EXPLORATORY ANALYSIS FOR COCOA BEAN VARIETY IDENTIFICATION USING COLORIMETRIC AND ANTHOCYANIN CONTENT VARIABLES

T.C. Pires¹, T.S. Melo¹, L.C.R. Miranda², L.F.M.¹, E.S. Bispo¹

¹ Bromatological Analysis Department, School of Pharmacy, Federal University of Bahia - UFBA, Rua Barão de Jeremoabo, s/n, Ondina, CEP: 40171-970, Salvador, Brazil.
² Food Science Department, Food Engineering College, State University of Campinas - UNICAMP.

ABSTRACT - The aim of this paper is to identify different cocoa varieties (*Theobroma cacao* L.) using exploratory analysis. The main variables are colorimetric parameters L *, C *, h and anthocyanin content, quantified via the differential pH method. The varieties identified are PH16, TSH1188, CEPEC2002, BN34 and Catongo. We used these variables to apply Principal Components Analysis (PCA), which found that PC1 and PC3 components constituted 71.95% of the total data variance, and that PC1 and PC3 constituted three and two groups, respectively. The trends observed in the PCA analysis were confirmed by the dendogram obtained by hierarchical grouping analysis (HCA).

RESUMO - O objetivo deste trabalho foi identificar diferentes variedades de cacau (*Theobroma cacao* L), sendo estas PH16, TS1188, CEPEC2002, BN34 e Catongo através da análise exploratória, utilizando como variáveis principais, parâmetros colorimétricos L *, C *, h e teor de antocianinas, quantificadas pelo método de pH diferencial. Por meio dessas variáveis aplicou-se as análises de componentes principais (PCA), onde foi verificado que as componentes PC1 e PC3 descreveram 71,95% da variância total dos dados, sendo que a PC1 e PC3 discriminaram três e dois grupos, respectivamente. As tendências observadas na análise de PCA foram confirmadas através do dendrograma, obtido pela análise hierárquica de grupamento (HCA).

KEYWORDS: *Theobroma cacao*, multivariate analysis, flavonoids.

PALAVRAS-CHAVE: *Theobroma cacao*, análise multivariada, flavonoides.

1. INTRODUCTION

The cacao tree (*Theobroma cacao* L.) is a plant belonging to the Malvaceae family. Interest in its cultivation stems from the use of cocoa beans in chocolate production (Alves, 2002). There are three cocoa groups: Forastero, Criollo and Trinitarian. The Forastero group has purple cotyledons, due to its having pigmented cells and is commercially the most abundant. Cocoa Criollo features non-pigmented cotyledons. The Trinitarian group is a hybrid between the Criollo and Forastero and its cotyledons have color ranging from white to purple (Hancock, 1994; Bart-Plange & Baryeh, 2003).

With the advent of “witch's broom disease” caused by *Moniliophthora perniciosa* fungus that attacks cocoa plantations, breeders began to invest in breeding programs aiming to create disease-resistant varieties that also produced industrial quality raw materials.

Anthocyanins are a family of flavonoid compounds and are the group of pigments responsible for most of the colors in flowers, fruits, leaves, stems and plant roots, in addition to having
properties that associate their consumption with healthy eating habits (Pastrana-Bonilla et al., 2003; Falcão et al. 2007). According to Wollgast & Ankla (2000), depigmented seeds of cacao cotyledons (white or violet-light) have 33% lower phenolic compound content in comparison with pigmented seeds (intense purplish color).

Colorimetry uses instruments to numerically itemize, each component of a color’s composition. Among the most commonly used systems, one can cite the CIELAB (International Commission on Illuminant) 1976, which is obtained by chromaticity L* coordinates (brightness or lightness), a* and b* (hue or hue), C* (saturation or chroma) and h (ink angle). (Gouveia, 2008; Mori et al., 2004).

The derivation of information from an experiment’s data typically involves the analysis of a considerable number of variables. This can result in a large data set that is at times redundant or irrelevant to the experiment’s objective (Lopes et al., 2010). This leads to an increased need for more complex data processing techniques from a mathematical point of view.

According to Souza et al. (2006), the two best-known multivariate analysis techniques are hierarchical cluster analysis (HCA) and principal component analysis (PCA). HCA and PCA are exploratory methodologies aimed at highlighting similarities or differences between samples in a given data set. Thus, techniques of multivariate analysis were applied to the data set formed by variables measured through anthocyanin content analysis and colorimetric descriptors in order to identify the different cocoa bean varieties studied, thus contributing to the characterization of cocoa bean varieties.

