Food Control 131 (2022) 108347

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy

Carolina Hernández-Hernández [a], Víctor M. Fernández-Cabanás [b], Guillermo Rodríguez-Gutiérrez [c], África Fernández-Prior [c], Ana Morales-Sillero [b,*]

[a] Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Campo Experimental Huimanguillo, Km 1. Ctra. Federal Huimanguillo- Cárdenas, C.P. 86400, Huimanguillo, Tabasco, Mexico
[b] Departamento de Agronomía, ETSIA, Universidad de Sevilla, Ctra. Utrera Km 1, 41013, Seville, Spain
[c] Instituto de la Grasa, Consejo Superior de Investigaciones Científicas (CSIC), Campus Universitario Pablo de Olavide, Edificio 46, Ctra, de Utrera, Km. 1, 41013, Seville, Spain

ARTICLE INFO

Keywords:
Theobroma cacao L.
NIR spectroscopy
Unshelled beans
Phenols
Theobromine
Germplasm bank

ABSTRACT

Near Infrared (NIR) spectroscopy was evaluated as a fast and easy method for identifying the most interesting cocoa genotypes according to chemical composition, including bioactive compound contents, in more than 80 samples of cocoa beans from the Mexican Germplasm Bank, which were harvested over three years. Clear differences in chemical composition were detected in fermented and dried samples among genotypes. The contents in fat, protein, total sugars, total phenols, phenolic compounds and theobromine were determined for both whole cocoa beans and ground cotyledon samples; and calibration models were developed from the spectra of intact beans, nibs and ground cotyledons. In general, the calibration models obtained for cotyledon composition from the spectra of cocoa nibs and ground beans were better than those obtained from the spectra of intact beans. Fat content showed better calibration statistic values from the spectra of nibs and ground cotyledon ($r^2 = 0.70$). Bioactive compounds, such as theobromine ($r^2 = 0.77$), total sugars ($r^2 = 0.74$), total phenols ($r^2 = 0.66$) and derivatives of epicatechin ($r^2 = 0.88$), together with fat ($r^2 = 0.70$), protein ($r^2 = 0.64$) and husk content ($r^2 = 0.82$), were well-predicted using NIR spectroscopy in intact beans, cocoa nibs and/or ground cotyledon. The potential of NIRS technology was confirmed to support germplasm banks and breeding programs for the rapid identification of interesting genotypes based on their contents in bioactive compounds.

1. Introduction

Cocoa (*Theobroma cacao* L.) beans are currently produced in the tropics. Total production (5.253.376 tonnes) continues to increase due to the sustained demand for the consumption of cocoa products, popular for their sensory qualities and substantial health benefits (FAOFAST, 2020). The dry weight, grain size, color, fat content and sensory characteristics of fermented and dried cocoa cotyledons are of particular importance as they determine quality (BCCCA, 1996). Besides, cocoa beans have a high sugar content, which favor the growth of yeasts during fermentation and hydrolyse and degrade pectin and oligosaccharides and ferment monomers (Chaves-Lopez et al., 2014). Hydrolizated pectic and oligosaccharides are reported to have functional activities such as bifidogenic, prebiotic immunologic protection or cancer prevention (Bermúdez-Oria et al., 2019).

Cocoa beans are also especially rich in phenols (≈10% dry weight). These compounds contribute to cocoa's astringency and bitterness and numerous health benefits associated to the consumption of cocoa have been linked to them, i.e. antioxidant and anti-inflammatory properties, and modulation of atherosclerosis and hypertension. Three groups of phenols have been described in cocoa beans: catechins or flavan-3-ols, anthocyanins, and proanthocyanidins. Cocoa beans also contain methylxanthine alkaloids, such as theobromine and caffeine that contribute to the bitterness of cocoa and have several pharmacological effects (Carrillo et al., 2014). Genotype, geographical origin, maturity and also processing, in particular the fermentation and drying method, influence the bioactive composition of cocoa beans (Hernández-Hernández et al., 2018).

* Corresponding author.
E-mail addresses: hernandez.carolina@inifap.gob.mx (C. Hernández-Hernández), victorf@us.es (V.M. Fernández-Cabanás), guirogu@cica.es (G. Rodríguez-Gutiérrez), mafprior@ig.csic.es (Á. Fernández-Prior), amorales@us.es (A. Morales-Sillero).

https://doi.org/10.1016/j.foodcont.2021.108347
Received 12 February 2021; Received in revised form 9 June 2021; Accepted 10 June 2021
Available online 13 June 2021

The identification of the best cocoa genotypes from germplasm banks or breeding programs based on high-quality cocoa production has become a priority in traditional and new cacao producing countries, such as Mexico and Peru (Hernández-Hernández et al., 2018; Peláez et al., 2016; Teye et al., 2015). However, cocoa bean quality analyses, in particular the composition of bioactive compounds, require a lengthy sample preparation stage, a significant quantity of solvent, and are sample-destructive and expensive. On the other hand, the food industry requires fast, cheap and, if possible, non-destructive techniques for the selection of the most interesting bean samples in terms of sensory quality and nutraceutical characteristics. In this sense, near infrared spectroscopy (NIRS) has been proven to be a valuable tool for the control of cocoa quality attributes (Davies et al., 1991). NIRS has been applied for the quantification of different chemical constituents, the determination of the fermentation level and pH (Kutsanedzie et al., 2017), and for qualitative analyses, such as the classification of cocoa samples by origin and by pre- and post-harvest management (Teye et al., 2014). Quantitative calibrations have been reported for fat, moisture, protein and husk contents and, to a lesser extent, for phenols and methylxanthines (Álvarez et al., 2012; Hashimoto et al., 2018; Krähmer et al., 2015; Teye et al., 2015). However, most studies analyzed ground cotyledon more than whole beans (cotyledon plus husk), mainly to minimize the effects of the sample's physical properties in measured spectra (Quel-al-Vascónez et al., 2020).

The feasibility of employing NIR to determine cocoa bean and cotyledon quality from the spectra of intact cocoa beans, with special emphasis on bioactive compounds, has not yet been fully explored. Sunoj et al. (2016) demonstrated the potential of Fourier Transform NIRS (FT-NIRS) for the prediction of total phenols, fermentation index, and to a lesser extent pH, with the spectra of 24 different types of samples of intact beans. Recently, Barbin et al. (2018) highlighted the potential of NIR for the classification and differentiation of cocoa varieties from the spectra obtained for intact and ground cocoa beans (with husk), although the study involved just five varieties and bioactive compounds were not analyzed.

