Genotype characterization of cocoa into genetic groups through caffeine and theobromine content predicted by near infra red spectroscopy

Davrieux, F. [a]; Assemat, S. [a]; Sukha, D. [b]; Portillo, E. [c]; Boulanger, R. [a]; Bastianelli, D. [d] and Cros E. [a]

[a] *CIRAD, TA 80/16, 34398 Montpellier Cedex 5, France. E-mail: davrieux@cirad.fr*

[b] *Cocoa Research Unit, University of the West Indies, St. Augustine, Trinidad and Tobago*

[c] *Facultad de Agronomica, Universidad del Zulia, Maracaibo, Estado Zulia, Venezuela*

[d] *CIRAD, Laboratoire d'Alimentatation Animale,TA 30/A, 34398 Montpellier Cedex 5, France*

Keywords: cocoa, caffeine, theobromine, purine, linear discriminante analysis, near infrared spectroscopy, genotype, characterization

Introduction

There are three main types of cultivated cocoa trees [1], Forastero, Criollo, and Trinitario. Trinitarios are a natural hybrid between Fostasteros and Criollos. Annual world production of cocoa amounts to 3 million tonnes, comprising approximately 85% Forastero (mainly West Africa), 15% Trinitario (mainly Caribbean region) and less than 1% Criollo (Venezuela, Mexico, and Madagascar). In terms of aroma and sensory quality, some Trinitario cocoa is considered to be "fine cocoa", unlike Forastero cocoa, which is considered as bulk cocoa. "Fine" grade cocoa sells for a higher price than bulk cocoa [2]. Being able to authenticate "fine" cocoas is therefore an asset for chocolate makers. In the absence of any visual traits to recognize them, analytical tools are needed in addition to an administrative traceability approach.

Among cocoa constituents, caffeine and theobromine are involved in cocoa flavour development [3]. Furthermore, the relative level of these two compounds is supposed to be linked to genetic and/or geographical origin [4]. This study was undertaken to test the potential of near infrared spectroscopy for:

- quantifying purines in cocoa,
- distinguishing between genotypes according to their predicted purine contents and
- distinguishing between genotypes according to their spectral fingerprint.

Experimental procedure

The study was carried out on 322 cocoa samples over six production years (1999 to 2004). The number of samples of each genotype and their geographical origin is given in Table 1. There were no samples of Trinitario and Criollo cocoa from the Ivory Coast and no samples of Criollo cocoa from Trinidad.

Table 1. Number of cocoa samples of each genotype and their geographical origin.

Geographical origin	Genotype			Total
	Forastero	Trinitario	Criollo	
Ivory Coast	77			77
Venezuela	4	87	57	148
Trinidad	36	61		97
Total	117	148	57	322

Methods

Wet chemical methods

After reflux extraction in water, caffeine and theobromine contents were determined by HPLC (high-performance liquid chromatography) using an Agilent system series 1100 with a UV-Vis diode array detector. The detection and quantification were performed at the maximum absorption wavelength (280 nm).

Near infrared spectroscopy

Near infrared (NIR) spectroscopy acquisitions were obtained on a Foss-Perstorp 6500 analyser using a spin cell. Spectral data were collected and processed using WinISI 1.5 software (InfraSoft International). 3 g of cocoa taken from 100 g of shelled, ground and sieved beans (0.5 mm) was analysed by diffuse reflectance from 400 nm to 2,500 nm in 2 nm steps.

Results and discussion

Chemical discrimination

Partial least square models were used to establish quantitative relations between NIR spectral bands (900 nm to 2,500 nm) and both caffeine and theobromine content [5]. The models developed (Table 2) fitted the data well; the coefficients of determination (R^2) were equal to 0.94 for caffeine and 0.88 for theobromine.

Table 2. Equation statistics for prediction of caffeine and theobromine content.

Constituent	n	Mean	SD	SEC	R^2	SECV	RPD.
Caffeine	309	0.28	0.14	0.034	0.94	0.043	3.2
Theobromine	313	0.97	0.20	0.070	0.88	0.084	2.4

n: number of samples chosen by the model (Student *t* test at $P < 0.05$)
SD: standard deviation of the calibration population
SEC: standard error of calibration
R^2: coefficient of determination
SECV: standard error of cross-validation
RPD: ratio of performance to deviation ($SD.SECV^{-1}$)

