principal component regression, PCA - principal component analyses, LDA - linear discriminant analysis, CA - cluster analyses.

[11]. The stability of the sterols within the production of the ground brewed coffee was proved by Jham [12], who investigated the effects of coffee type and drying procedures on the lipid classes and triacylglycerols in the coffee samples from Brazil. For six different treatments, the significant result was found for the composition of triacylglycerols and diacylglycerols, but not for sterols, terpene esters, monoacylglycerols and free fatty acids. Due to the facts mentioned above, the study of sterolic fraction of the lipids present in coffee oil seems to be the most promising approach.

The aim of the presented work was to differentiate between two important coffee varieties - *arabica* and *robusta* by means of chemical analyses of sterolic fraction of 27 roasted coffee samples. After the extraction of the coffee oil, the lipids have been saponified and sterols converted into trimethylsilyl derivatives and analysed by gas chromatography with mass spectrometric detection (GC/MS).

MATERIALS AND METHODS

Samples

The set of 13 coffee samples of known geographic origin was selected for the analyses. Six samples belonged to the *arabica* variety, four were of the *robusta* variety, and three samples were different *arabica-robusta* mixtures (25% *arabica* + 75% *robusta*, 50% *arabica* + 50% *robusta*, 75% *arabica* + 25% *robusta*) prepared in the laboratory.

Fourteen samples of commercial roasted coffee were purchased from the market.

All coffee samples were ground and powdered and than stored in polyethylene flasks.

Analytical procedure

About 7 g of the coffee sample was extracted with hexane in a Soxhlet apparatus for 4 h. The extract was dried over anhydrous sodium sulfate, evaporated using a vacuum rotary evaporator and dried at 105 °C.

Saponification of lipids and derivatization were performed according to the modified AOAC method. The obtained oil was treated with 40 ml ethanol and 8 ml 50% potassium hydroxide, refluxed for 70 min, then 60 ml of ethanol was added and the mixture refluxed for additional 15 min. Sterolic fraction was extracted with benzene, washed three times with water, dried over anhydrous sodium sulfate and evaporated to dryness. Next trimethylsilyl derivatives were prepared: the residue was dissolved in 0.8 ml of pyridine, transferred into cuvette, where 0.2 ml trimethylchlosilan (TMCS) was added, the cuvette closed, shaked vigorously for 30 s, let stay for 15 min and centrifugated. A 1 μl aliquot of this solution was injected into the gas chromatograph; the content of individual sterols was expressed as the percentage of the sterolic fraction obtained.

GC conditions: GC/MS (Hewlett Packard 5890 Series II GC with MS, Agilent Technologies, Santa Clara, USA), operating conditions: column CP Sil 8 CB/MS (Varian, Palo Alto, USA)

30 m × 0.25 mm × 0.25 μm, carrier gas: helium, constant flow rate 0.5 ml.min^{-1}, temperature program: isothermal 265 °C, injector: 280 °C, split 1:1, detector: 300 °C.

RESULTS AND DISCUSSION

The majority of the recent works that dealt with the evaluation of the *arabica*/*robusta* composition of roasted coffee are based on the determination of the composition of unsaponifiable lipid components after the saponification, their isolation and determination by GC.

Fig. 1 shows the characteristic chromatogram of the sterolic fraction of coffee oil. Six main sterols were identified in the coffee samples (Tab. 2). The obtained results confirmed a relatively stable composition of coffee oil from *arabica* variety that consists in average of 14.1%, 21.2%, 52.2%, 1.7% and 9.5% of campesterol, stigmasterol, sitosterol, Δ^5-avenasterol and stigmastenol, respectively, and the coffee oil from *robusta* variety that consists in average of 14.8%, 19.3%, 42.2%, 9.5% and 8.9% of campesterol, stigmasterol, sitosterol, Δ^5-avenasterol and stigmastenol, respectively. The results correspond with the previous findings for coffee sterols as summarised by CARRERA et al. [5], with sitosterol, campesterol and stigmasterol being the main components. As can be seen, the difference in the sterolic profile in authentic samples manifests itself mainly in the case of Δ^5-avenasterol, which is in average five times more abundant in the *robusta* variety.

Differences between *arabica*/*robusta* varieties

Several statistical approaches (multilinear regression, principal component analyses and cluster analyses) were used for the evaluation of authenticity of coffee samples.