The hypothesis of this work was that NIRS might be a powerful tool for cocoa germplasm banks and breeding programs, allowing for the fast identification of interesting genotypes based on whole bean and cotyledon composition looking for nutritional and functional improvements, avoiding the troublesome step of milling a product with a high fat content. Our aim was to assess the potential of NIRS to classify intact cocoa beans according to the composition of both types of samples (beans and cotyledons) including bioactive compounds. Over a three-year period, more than 80 samples belonging to 63 different genotypes were obtained from the Mexican Germplasm Bank. Calibration models were developed and compared for intact beans, cocoa nibs and ground cotyledons.

2. Materials and methods

2.1. Samples

Dry cocoa bean samples were obtained from the Germplasm Bank of Mexico located in the experimental fields of Huimanguillo (Tabasco) and Rosario Izapa (Chiapas) of the National Institute of Agricultural and Livestock Forestry Research (INIFAP). A total of 85 samples were collected in 2013, 2014 and 2015. Cocoa pods were harvested manually at the optimum maturity stage when the placenta was completely detached from the husk. Subsequently, the pods were split to extract the grain, which was subjected to a microfermentation treatment. Microfermentation consisted of immersing a cloth mesh with 0.5 kg of fresh grain (sample) at the center of a 100 kg wooden box to follow the normal fermentation process for six days. The sample was then sun dried with constant stirring until a grain with a moisture content between 6 and 7% was obtained. For each sample, the weight was determined from the average value of 50 beans and a sub-sample of 250 g was transferred to the laboratory to determine the chemical analysis and the NIR spectra.

2.2. Chemical analysis

2.2.1. Sample preparation

After manual separation of the seed coats (i.e. the husks) from the beans, samples of cocoa cotyledon and cocoa bean husk were ground separately in a mill (IKA A11basic, Germany) and sieved to a 0.5 μm particle size. The contents in fat, protein, total sugars, total phenols, phenolic compounds, methylxanthines, as well as the antioxidant activity, were determined in both the cotyledon and husk samples.

2.2.2. Fat and protein content

The fat (g/100 g^{-1}) was extracted in 10 g of milled sample, in a semicontinuous extraction with hexane in Soxhlet equipment for 16 h. Total nitrogen content (g/100 g^{-1}) was determined with an elementary analyzer (LECO, USA) by the Dumas method, which consists of an internal combustion and thermoconductivity at 1000 °C. The protein content (g/100 g^{-1}) was calculated using the conversion factor 6.25 (ISO, 2016).

2.2.3. Total sugars and total phenols content

For the analysis of total sugars, total and individual phenols, and antioxidant activity in cotyledon and cocoa husk samples, a liquid extracts were obtained from 1 g of dry defatted sample, as described in a previous work by Hernández-Hernández et al. (2018) using ethanol: water at pH 3. Samples were stored at −18 °C until analysis.

Total sugars (mg of glucose g^{-1} of defatted sample) were determined by the anthrone method (Witham et al., 1971), at an absorbance of 655 nm, in a spectrophotometer (BIO-RAD iMark Microplate Reader, USA). Total phenols (mg of gallic acid g^{-1} of defatted sample) were determined by the Folin-Ciocalteu colorimetric method (Singleton & Rossi, 1965).

2.2.4. Theobromine and phenolic compounds

Theobromine and the phenolic compounds as catechin, epicatechin and its derivatives were analyzed using a Varian ProStar liquid chromatography system with a C-18 column (Kinetex Biphenyl 100 A, 250 mm × 4.6 mm, i.d. 5 μm) and diode array detector (DAD, the wavelength used for quantification was 280 nm) with automatic Rheodyne injection valves (20 μL loop). The method was previously described (Hernández-Hernández et al., 2018).

2.2.5. Antioxidant activity

The antioxidant activity of each phenolic extract was determined by the 2,2-diphenyl-1-picrylhydrazyl (DPPH) method, as previously described (Hernández-Hernández et al., 2018; Mrabet et al., 2016). The amount of antioxidant needed to reduce the initial absorbance by 50% (EC_{50}, effective concentration, mg mL^{-1}) was calculated from a calibration curve according to the linear regression of each antioxidant solution.

2.3. Scanning of NIR spectra

The reflectance spectra were obtained for the cocoa samples from intact beans, cocoa nibs and ground cotyledons. A Foss-NIRSystem 6500 SY-II monochromator device (Foss NIRSystems, Silver Spring, MD) was used to scan samples from 400 to 2498 nm every 2 nm (spectral band pass 10 nm ± 1 nm). Analyses were carried out using a transport module. Intact cocoa beans and cocoa nib samples were scanned using a rectangular sample cup (Natural Product Sample Cup IH-0331) with dimensions of 4.7 cm in width and 20 cm in length. The samples of ground cotyledon were scanned using a microcapsule with dimensions of 6 cm in width and 10 cm in length. The spectrum of each type of sample was the average of the spectra of two sub-samples.

The spectra were manipulated and processed to obtain the calibration equations using WINISI software v. 1.5 (Infrasoft International,

State College, PA).

2.4. Calibrations development

For development of calibrations, the spectra of intact cocoa beans were correlated with the compositions of both intact bean and dehusked cotyledon. Cocoa nibs and ground cotyledon spectra were also correlated with data of dehusked cotyledon composition.

Scatter phenomena were corrected by Standard Normal Variate (SNV) and Detrending (DT) mathematical pre-treatments (Barnes et al., 1989). To select the best calibration, five spectral derivatives were compared: 1,5,5,1; 1,10,5,1; 1,10,10,1; 2,5,5,1 and 2,10,5,1. The first digit refers to the order of the derivative, the second is the gap where the derivative was calculated, the third is the smoothing segment, and the fourth is the second smoothing segment (Marten et al., 1989). All calibrations were obtained from two different spectral ranges: visible + NIR (400–2498 nm) and NIR (1100–2498 nm).

Modified Partial Least Square Regression (MPLSR) was used to develop the calibration models (Shenk & Westerhaus, 1996) and cross-validation method was used to determine the optimum number of PLS terms in the regression models and to avoid overfitting.