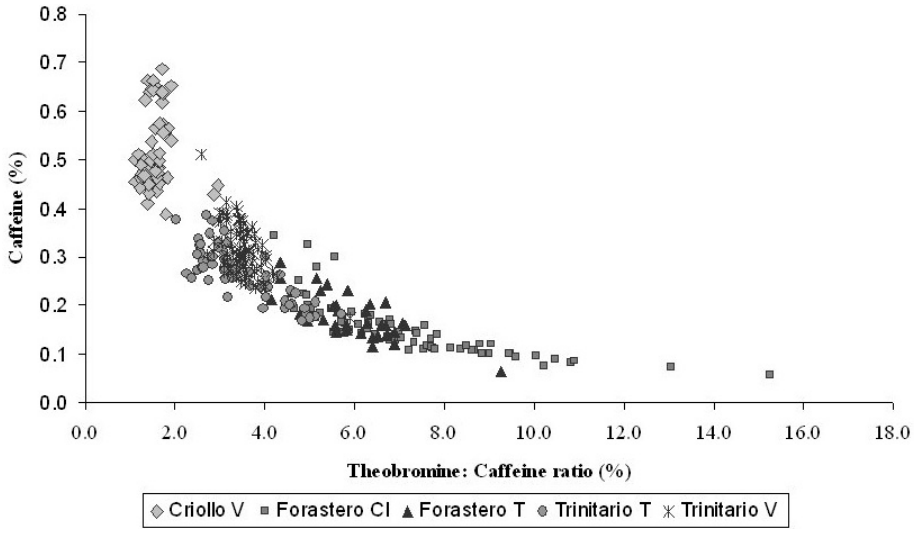

Figure 1. Representation of the calibration samples according to their predicted purine content. V = Venezuela, CI = Côte d'Ivoire (Ivory Coast), T = Trinidad.

Figure 1 shows the relationship between theobromine to caffeine ratio, and caffeine content for the calibration samples. There was a clear separation between the Criollos on the one hand, and the Trinitarios and Forasteros on the other hand. Within the samples from Venezuela, a separation was found between the Criollo and Trinitario types. The samples ranged from the Forasteros (caffeine content 0.15% and theobromine 1.02%) to the Criollos (caffeine content 0.53% and theobromine 0.88 %), with the Trinitarios being intermediate between the other two types. This finding aligned with a genetic interpretation of the samples, as Trinitarios are a natural hybrid between Forasteros and Criollos. The sampling method brought out the genetic variability within the Trinitario type. A similar relationship was observed for the reference values.

Sample validation

In order to validate these observations, the model was tested on a set of 50 samples selected at random. The same partial least square analysis procedure was used to develop a calibration equation on the remaining 272 samples. The standard errors of prediction (SEP) were equal 0.06% and 0.1% and for caffeine and theobromine respectively.

Assignment of sample to type based on caffeine and theobromine content

The minimum, maximum and average values for caffeine and theobromine contents, and the reference analysis ratios and predictions, per genotype, are shown in Table 3.

Table 3. Descriptive statistics for caffeine and theobromine and content of the reference and NIR predicted values for the 50 validation samples.

Sample value	Cocao type	Caffeine content (%)			Theobromine content (%)			Ratio (Theobromine/Caffeine)		
		Min. value	Max. value	Mean value	Min. value	Max. value	Mean value	Min. value	Max. value	Mean value
Reference (n = 50)	Criollo	0.40	0.81	0.53	0.59	1.02	0.77	1.16	1.95	1.48
	Forastero	0.10	0.23	0.15	0.68	1.38	1.01	5.30	9.00	6.74
	Trinitario	0.18	0.43	0.30	0.69	1.27	0.99	2.00	4.44	3.38
Predicted (n = 50)	Criollo	0.46	0.65	0.52	0.58	1.10	0.79	1.22	1.78	1.49
	Forastero	0.06	0.30	0.16	0.84	1.41	1.07	4.37	14.45	7.95
	Trinitario	0.18	0.42	0.28	0.61	1.34	1.01	2.12	4.91	3.63

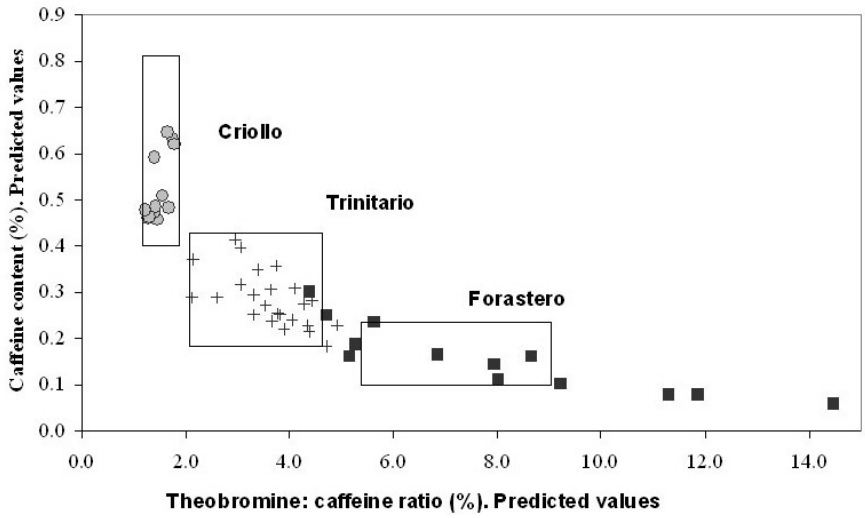

Figure 2. Distribution of the 50 validation samples depending on their predicted caffeine contents and their theobromine to caffeine ratios. Display of the zones defined by the reference values.