The principal component analysis was applied to the data matrix. Fig. 2 shows the plot of the loadings obtained for the two first principal components (PC) of the roasted *arabica* and *robusta* authentic coffees used. Two principal components account for the 74% of the total variance. The variables Δ^5-avenasterol, campesterol and sitosterol are the descriptors with the highest contribution to the PC1. The scores of the authentic samples are in Fig. 3. The grouping is not very clear, but it can be observed that all *arabica* samples are on the left (negative) side of the PC1, while the *robusta* samples appear on the right (positive) side. This result is in a good agreement with the fact that the PC1 differentiates *arabica* and *robusta* samples according to the concentration of sitosterol (high for *arabica*, low for *robusta*), 5-avenasterol and campesterol (low for a*rabica*, high for *robusta*).

Similar results, only in the reverse order, were obtained from the analyses of the entire data set, i.e. of the commercial roasted coffee samples, model coffee blends and authentic coffee samples. Two principal components account for the 65% of the total variance. Variables Δ^5-avenasterol and sitosterol are the descriptors with the highest contribution to the PC1. The scores of the authentic samples are depicted in Fig. 4. When considering the sterolic profile, differences between coffees become apparent. It can be seen that all *robusta*

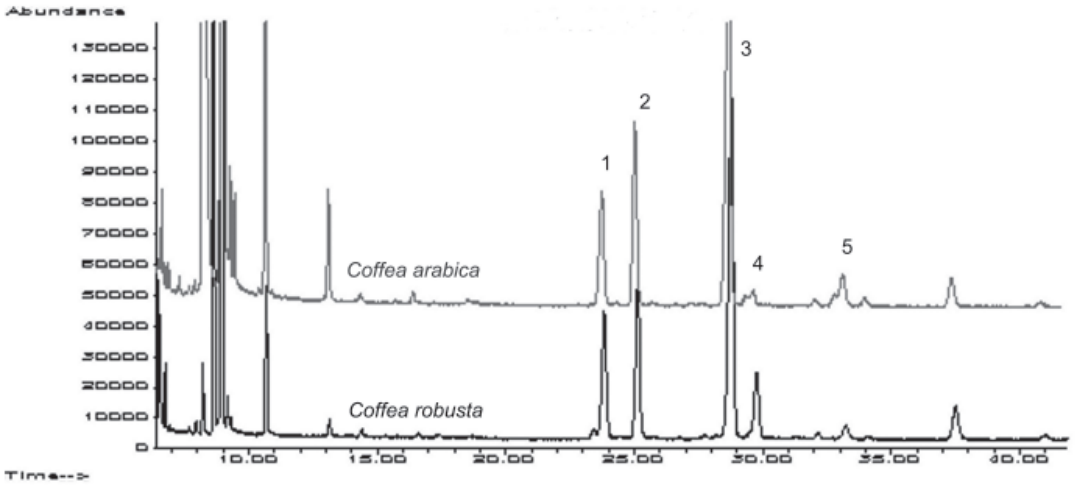

Fig. 1. Overlaid chromatograms of the sterolic fraction of the roasted coffee samples: a) 100% *arabica*, b) 100% *robusta*.
Peak assignment: 1 - campesterol, 2 - stigmasterol, 3 - beta-sitosterol, 4 - Δ^5-avenasterol, 5 – stigmastenol.

Tab. 2. Composition of sterolic fraction [%] of commercial roasted coffee samples (1X-14X), model coffee blends (22A25, 23A50, 24A75) and authentic coffee samples (A, R).

Sample	campesterol	stigmasterol	sitosterol	Δ^5-avenasterol	stigmastenol
1X	14.4	20.5	43.3	8.7	10.5
2X	14.9	20.3	44.0	8.3	9.9
3X	13.9	20.3	43.1	10.0	10.1
4X	12.9	17.4	41.4	8.7	6.9
5X	17.0	21.0	50.0	2.0	10.0
6X	14.5	21.6	42.1	12.2	9.5
7X	12.7	21.4	52.9	3.7	9.2
8X	14.7	22.3	48.2	7.1	7.8
9X	12.7	20.5	47.3	10.0	9.6
10X	14.6	21.2	53.3	1.4	9.7
11X	14.3	22.2	52.3	2.0	9.4
12X	14.2	21.2	50.5	4.9	9.3
13X	14.0	21.5	50.2	4.1	10.5
14X	14.0	19.8	54.1	3.1	9.2
15A	14.1	20.2	49.6	2.2	10.9
16A	13.5	21.0	49.7	2.5	10.3
17A	14.0	22.4	49.5	1.5	9.6
18R	14.1	20.6	35.0	9.6	7.7
19R	13.6	19.4	37.8	10.1	9.4
20R	15.5	18.5	48.3	9.0	9.0
21R	16.0	18.5	47.7	9.4	9.5
22A25	14.6	19.4	48.3	8.3	9.6
23A50	14.6	22.2	50.2	5.6	7.3
24A75	14.9	20.9	51.0	4.7	9.2
25A	14.8	22.0	53.8	1.3	8.7
26A	13.9	21.4	54.4	1.2	9.0
27A	14.5	20.4	55.9	1.2	8.2
Average composition of authentic samples					
arabica	14.1	21.2	52.2	1.7	9.5
robusta	14.8	19.3	42.2	9.5	8.9
Standard deviation					
arabica	0.5	0.9	2.9	0.6	1.0
robusta	1.1	1.0	6.8	0.5	0.8