Equations were developed by carrying out two outlier filters (T and H) before completing the final calibration, since anomalous data have a significantly greater effect in the model than standard samples, and thus can result in a poor performance for most samples. T outliers are samples with significant differences between laboratory values (reference values) and predicted values, while H outliers are samples with spectra showing excessive distance ($H > 3$) from the spectra center of the calibration set (Shenk & Westerhaus, 1996).

The predictive ability of the models was evaluated by examining the values of the calibration statistics obtained: the standard error of calibration (SEC), the coefficient of determination for calibration (R^2), the standard error of cross-validation (SECV), the coefficient of determination for cross-validation (r^2), the RPD, which is the ratio of the standard deviation of the original data (SD) to the SECV, and the RER, which is the ratio of the range in the reference data to the SECV. Values of RPD between 2.5 and 3 and higher than 3 indicates good and excellent prediction accuracy, respectively; between 2 and 2.5, coarse quantitative predictions are possible; between 1.5 and 2.0, it is possible to distinguish high and low values, and below 1.5 it means that calibrations are not useable. RER values should be 10 or higher for good predictive abilibility (Williams & Sobering, 1996).

Five mathematical pre-treatments were applied, and the best calibration model was selected based on the highest r^2 value and the lowest SECV for each of the studied constituents. Although the use of external validations is a common tool for the evaluation of NIR predictive models, we found some difficulties trying to perform it with the collective as the reduced number of samples (only 85 selected according to a similar post-harvest fermentation procedure; we had to discard other 60 samples for this reason) and the variability in genotypes (63), cropping seasons (3) and growing places (2) led to important differences intra and inter cropping seasons. Due to this extremely high variability in the studied samples and taking into account that the estimation of bioactive principles from the NIR spectra is complex and requires a broad calibration data set, it was not possible to select an adequate validation group without altering the variability in the calibration group. We reported this same problem in a previous work with cocoa husk (Hernández-Hernández et al., 2021), where we also had to perform a comparison of results based on SECV. For this reason and considering that SECV is the single best estimate of the prediction capability of the equation, being this statistic similar to the average Standard Error of Prediction (SEP) from 10 randomly-chosen prediction sets (Shenk & Westerhaus, 1996), it was decided to use cross-validation instead of an external validation to estimate the predictive ability of the models in this feasibility study.

2.5. Statistical analysis

The mean values, range, standard deviation, and coefficient of variation were determined for each parameter, and the correlation coefficients among the mean values were assessed by the Pearson's Correlation test. Statgraphics Centurion XVI v. 16.1.15 software was used.

3. Results and discussion

3.1. Bean weight, percentage of husk and chemical composition

Cocoa beans and cotyledons had high contents in fat and proteins and were rich in sugars, phenols and theobromine (Table 1). As reported in previous works (Hernández-Hernández et al., 2018, 2019), phenols such as catechin, epicatechin, derivatives from epicatechin, which are compounds that maintain most of their chemical structure after a chemical or enzymatic reaction, and methylxanthines such as theobromine, were identified. Theobromine was the major bioactive compound detected in our samples and the only methylxanthine. Caffeine, which has also been described as a methylxanthine present in cocoa (Carrillo et al., 2014), was not identified in any of the 85 samples tested. Differences in the composition of bioactive compounds in intact beans compared to the cotyledon can be associated to the husk, the main by-product from the production of cocoa products, which represented around 18% of the seed weight.

Most of the parameters analyzed in both the whole bean and cotyledon samples showed a high variability in their contents in bioactive compounds. This variability is probably due to genotype and to the influence of environment conditions, as samples came from different genotypes (63), growing places (2) and were harvested over three years. The processing of the beans may also influence this composition (Oracz et al., 2015), although in this study, all the cocoa bean samples were processed in a similar way. Epicatechin derivatives I, II and III presented the highest coefficients of variation (CV) in both cases (>100%), while the CV for catechin, total sugars, total phenols, epicatechin, antioxidant activity, and theobromine varied between 40 and 70%, approximately. These CV values are important for developing NIRS calibrations. In contrast, the bean weight, husk proportions, and the contents in fat and protein showed the lowest coefficients of variation (<30%). The mean values for fat and protein varied from 19 to 64 g/100 g^{-1} and from 11 to 32 g/100 g^{-1}, respectively. Despite their low CV, some samples (data not shown) were identified with higher contents in fat (Afoakwa et al., 2013; Liendo et al., 1997; Vera et al., 2014) and protein (Álvarez et al., 2012; Hashimoto et al., 2018; Liendo et al., 1997) than those previously reported for cocoa genotypes grown in other producing countries. Regarding bioactive compounds, total sugars varied from 29 to 271 mg g^{-1}, total phenols fluctuated between 16 and 186 mg g^{-1}, and theobromine between 10 and 110 mg g^{-1}. For phenols and theobromine, some samples also stood out for their higher contents in comparison with those reported in samples from other countries (Carrillo et al., 2014; Oracz et al., 2015). Epicatechin, the major phenol found in the analyzed samples, showed contents of 6–61 mg g^{-1}, with the content in some samples also being higher than the 2–17 mg g^{-1} reported by Caligiani et al. (2010) and Kim and Keeney (1984), respectively.

Cross correlation coefficients between the values of all the physical-chemical parameters determined in intact beans and cotyledons showed quite similar results (Tables 2 and 3). Fat was moderately and significantly correlated (0.4–0.6) with total phenols, derivatives I, II and III, and theobromine, and negatively correlated with protein and total sugars (−0.3 to −0.5). Protein was moderately and positively correlated with catechin (0.4–0.5) and negatively correlated with the contents of derivatives I, II and III (−0.4 to −0.5). Highly significant and positive correlations (0.8) were found between derivative I and derivatives II and III, and between derivative II and derivative III. Total phenols showed a positive correlation with antioxidant capacity, or in other words, a negative correlation with EC_{50} values (−0,6). Thus, a high antioxidant

Table 1
Descriptive statistical analysis of the parameters analyzed in cocoa beans and cotyledons (N = 85).