2. MATERIAL AND METHODS

This work was developed in the Laboratory for Research and Analysis of Food and Contaminants (LAPAAC), Bromatological Analysis Department of the School of Pharmacy, Federal University of Bahia, Brazil.

We analyzed five cocoa bean varieties (PH16, TSH1188, CEPEC 2002 Catongo, BN34), obtained in a production farm located in the Ibirataia municipality, Bahia. All varieties were grown and pre-processed under the same conditions.

2.1 Extraction and Quantification of Anthocyanins

Extraction and quantification of the pigments was performed according to Teixeira et al. (2008). The samples were mashed and weighed before-hand giving 5.0 g per sample. 80 mL of extractor solvent was subsequently added (methanol-water (70:30)). Following from this, we added enough HCl to be able to adjust the medium’s pH to 2.0. The material was allowed to stand for 24 hours at 5 °C, protected from light. The material was manually filtered on filter paper and transferred to volumetric flask (100 mL), and then the volume was topped up with the extractor solvent, forming a concentrated extract.

The differential pH method was used for anthocyanin quantification, using UV/VIS spectrophotometer (BIOESPECTRO brand SP-220 model), making readings at a wavelength of 535 nm. The total anthocyanin content was expressed as mg anthocyanins.100g⁻¹ of cocoa beans. We used the average Extinction Coefficient (E1% 1cm) of several anthocyanins, adopting 873 and 775 respectively for the pH 1.0 and pH 4.5 (Fuleki & Francis, 1968a). The pH 1.0 solution was prepared from a mixture of KCl solution (0.2N) and HCl (0.2N) in the ratio (27:73). The pH 4.5 buffer was prepared from sodium acetate solution (0.2N) and HCl (0.2N) in the ratio (27:73). The pH 4.5 buffer was prepared from sodium acetate solution (1N) and water, in the ratio (40:60) and HCl (enough to adjust the medium to pH 4.5). Aliquots of 12.5 mL and 5 mL of concentrated extract were transferred to 25
mL and 10 mL volumetric flasks, respectively, having topped up its volume with pH 1.0 buffer solution in 25 mL flasks and the solution pH 4.5 in 10 mL flasks.

Calculation of Total anthocyanin content per 100 grams of the sample evaluated was obtained according to the following formula, adapting the values for the difference between the two pH reading methods. Equation quantification of anthocyanins:

\[ \text{Ant}^T_{mg/100g} = \frac{DO_{235nm} \times V_{E1} \times V_{E2} \times 1000}{V_{Alq} \times m \times E_{1cm}^{1%}} \]  

(Equation 1)

Where,  
DO: Optical Density of the diluted extract;  
V\text{E1}: Total Volume of the concentrated extract;  
V\text{E2}: Total Volume of diluted extract;  
V\text{Alq}: Aliquot Volume used in the dilution of the concentrated extract;  
m: Sample weight;  
\( E^{1\%}_{1cm} \): Average Extinction Coefficient.

2.2 Colorimetric analysis

For color, cocoa bean samples were placed in buckets in CIELAB, illuminant D65 and 10 observer, using 3 replicates per sample. In this color system, \( L^* \) represents the lightness (\( L^* = 0 \) - Black and \( L^* = 100 \) - white), \( C^* \) is the chromaticity and \( h^* \) is the angle of paint.

2.3 Multivariate analysis

The principal component analysis (PCA), hierarchical groupings (HCA) were applied to the data matrix using the variables \( L^* \), \( C^* \) and anthocyanin content; the \( h \) parameter was not used due to the high correlation (\( r = 0.98 \)) with \( C^* \). Auto-scaling was employed for data preprocessing. To obtain the dendogram we used the Euclidean distance and incremental connection methods.

3. RESULTS AND DISCUSSION

Cocoa bean varieties were evaluated for the parameters \( L^* \), \( C^* \), \( h \) and anthocyanin content. The averages of these parameters are on Table 1.