samples are situated at left side of the plot being well separated from the *arabica* coffees, which are on the right (PC1 > 1), although the mixtures of both model and real *arabica*/*robusta* blends are not clearly differentiated by their *arabica* content.

The cluster analysis of all samples (commercial roasted coffee samples, model coffee blends and authentic coffee samples) based on the concentrations of the measured sterolic profile was performed (Fig. 5). The samples could be simply divided into three groups according to their respective *arabica* content:

– samples with the high content of *arabica*,
– mixtures of *arabica* and *robusta* varieties, and
– samples with the majority of *robusta*.

However, the discriminating ability inside the main clusters is insufficient and several outliers (6X, 8X) were be found.

The results obtained from the analyses of au-

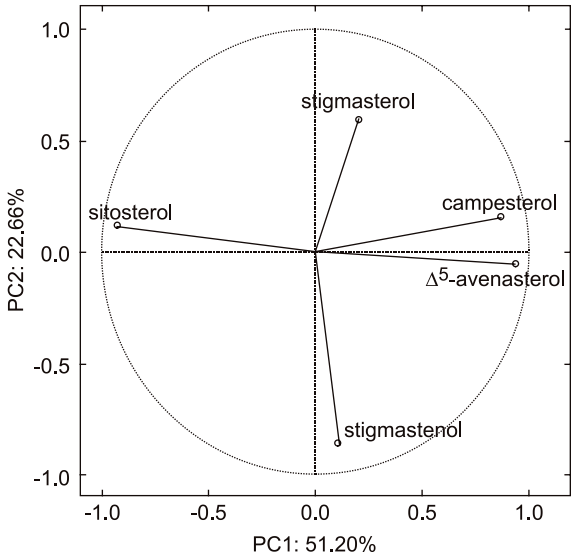

Fig. 2. Plot for the PCA loadings for the set of authentic samples.

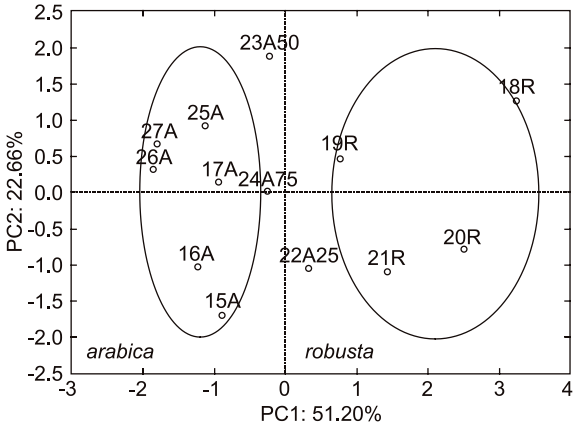

Fig. 3. Plot for the PCA scores for the set of authentic samples.

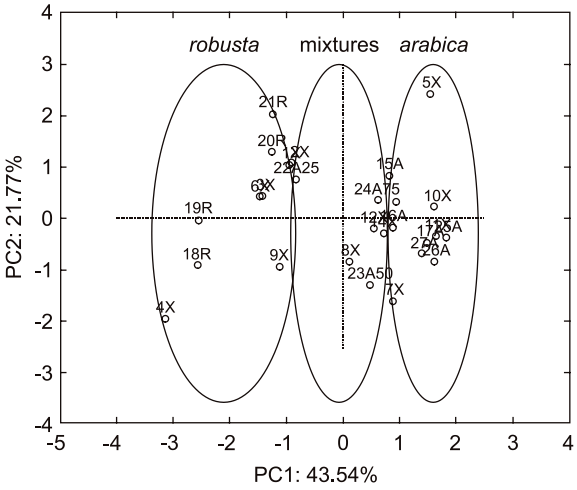

Fig. 4. Plot for the PCA loadings for the set of all samples.