	Intact beans					Cotyledon				
Constituent[a]	Range	Mean	SD	CV	SEL	Range	Mean	SD	CV	SEL
Bean weight (g)	0.65–2.20	1.18	0.29	24.51	–	0.50–1.88	0.96	0.24	24.77	–
Husk (g/100g^{-1})	9.76–31.19	18.13	4.44	24.51	–	–	–	–	–	–
Fat (g/100g^{-1})	18.65–49.48	37.00	7.04	19.03	3.84	22.53–64.30	44.90	9.37	20.86	4.44
Protein (g/100g^{-1})	11.93–29.13	20.57	4.98	24.20	2.47	11.13–31.67	21.63	5.91	27.33	3.18
Total sugars (mg g^{-1})	31.34–239.48	101.73	54.24	53.31	11.41	29.27–271.28	105.88	60.12	56.78	12.98
Total phenols (mg g^{-1})	15.22–167.59	53.49	25.58	47.82	6.07	15.93–186.15	59.75	28.91	48.38	7.37
Catechin (mg g^{-1})	0.68–9.14	2.32	1.42	61.14	0.52	0.34–10.57	2.40	1.68	70.00	0.63
Epicatechin (mg g^{-1})	6.61–54.64	20.73	8.97	43.26	3.17	6.13–60.51	21.65	10.12	46.77	3.79
Derivative I (mg g^{-1})	0.00–8.02	2.13	2.15	100.87	0.22	0.00–10.11	2.53	2.63	103.70	0.27
Derivative II (mg g^{-1})	0.00–11.97	1.80	2.35	130.27	0.22	0.00–14.74	2.20	2.91	132.53	0.26
Derivative III (mg g^{-1})	0.00–12.84	2.00	3.26	162.80	0.25	0.00–15.26	2.50	4.11	164.17	0.33
Theobromine (mg g^{-1})	9.71–95.71	37.67	17.77	47.19	2.54	9.79–109.94	41.09	20.64	50.23	3.09
Antioxidant activity (EC_{50})	2.05–22.99	10.22	4.72	46.21	0.76	0.00–16.85	6.78	3.54	52.27	0.71

SD: Standard deviation.
CV: Coefficient of variation (100*SD/Mean).
SEL: Standard Error of Laboratory.
[a] All chemical constituents were determined in triplicate.

capacity could be indicative of a higher phenolic content as it has been previously reported (Hernández-Hernández et al., 2018). Theobromine concentration also showed a positive correlation with antioxidant activity, or negative with EC50 (−0.5), which reflects the good antiradical properties of this compound, as previously described (Azam et al., 2003).

3.2. Spectral characteristics

Fig. 1 illustrates the visible and NIR average spectrum of intact cocoa beans, beans husk, cocoa nibs and ground cotyledon. Absorption peaks in the visible region of 446 and 678 nm could be related to the presence of chlorophyll and carotenoid pigments. Main absorption peaks found in the NIR region coincided with those described by Caporaso et al. (2018) at 1208, 1397–1437, 1724–1743, 1919 and 2307–2326 nm. Absorbance around 1208 nm could be associated with carbohydrates, but differences were observed in this zone for intact and shelled beans, suggesting the presence of carbohydrates linked with fat (Davies et al., 1991). In fact, fat is the main constituent in cocoa, and responsible for many of the aforementioned absorption peaks with 1724–1743 and 2307–2326 nm (Caporaso et al., 2018). Peaks in these regions are representative of lipids and are sharper for shelled products, indicating a higher fat content in the absence of husks. Carbohydrates are the second largest constituent in cocoa and are responsible for the shoulder located at 1397 nm, assigned to the absorption of $–CH_2$ groups, the peaks around 1440 nm, linked to the absorption of C-H bond, and the peaks at 1919 nm, associated with carbonyl groups (Caporaso et al., 2018). However, this former band is close to 1923 nm, where the water O-H bond has a high absorbance, and so should be assigned to moisture content as is the case for the peak at around 1940 nm (Davies et al., 1991). Absorbance at around 2057 nm is assigned to amides, and so can be attributed to protein content (Caporaso et al., 2018). Bands associated with minor constituents like bioactive compounds are difficult to detect by simple visual inspection. However, reference has been made to spectral regions associated with total phenols in the ranges of 1000–1041, 1349–1386, 1661–1718, 2162–2258 nm and to theobromine contents in the range of 2222–2488 nm (Hashimoto et al., 2018). The main absorption bands for theobromine have also been described at 1292, 1628, 1764, 1836, 2092 and 2228 nm; and epicatechin absorption bands at 1388, 1492, 1658, 1916, 2260 and 2324 nm (Álvarez et al., 2012).

The spectral patterns of the average spectra for each type of cocoa sample (intact cocoa beans, beans husk, cocoa nibs and ground cotyledons) were highly variable. The differences in spectra could be attributed to discrepancies in particle size and to the presence or absence of the husk, as this principal by-product from the cocoa industry is rich in fiber and bioactive compounds (Hernández-Hernández et al., 2019). In particular, the peaks corresponding to fat are more accentuated in the intact and ground cotyledon, while there is some similarity between the spectrum of the husk and the intact grain. The latter is because the radiation interacts first with the husk that covers intact grain and its penetration decreases exponentially with the depth (Pissard et al., 2018). Almost similar spectral patterns were shown by Hashimoto et al. (2018) for ground cocoa beans (cotyledons plus husks) and ground cocoa nibs (cotyledons).

3.3. NIR calibration results

Thirty-four regression models were developed for the three available spectral sets (intact beans, cocoa nibs and ground cotyledons). Calibration statistics obtained from the spectra of intact cocoa beans are displayed in Table 4; whereas calibration statistics for cocoa nibs and ground cotyledon are shown in Table 5. The spectral region that presented the best results was the visible + NIR region of the spectrum (400–2498 nm). For mathematical derivatives, the second derivatives (2, 5, 5, 1 and to a lesser extent 2, 10, 5, 1) were the most interesting, specifically for cotyledons.

The cocoa industry includes commercial standards concerning bean weight and husk content among others qualitative attributes because they can influence the manufacturing performance or flavor. A high proportion of husk can negatively affect the flavor and taste of some final products of the cocoa transformation, and cause equipment abrasion during the grinding process (Quelal-Vásconez et al., 2020). However, cocoa bean husk is a potential source of natural bioactive compounds and nowadays many industries are interested in their extraction. Recently, the feasibility of NIR for the prediction of these compounds in cocoa bean husk has been demonstrated (Hernández-Hernández et al., 2021).