<table>
<thead>
<tr>
<th>VARIETY</th>
<th>Anthocyanin Content (mg.100g(^{-1}))</th>
<th>( L^* )</th>
<th>( C^* )</th>
<th>( h )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATONGO</td>
<td>2.82</td>
<td>38.52</td>
<td>29.15</td>
<td>54.75</td>
</tr>
<tr>
<td>BN34</td>
<td>10.59</td>
<td>31.25</td>
<td>20.04</td>
<td>48.90</td>
</tr>
<tr>
<td>TSH1188</td>
<td>6.56</td>
<td>32.07</td>
<td>17.27</td>
<td>45.69</td>
</tr>
<tr>
<td>PH16</td>
<td>11.22</td>
<td>36.86</td>
<td>24.51</td>
<td>50.85</td>
</tr>
<tr>
<td>CEPEC 2002</td>
<td>9.54</td>
<td>39.06</td>
<td>23.18</td>
<td>51.13</td>
</tr>
</tbody>
</table>

The \( L^* \) values ranged between 31.25 and 39.06. The \( C^* \) value ranged between 17.27 and 29.15. Anthocyanin content values were between 2.82 and 11.22, leading to greater variation in these
parameters. $h$ values ranged between 45.68 and 54.75. These were excluded to carry out the exploratory analysis because it was observed that in the projection of the PCA plans they indicated the same information as the C* parameter.

From the Principal Components Analysis (PCA), we found that PC1 and PC3 components accounted for 71.95% of the total variance of the data and provided discriminatory information for the samples. This was because the first principal component (PC1) accounted for 66.1% and the third 5.84% of the total data variance. PC1 was responsible for discriminating three of the five varieties due to color parameters and anthocyanin content. PC3 promotes the separation of two groups where the main weight of this component is in the C* and L* parameters. The joint projection of PC1 and PC3 is able to discriminate the five cocoa bean varieties evaluated in this study (Figure 1).

**Figure 1 - Scores of graphs (A) and weights (B) to principal component analysis**

The behavior presented by the BN34 and TSH1188 samples, ie, the discrimination of these varieties in relation to the others, was due to their presenting intermediate values for anthocyanin content and smaller values for L* and C* when compared to other cocoa bean varieties. However, the samples for the PH16 and CEPEC 2002 varieties showed high values for L* and C*, while PH16 had the highest anthocyanin content of all cocoa bean varieties. The group of the Catongo variety however, presented the highest values for C* and the lowest anthocyanin content.

The trends observed by the Principal Components Analysis (PCA) were confirmed by the dendogram obtained by HCA (Figure 02).

**Figure 02 - Dendogram (HCA) shows the dissimilarity between cocoa bean varieties**
It is possible to observe the formation of 6 groups on a relative distance of 20% since a sample of the variety shown BN34 is inconsistent with the other replicates.

The reason why the Catongo differed from the other groups is due to its being considered an albino mutation of the Forastero cocoa. Its main characteristic from a morphological point of view is a depigmentation of the cotyledons, which could lead to its classification as a Criollo cocoa. However, since the fruit has a hard skin, and the number of seeds per fruit is above 30, it is typically included in the Forastero population (Veríssimo, 2012). The depigmentation of the cotyledons is explained by the low anthocyanin content.

The varieties of existing cocoa trees in Southern Bahia are mostly cocoa Forasteros, introduced more than 260 years ago and cultivated by growers (Mello & Gross, 2013). According to Efraim (2009) the THS 1188 clone is a descendant of the Forastero and Trinitario groups. The PH 16 clone is a variety derived from selections made for commercial reasons. It has no known parents, and was originally identified in 1996 as belonging to a cocoa hybrid population, from crosses between interclonal cacao trees of the Amazon and Trinitario groups (Cruz, 2012). The groups that fit the CEPEC 2002 and BN34 varieties were not found in the literature. As noted in the test results, even belonging to the same Forastero and Trinitario hybrid group, the PH16 and TSH1188 varieties differed, showing that there are sets of characteristics that identify each variety.

4. CONCLUSION

Using exploratory analysis of data applied to results from simple and rapid analytic techniques, it was possible to obtain information about the dissimilarity among the cocoa bean varieties. This corroborated previous studies about the same differences. In terms of future research, the authors propose to use calibration models via multivariate techniques for the identification and discrimination of cocoa bean varieties as well as the study of other parameters for discrimination cocoa varieties.

5. BIBLIOGRAPHIC REFERENCES


Cruz, J. F. M. (2012). Caracterização das sementes de variedades de cacau Theobroma cacao L. resistentes à vassoura de bruxa durante a fermentação e após a secagem. (Dissertação de Mestrado), Universidade Federal da Bahia, Salvador.

Efraim, P. (2009). Contribuição à melhoria de qualidade de produtos de cacau no Brasil, por meio da caracterização de derivados cultivares resistentes à vassoura de-bruxa e de sementes danificadas pelo fungo. (Tese de Doutorado), Universidade Estadual de Campinas, Campinas.