thentic samples were used to set up a multiple regression equation. All variables were included into the model in order to estimate the *arabica* content in real samples of coffee. The regression equation can thus be written as: *arabica* content [%] = = 21 – 3.4 × campesterol [%] + 3.3 × stigmasterol [%] + 0.5 × sitosterol [%] – 10.1 × Δ^5-avenasterol [%] + 5.1 × stigmastenol [%].

This equation correlates well with the real *arabica* variety content with an acceptable observed accuracy ($R^2 = 0.90$, $RSD = 5.8$).

However, in spite of the working prediction model, our principal aim was to find the relationship between *y* variable (*% arabica* in the coffee mixture) and the chemical descriptor with the highest contribution to the model. According to further correlation study the best fit was achieved with Δ^5-avenasterol:

$$arabica\ [\%] = 117 - 11.1 \times \Delta^5\text{-avenasterol}\ [\%]$$

with a high correlation coefficient ($R^2 = 0.98$) and precision ($RSD = 7.9$) obtained. However, for the limited number of the model samples ($N = 13$). The correlation between the Δ^5-avenasterol content and the actual value of *% arabica* in the *arabica-robusta* mixtures is given in Fig. 6. The results showing that Δ^5-avenasterol alone as a very good predictor are in good accordance with the findings of VALDENEBRO et al. [2], who calculated the following predictive model for the Kraft Jacobs Suchard coffees:

$$arabica\ [\%] = 129 - 10.5 \times \Delta^5\text{-avenasterol}\ [\%]$$

To validate the results obtained for the model samples by the statistical analyses (multilinear regression, principal component analyses a cluster analyses) a test set containing the real coffee samples purchased at the market was considered. Achieved results of all statistical approaches were in a good agreement (Tab. 3); the coffees claimed to be of the *arabica* variety were proved to be authentic. Labelling and composition of all commercial roasted coffee samples fulfilled the requirements of the recent Czech legislation. The analyses of sterolic profile of the oil of roasted coffee provide a very useful tool for the evaluation of coffee variety using Δ^5-avenasterol as the main marker of *arabica*/*robusta* ratio in the sample. Prediction error of classifying the coffee samples according to the suggested correlation of *arabica* content with the percentage of Δ^5-avenasterol is about 8%. However, the analysed set of the model samples was rather limited and its extension would allow for a more reliable and better discrimination.

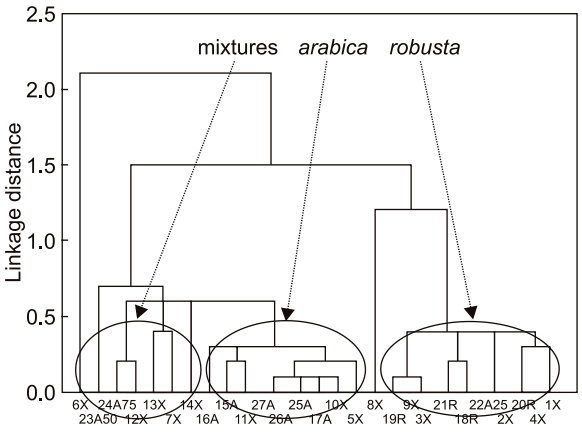

Fig. 5. Cluster analysis of the coffee samples.

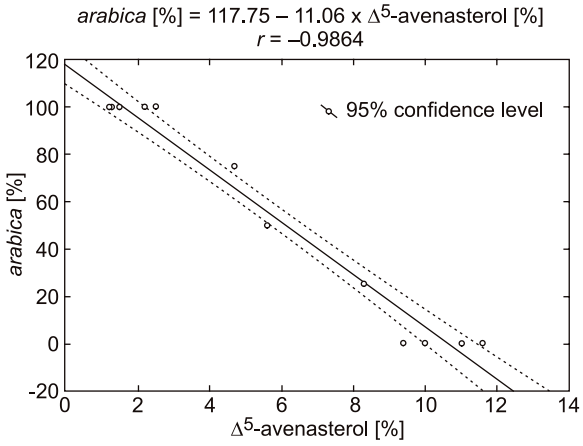

Fig. 6. Correlation between the Δ5-avenasterol content and the actual value of % *arabica* in *arabica*/*robusta* mixtures.