Model calibrations developed in this work for bean and cotyledon weight showed limited results, giving an r^2 of 0.49 and 0.43, SECV of 0.18 and 0.15, RPD of 1.38 and 1.33, and RER of 8.61 and 9.20, respectively. This means that these parameters were not well predicted. It is important to state that both the cocoa bean and cotyledon weights were not estimated using NIR techniques in previous studies. Probably, more analyses should be done using a larger number of samples in order to improve the performance of the models. Other works as that carried out by Morales-Sillero et al. (2011) reported values of r^2 between 0.86 and 0.92, SECV between 0.29 and 0.55, and RPD 2.68 and 3.44, when analysing more than 400 samples to predict olive fruit weight.

The husk content in the intact bean were better predicted as r^2, SECV, RPD and RER reached values of 0.82, 1.68, 2.33 and 12.88, respectively

Table 2
Pearson's correlation among different traits in intact beans.

Constituent	Bean weight	Husk content	Fat	Protein	Total sugars	Total phenols	Catechin	Epicatechin	Derivative I	Derivative II	Derivative III	Theobromine
Husk content	0.03ns											
Fat	0.04ns	0.05ns										
Protein	0.02ns	−0.44****	−0.41***									
Total sugars	0.29**	−0.07ns	−0.34**	0.33**								
Total phenols	0.12ns	−0.12ns	0.42****	−0.07ns	0.00ns							
Catechin	−0.03ns	−0.17ns	−0.15ns	0.46****	−0.04ns	−0.14ns						
Epicatechin	0.34**	−0.07ns	0.20ns	0.16ns	0.06ns	0.16ns	0.16ns					
Derivative I	0.32**	0.34**	0.47****	−0.50****	0.11ns	0.30**	−0.27**	0.19ns				
Derivative II	0.22*	0.25*	0.46****	−0.40****	0.12ns	0.36***	−0.25*	0.20ns	0.79****			
Derivative III	0.30**	0.26*	0.53****	−0.47****	0.02ns	0.43****	−0.23*	0.22*	0.79****	0.84****		
Theobromine	0.04ns	−0.13ns	0.46****	−0.11ns	−0.47****	0.46****	0.03ns	0.18ns	0.09ns	0.09ns	0.12ns	
Antioxidant activity (EC_{50})	−0.25*	0.29**	−0.37***	0.14ns	0.01ns	−0.64****	0.20ns	−0.18ns	−0.31**	−0.34**	−0.39***	−0.46****

ns: non-significant*, **, ***, and **** Significant at $P \leq 0.05$, 0.01, 0.001, and 0.0001, respectively.

(Table 4), which means that coarse quantitative prediction is possible with the model. Hashimoto et al. (2018) showed the feasibility of NIR for predicting the content of cocoa bean husk after analyzing 81 intact samples of fermented cocoa beans from Brazil supplied by the industry over the period of one year, although the spectra were obtained from milled and sieved bean samples. The calibration model that they obtained after cross-validation showed an r^2 value of 0.76. They determined statistics as RMSECV (root mean square error of cross-validation) and RER, with values of 1.37 and 6.7, respectively, which means that the model was suitable for screening purposes.

The calibration models obtained to estimate the contents in fat and protein from the spectra of intact cocoa nibs and ground cotyledons were rather similar (Table 5). For fat, the best r^2, SECV, RPD, and RER values were found with the spectra for intact nibs (0.70, 5.13,1.82, and 8.14, respectively). In the case of protein, no differences were observed between both type of spectra, with r^2, SECV, RPD and RER values around 0.6, 3.6, 1.6, and 5.5, respectively. For both traits, the values for RPD (close to 2), and those for r^2 and SECV suggest that quality control could be carried out with NIRS for screening purposes, with more accuracy in this kind of samples as compared to intact beans (Table 4). Nevertheless, the models obtained from the spectra of ground cotyledons have no real advantage compared to the cocoa nib models, since highly time-consuming sample preparation would be required.

The potential for NIR to predict fat and protein contents has been also demonstrated by Hashimoto et al. (2018) in the abovementioned work. The calibration models for fat content (with 0.67, 0.97 and 8, as the values of r^2, RMSECV and RER, respectively) were developed with the spectra of ground beans (cotyledons + husk). A calibration model for protein was developed with the spectra from ground cotyledons and it also showed good results (statistic values were $r^2 = 0.75$, RMSECV = 0.34 and RER = 9). Barbin et al. (2018) indicated the feasibility of NIR for predicting fat and protein contents with the spectra of intact and ground fermented beans. For whole beans, their fat models after cross-validation gave r^2, RER and RPD values higher than 0.9, 10, and 3.2, while protein models gave values higher than 0.93, 12, and 3.8, respectively. Nevertheless, those results are not properly comparable with ours because that study was carried out with samples from just five varieties. Thus, the authors indicated that the fat content varied between 41 and 48% and the protein content varied between 12 and 15%. A much higher variability was considered in our work (Table 1). Álvarez et al. (2012) showed good prediction of fat (r^2, SEV and RPD values of 0.94, 0.89 and 3.4, respectively) by analyzing 247 samples from Criollo cocoa beans with the spectra of ground cotyledons. However, the samples were unfermented and, again, the variability was lower than in our study as all the samples came from a single plantation.

The calibrations selected for total sugars showed better results with the spectra of nibs or ground cotyledons (Table 5). The values for r^2, SECV, RPD and RER were around 0.7, 30, 1.90 and 8.0 respectively, in contrast with the models developed with the spectra of intact beans, in which the values for both statistics were around 0.5, 40, 1.39 and 5.5, respectively (Table 4).

For total phenols and the methylxanthine theobromine, the calibrations statistics obtained from the spectra of intact beans presented better results than those developed from the spectra of cotyledons (nibs or ground) (Tables 4 and 5). During fermentation, phenolic compounds and theobromine migrate out of the cotyledon and into the husk. This is why the concentration of these components in the husk and in the uppermost parts of the cotyledon increases (Hernández-Hernández et al., 2019). In the case of total phenols, the r^2 and SECV were 0.66 and 12.49 respectively for whole beans (cotyledons plus husk), and 0.57 and 16.23 for cotyledons (Table 4). For theobromine, the value for r^2 was 0.77 in beans and 0.69 in cotyledons and the SECV values were 7.37 and 9.9, respectively. RPD and RER values in beans were 1.69 and 12.20 respectively for total phenols, and 2.11 and 11.67 respectively for theobromine, which means that the models developed from the spectra of intact beans were able to discriminate between low and high values

Table 3
Pearson's correlation among different traits in cotyledons.