For each genotype, a zone was defined using the minimum and maximum reference values for the caffeine content and the theobromine to caffeine ratio. The validation samples were then projected onto those zones (Figure 2) according to their predicted caffeine content, and to their theobromine to caffeine ratio calculated from the predicted values.

Assignment of the Criollos based on their predicted values corresponded well to the zone defined by the reference values, and the Trinitarios were also well assigned, with three samples outside the zone and one borderline. Assignment of the Forasteros was clearly less obvious; indeed the variability zone was much more extensive than that defined by the reference values, the average predicted theobromine content was much higher.

Separation based on NIR spectral values was less marked between Trinitarios and Forasteros, but in both cases the separation of Criollos from the other two genotypes was clear.

Spectral discrimination

Forasteros can therefore be precisely distinguished from Criollos based on their purine contents predicted by NIR spectroscopy. This distinction was quite naturally tested on the basis of the spectral data, keeping the same set of validation samples. Using the first fifteen principal components extracted by principal component analysis from the spectra data, a step-by-step linear discriminant analysis enabled classification of the cultivars with a rate of 89.4 %.

In the step-by-step linear discriminant analysis the variable that maximized the Mahalanobis distance between the two closest groups was progressively introduced into the model [6]. In this analysis 10 principal components were introduced.

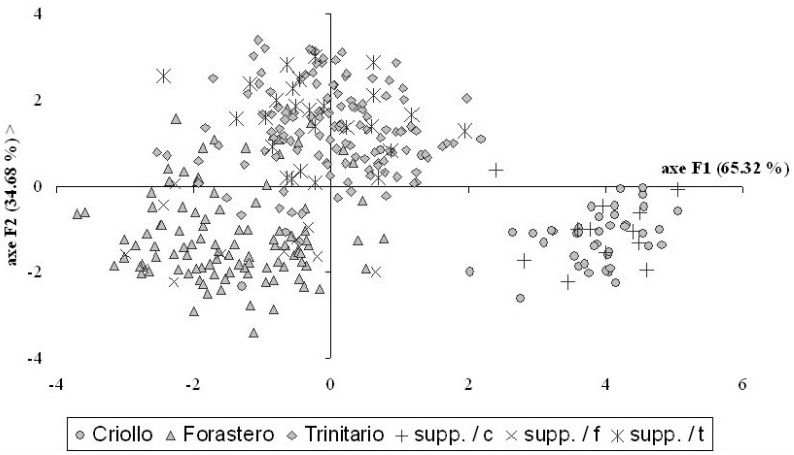

Figure 3. Scatter plot of the learning and validation (suppl./)) sample scores for the two discriminant functions.

The additional 50 samples were assigned to the correct genotype (Figure 3) group with a success rate of 96 %, according to their principal component analysis scores obtained by projection on the principal component analysis axes; two Forastero samples were misclassified as Trinitarios.

Conclusion

The purine content of fermented dried cocoa can be predicted by NIR spectroscopy. Criollo cocoa samples can be characterized by predicting their caffeine and theobromine contents.

Discrimination of the three genotypes directly based on spectral data was efficient (89% for learning and 96% for validation), though constructing a reference library based solely on qualitative data (certification of samples) would be dependent on the problem of distinguishing between pure Criollos and Trinitarios that were genetically close to Criollos, hence the merits of having a more conventional approach available (prediction of purine content), making it possible to interpret the results more objectively, and to check inconclusive samples in the laboratory.

In practice, the approach taken to characterize genotypes will have to combine these two approaches with the initial discrimination based on spectra validated by predicting purine content.

Consideration could be given to completing fine cocoa characterization by combining this spectral information with a sensory analysis of the samples, defining a standard profile which would then include "fine" Trinitarios.

When spectral mapping of samples for their purine content, their spectral fingerprints would need to be completed by identifying their geographical origin such as Criollos from Madagascar or Mexico; Forasteros from Venezuela, Indonesia or Ecuador; or Trinitarios from Cameroon.

The possible existance of a genotype by region of origin interaction also needs to be studied to find out, for example, how to determine the spectral profile of a specific genotype grown under different environmental conditions.

References

1. E.E. Cheesman, *Trop. Agric. Trin*. **22 (8)**, 144 (1944).

2. P. Lechevalier, in *Cacao et chocolat. Production, utilisation, caractéristiques*. Technique et Documentation, Lavoisier, Paris, France, p. 5 (1998).

3. E. Cros, in *Cacao et chocolat. Production, utilisation, caractéristiques*. Technique et Documentation, Lavoisier, Paris, France, p. 240 (1998).

4. J.N. Wintgens, in *2nd International Congress on Cocoa and Chocolate, Munich, Germany*. Hambourg: Behr's Seminars, p.40 (1991).

5. F. Davrieux, S. Assemat, R. Boulanger, E. Cros, in *Proceedings of the 14th International Cocoa Research Conference, Accra, Ghana*, **II,** 861 (2003).

6. M.F. Devaux, D. Bertrand, P. Robert, M. Quannari, *Appl. Spectrosc* **42**, 1015 (1988).