Tab. 3. Variety types of commercial coffees obtained from chemical analyses of sterolic fraction according to different statistical approaches.

Sample	Claimed	PCA	CA	Multilinear regression *arabica* [%] = 117 – 11.1 × Δ5-avenasterol [%]
1X	mixture	*robusta*	*robusta*	21
2X	mixture	*robusta*	mixture	26
3X	*robusta*	*robusta*	*robusta*	7
4X	mixture	*robusta*	*robusta*	21
5X	*arabica*	*arabica*	*arabica*	99
6X	unknown	*robusta*	outlier	–18
7X	unknown	*arabica*	mixture	77
8X	mixture	mixture	outlier	39
9X	unknown	*robusta*	*robusta*	7
10X	*arabica*	*arabica*	*arabica*	102
11X	*arabica*	*arabica*	*arabica*	96
12X	mixture	mixture	mixture	63
13X	mixture	*arabica*	mixture	72
14X	mixture	mixture	mixture	83

PCA - principal component analyses, CA - cluster analyses.

Acknowledgements

The author is grateful for financial support provided by the Ministry of Education, Youth and Sports of the Czech Republic (research project MSM 6046137305).

REFERENCES

1. Singhal, R. S. - Kulkarni, P. R. - Rege, D. V.: Handbook of indices of food quality and authenticity. Cambridge : Woodhead Publishing, 1997. 562 pp. ISBN 1-85573-299-8.
2. Valdenebro, M. S. - Leon-Camacho, M. - Pablos, F. - Gozales, A. G. - Martin, M. J.: Determination of the *arabica*/*robusta* composition of roasted coffee according to their sterolic content. Analyst, *124*, 1999, pp. 999-1002.
3. Frega, N. - Bocci, F. - Lercker, G.: High resolution gas chromatography method for the determination of *Robusta* coffee in commercial blends. Journal of High Resolution Chromatography, *17*, 1994, pp. 303-307.
4. Lercker, G. - Frega, N. - Bocci, F. - Rodriguez-Estrada, M. T.: High resolution gas chromatographic determination of diterpenic alcohols and sterols in coffee lipids. Chromatographia, *41*, 1995, pp. 29-33.
5. Carrera, F. - Leon-Camacho, M. - Pablos, F. -

Gonzales, A. G.: Authentication of green coffee varieties according to their sterolic profile. Analytica Chimica Acta, *370*, 1998, pp. 131-139.
6. Gonzales, A. G. - Pablos, F. - Martin, M. J. - Leon-Camacho, M. - Valdenebro, M. S.: HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. Food Chemistry, *73*, 2001, pp. 93-101.
7. Martin, M. J. - Pablos, F. - Gonzales, A. G. - Valdenbro, M. S. - Leon-Camacho, M.: Fatty acid profile as discriminant parameters for coffee varieties differentiation. Talanta, *54*, 1999, pp. 291-297.
8. Freitas, C. A. M. - Mosca, A. I.: Coffee geographic origin - an aid to coffee differentiation. Food Research International, *32*, 1999, pp. 565-573.
9. Martin, M. J. - Pablos, F. - Gonzales, A. G.: Characterization of *arabica* and *robusta* roasted coffee varieties and mixture resolution according to their metal content. Food Chemistry, *66*, 1999, pp. 365-370.
10. Pertoldi Marletta, G. - Procida, G. - DePalo, D. - Weber, A.: Characterisation of *Arabica* and *Robusta* coffee varieties according to their trace heavy metal content. In: Proceedings of 19th International Scientific Colloquium on Coffee. Trieste : Association for Science and Information on Coffee, 2001, pp. 386-390.
11. Ratnayake, W. M. N. - Hollywood, R. - O'Grady, E. - Stavric, B.: Lipid content and composition of coffee brews prepared by different methods. Food and Chemical Toxicology, *31*, 1993, pp. 263-269.
12. Jham, G. N. - Vilikova, R. - Miller, H. V. - Nikolova-Damymanova, B. - Devon, P. R: Lipid classes and triacylglycerols in coffee samples from Brasil: effect of coffee type and drying procedures. Food Research International, *34*, 2001, pp. 111-115.

Received 6 November 2006; revised 19 January 2007; accepted 24 January 2007.