Constituent	Cotyledon weight	Fat	Protein	Total sugars	Total phenols	Catechin	Epicatechin	Derivative I	Derivative II	Derivative III	Theobromine
Fat	−0.04ns										
Protein	0.11ns	−0.53****									
Total sugars	0.22*	−0.45****	0.42****								
Total phenols	0.13ns	0.36***	−0.08ns	−0.05ns							
Catechin	0.01ns	−0.17ns	0.44****	0.02ns	−0.09ns						
Epicatechin	0.32**	0.17ns	0.18ns	0.06ns	0.15ns	0.18ns					
Derivative I	0.24*	0.53****	−0.48****	−0.03ns	0.26*	−0.23*	0.18ns				
Derivative II	0.14ns	0.51****	−0.40***	0.02ns	0.34**	−0.20ns	0.19ns	0.76****			
Derivative III	0.21ns	0.58****	−0.47****	−0.09ns	0.41***	−0.18ns	0.20ns	0.75****	0.83****		
Theobromine	0.06ns	0.34**	−0.08ns	−0.51****	0.44****	0.11ns	0.20ns	0.04ns	0.06ns	0.08ns	
Antioxidant activity (EC_{50})	−0.25*	−0.15ns	−0.07ns	0.05ns	−0.57****	0.08ns	−0.19ns	−0.18ns	−0.12ns	−0.18ns	−0.47****

ns: non-significant*, **, ***, and **** Significant at P ≤ 0.05, 0.01, 0.001, and 0.0001, respectively.

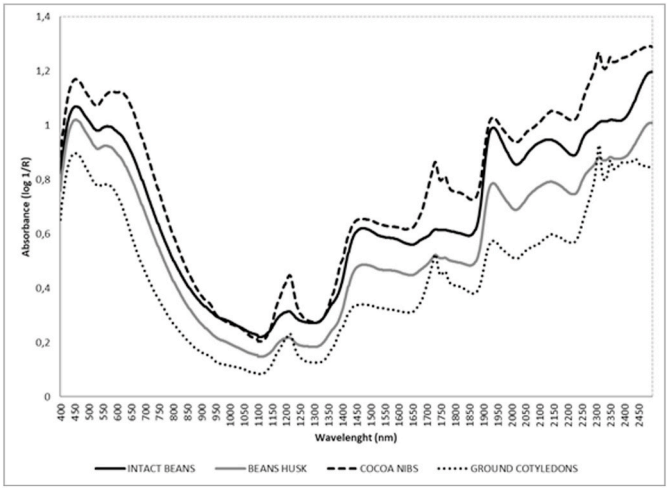

Fig. 1. Average spectra (visible + NIR) of intact cocoa beans, beans husk, cocoa nibs and ground cotyledon.

for total phenols and that a coarse quantitative prediction was possible for theobromine.

Regarding phenolic compounds, the models obtained for derivatives of epicatechin I, II and III, showed r^2 values around 0.8 and SECV close to 1 with the spectra from nibs or ground cotyledons (Table 5). In addition, the derivatives I and II were also well predicted in intact beans and cotyledons with the spectra obtained from intact beans (Table 4). For derivative III, r^2 and SECV values were 0.6 and around 1.3, respectively with these spectra. The RPD values were around 2 or ever higher and the RER values were equal or higher than 10, indicating an adequate predictive ability for those parameters. Despite the fact that the highest RPD values were obtained with nibs or ground cotyledons, the use of models developed with intact product spectra would mean a significant saving in time. For catechin, the best values of r^2 and SECV were found in the models obtained from the spectra of nibs (0.47 and 0.71, respectively) as can be seen in Table 5. Lower values for these statistics were found in the models for ground cotyledon and with the spectra of intact beans (Table 4). In the case of epicatechin, despite being the major constituent, calibration models showed poor results. The best r^2 and SECV values were around 0.2 and 6.5, which were obtained from the spectra of nibs and ground cotyledons (Table 5).

Few works have studied the potential of NIR for the prediction of the above bioactive compounds in cocoa and only two used the spectra of intact beans. In particular, good predictions for carbohydrates (values of r^2, RER and RPD higher than 0.86, 7.4, and 2.7, respectively) were reported in the mentioned study of Barbin et al. (2018), in which the calibrations were developed with samples from just five cocoa varieties. In the work by Sunoj et al. (2016), the authors reported values for r^2, RMSECV and RPD close to 0.8, 0.9 and 2.5, respectively, by using FT-NIR to predict total phenols during the fermentation process of cocoa beans which proceeded from an Indian plantation. In other works, the calibrations were developed with the spectra from ground beans or ground cotyledons. Thus, with the ground beans' spectra, Krahmer et al. (2015) showed good performance after cross-validation for carbohydrates (r^2 = 0. 82, REMSECV = 0.32 and RPD = 2.35), total phenols (r^2 = 0. 93, REMSECV = 0.25 and RPD = 3.8) and theobromine (r^2 = 0.79, REMSECV = 0.14 and RPD = 2.19), although samples came from only nine biological replicates and were analyzed at different times during the fermentation process. Concerning the use of ground cotyledon spectra, Hashimoto et al. (2018) reported good predictions for total phenols (r^2 = 0.89, REMSECV = 0.63 and RER = 12) and theobromine (r^2 = 0.77, RMSECV = 0.13 and RER = 13), although the diversity of the samples used was lower than in our work; the main differences among samples were related in a great extent to growing regions and fermentation processing, but not to genetic variation. Álvarez et al. (2012) also reported the good performance of calibration models for theobromine (r^2 = 0.88, SECV = 0.08 and RPD = 2.5). However, the samples were unfermented and, again, the variability was lower than in our study as all the samples (247) came from the same plantation in Venezuela.

Literature regarding the prediction of phenolic compounds by NIRS has been mainly focused on epicatechin, cocoa's major component, followed by catechin. We have found only two published studies and this compound was well predicted, although most of samples were not fermented. Alvarez et al. (2012) reported an r^2 value of 0.96, SECV of 0.18, and RPD of 2.3; whereas Krähmer et al. (2015) reported an r^2 value of 0.93, RMSECV of 0.22 and RPD of 3.69. The results found in our work are, therefore, unforeseen, but they could be explained by the type of cocoa bean sample, which was unfermented in most of the above-mentioned studies. The concentration of phenols in the cotyledon diminishes during the fermentation process, being transferred to the husk, and probably linked to cell wall material like sugars. In this sense, both the concentration and the way in which the phenols are present in fermented and unfermented cotyledon are different.

The high predictability of models for the contents of total phenols, epicatechin derivatives and theobromine can be explained by the significant correlations that were found between them and the content in fat (Tables 2 and 3). The correlation of the fat content to the contents in total phenols, epicatechin derivatives I, II and III, and theobromine, were of 0.42, 0.47, 0.46, 053, and 0.46, respectively in intact beans, and of 0.36, 0.53, 0.51, 0.58, and 0.34, respectively in cotyledons. Besides, the catechin and epicatechin were not significantly correlated with fat (values of correlation coefficients were of −0.15 in intact beans and −0.17 in cotyledons for catechin; and of 0.20 and 0.17, respectively, for epicatechin).

Table 4
Calibration statistics for the selected equations obtained from the spectra of intact beans.

	Intact bean composition										Cotyledon composition									
Constituent	Derivative	N	Mean	SD	SEC	R^2	SECV	r^2	RPD	RER	Derivative	N	Mean	SD	SEC	R^2	SECV	r^2	RPD	RER
Weight	1,10,10,1	77	1.16	0.25	0.14	0.69	0.18	0.49	1.38	8.61	1,5,5,1	75	0.95	0.20	0.11	0.68	0.15	0.43	1.33	9.20
Husk	2,5,5,1	75	17.61	3.92	1.27	0.89	1.68	0.82	2.33	12.88	–	–	–	–	–	–	–	–	–	–
Fat	2,10,5,1	79	37.01	7.08	3.58	0.74	5.68	0.35	1.25	5.43	2,5,5,1	83	44.61	9.43	3.95	0.82	6.39	0.54	1.48	6.54
Protein	1,10,5,1	78	20.40	4.92	2.68	0.70	3.16	0.59	1.56	5.44	2,5,5,1	85	21.61	5.91	3.43	0.66	4.20	0.49	1.41	4.89
Total sugars	2,5,5,1	79	100.05	52.44	33.83	0.58	37.94	0.48	1.38	5.49	2.5.5.1	78	102.22	56.18	25.38	0.80	40.49	0.48	1.39	5.98
Total phenols	2,5,5,1	76	51.11	21.13	7.69	0.87	12.49	0.66	1.69	12.20	2,5,5,1	81	56.75	24.23	9.37	0.85	16.23	0.57	1.49	10.49
Catechin	2,10,5,1	73	1.98	0.81	0.56	0.52	0.71	0.23	1.13	11.92	2,5,5,1	73	2.02	0.97	0.63	0.58	0.81	0.31	1.20	12.63
Epicatechin	2,5,5,1	73	18.86	6.42	6.13	0.09	6.61	−0.07	0.97	7.27	2,10,5,1	75	20.42	8.57	5.34	0.61	7.78	0.19	1.10	6.99
Derivative I	2,10,5,1	72	1.83	1.86	0.43	0.95	0.81	0.81	2.29	9.90	2,5,5,1	71	2.09	2.17	0.66	0.91	0.97	0.80	2.23	10.42
Derivative II	2,5,5,1	69	1.41	1.79	0.54	0.91	0.84	0.78	2.13	14.25	2,5,5,1	74	1.88	2.38	0.89	0.86	1.26	0.72	1.90	11.70
Derivative III	2,10,5,1	71	1.20	1.92	0.96	0.75	1.22	0.60	1.58	10.52	2,10,5,1	70	1.37	2.24	1.16	0.73	1.44	0.58	1.56	10.60
Theobromine	2,5,5,1	73	35.21	15.55	4.19	0.93	7.37	0.77	2.11	11.67	2,5,5,1	76	38.62	17.76	5.83	0.89	9.90	0.69	1.79	10.10
Antioxidant activity (EC_{50})	2,5,5,1	78	9.94	4.42	2.18	0.76	3.43	0.41	1.29	6.10	1,5,5,1	79	6.77	3.75	1.94	0.73	2.34	0.61	1.60	7.20

N: Number of samples used for calibration.
Mean: Mean of the calibration series.
SD: Standard deviation.
SEC: Standard error of calibration (in actual values).
R^2: Determination coefficient for calibration.
SECV: Standard error of cross validation (in actual values).
r^2: Determination coefficient for cross-validation.
RPD: Ratio SD/SECV.
RER: Ratio (max-min)/SECV.

Table 5
Calibration statistics for the selected equations from the spectra of intact cocoa nibs and ground cotyledons.

	Cocoa nibs										Ground cotyledon									
Constituent	Derivative	N	Mean	SD	SEC	R^2	SECV	r^2	RPD	RER	Derivative	N	Mean	SD	SEC	R^2	SECV	r^2	RPD	RER
Fat	2,5,5,1	83	44.42	9.35	4.57	0.76	5.13	0.70	1.82	8.14	2,10,5,1	81	44.97	8.94	4.05	0.79	5.15	0.67	1.74	8.11
Protein	2,5,5,1	85	21.86	5.97	3.14	0.72	3.71	0.62	1.61	5.54	2,10,5,1	83	21.49	5.90	2.92	0.76	3.53	0.64	1.67	5.82
Total sugars	2,5,5,1	81	105.77	59.89	20.26	0.89	30.67	0.74	1.95	7.89	1,10,10,1	80	97.61	51.21	23.68	0.78	27.14	0.72	1.89	8.92
Total phenols	2,5,5,1	80	58.82	24.01	16.64	0.52	17.62	0.46	1.36	9.66	2,5,5,1	81	57.48	24.09	15.19	0.60	17.95	0.45	1.34	9.48
Catechin	1,10,10,1	73	2.04	0.97	0.55	0.68	0.71	0.47	1.36	14.41	2 5 5 1	77	2.24	1.36	0.69	0.74	1.16	0.28	1.18	8.82
Epicatechin	2,10,5,1	74	19.81	7.43	4.72	0.60	6.53	0.24	1.14	8.33	1 10 5 1	75	20.09	8.06	5.82	0.48	7.04	0.23	1.14	7.72
Derivative I	2,10,5,1	73	2.28	2.37	0.79	0.89	1.12	0.78	2.12	9.03	2 5 5 1	70	2.12	2.21	0.56	0.94	0.82	0.86	2.67	12.33
Derivative II	2,10,5,1	70	1.81	2.32	0.61	0.93	0.82	0.88	2.85	17.98	1 10 5 1	69	1.89	2.56	0.69	0.93	0.89	0.88	2.88	16.56
Derivative III	1,5,5,1	67	1.53	2.66	1.05	0.84	1.13	0.82	2.35	13.50	2 5 5 1	67	1.51	2.65	0.63	0.94	1.07	0.84	2.47	14.26
Theobromine	2,10,5,1	74	38.77	17.77	7.12	0.84	9.61	0.71	1.85	10.40	2 5 5 1	76	39.06	18.61	7.14	0.85	10.36	0.69	1.80	9.65
Antioxidant activity (EC_{50})	2,10,5,1	78	6.37	3.04	1.29	0.82	2.10	0.52	1.45	8.02	1,5,5,1	84	6.74	3.66	2.82	0.41	2.93	0.35	1.25	5.75

N: Number of samples used for calibration.
Mean: Mean of the calibration series.
SD: Standard deviation.
SEC: Standard error of calibration (in actual values).
R^2: Determination coefficient for calibration.
SECV: Standard error of cross validation (in actual values).
r^2: Determination coefficient for cross-validation.
RPD: Ratio SD/SECV.
RER: Ratio (max-min)/SECV.

The fermentation of the beans increases the solubilisation of minor compounds in the lipid phase. During this process there is a significant increase in temperature, a loss of moisture and a rupture of the cells. In the case of phenols, their polymerisation and binding with other components such as proteins is also favoured, which makes them more lipid-soluble (Bonvehi & Coll, 1997). These factors encourage the minority components such as phenols and theobromine to pass into the oil in greater quantities as they are in greater contact and are more lipid-soluble. This fact would explain the good correlation between the contents of phenols and theobromine and the amount of fat, since fermentation increases the solubilisation of these compounds in the lipid phase.

Antioxidant activity was best predicted in cotyledons and with the spectra from cocoa beans ($r^2 = 0.61$ and SECV = 2.34) (Table 4). Poorer results were found for intact beans ($r^2 = 0.41$ and SECV = 3.43), nibs ($r^2 = 0.52$ and SECV = 2.10) and in particular ground cotyledons ($r^2 = 0.35$ and SECV = 2.93) (Tables 4 and 5). No clear pattern for the prediction of this constituent could be found, probably due to the unspecific chemical basis underlying this constituent. However, NIR has been used successfully to determine antioxidant activity in different type of products, including, among others, blueberry and grapes, olive oil, red wine, coffee and tea (Cozzolino, 2015).

As previously mentioned, most published studies have focused on model calibrations developed from ground cocoa cotyledon and very few studies have been developed with samples of intact beans. The results of this work show that, in general, the models obtained to estimate cotyledon composition from the spectra of cocoa nibs and ground beans were better than those obtained from the spectra of intact beans. This is probably due to the similarity of the material sent to laboratory for chemical analysis and that used for spectral acquisition (cotyledons) and to the behavior of the husk, which acts as a wall and reflects the incident light, shadowing the inner constituents. Nevertheless, in this study, the best r^2 values for total phenols and theobromine were obtained from the spectra of intact beans, both in cotyledons and beans, confirming that this non-destructive analytical method predicts these parameters with confidence while saving time and reducing the costs of laboratory analysis. The calibrations obtained from intact beans (Table 4) would also serve to classify cocoa bean samples according to their total composition (cotyledon + husk) and the composition of the fraction used by industry (cotyledon). The use of intact cocoa beans for the determination of quality parameters including their content in bioactive compounds would suppose an important advance in the speed of the analytical response, avoiding the problematic step of milling products with a high fat content and the time required for the separation of the husk, which is usually carried out by hand in the laboratory. The implementation of rapid screening techniques at germplasm banks and breeding programs would allow for the identification of interesting genotypes with high contents in functional components for their use in the cocoa industry, something that is perceived as beneficial by a significant number of consumers of cocoa products (Badrie et al., 2015; Subhashini et al., 2010).

4. Conclusions

NIRS technique could be used in germplasm banks and breeding programs for the rapid identification of interesting genotypes based on their contents in the most important bioactive compounds besides routine analyses for fat content. The results obtained from the current study are especially valuable as they include a high variability of cocoa beans in terms of genetic and weather conditions over three years.

NIR analyses employing the spectra of intact beans offers a faster and cheaper alternative over standard analytical methods for predictions based on calibration statistics of bioactive and standard constituents. Results of this work obtained with these spectra show that coarse quantitative predictions are possible for the contents of husk, derivatives of epicatechin I and II, and theobromine (RPD values between 2.0 and 2.5 and RER values ≥ 10); whereas for the contents of fat, proteins, total sugars, total phenols, and derivative III, screening purposes are possible as the models discriminate between low and high values (RPD values between 1.5 and 2.0 and RER ≥ 5). Nevertheless, experiments using a larger number of samples would be recommended in the future in order to improve the accuracy of these predictions.

In addition, the application of NIR analyses could help to improve the fermentation process by providing information on how changes in variables like time or temperature can increase the antioxidant activity of cotyledon.

Funding

This research was supported by the Spanish Ministry of Economy and Competitiveness and co-funded by the European Social Fund (ESF) (project AGL2016-79088R), the Spanish Ministry of Economy and Competitiveness Ramon y Cajal Programme (RyC2012-10456), the National Institute of Forestry, Agricultural and Livestock Research (INIFAP-México) and the Mexican National Council of Science and Technology (CONACyT-Mexico).

CRediT authorship contribution statement

Carolina Hernández-Hernández: Methodology, Formal analysis, Resources, Writing – original draft. **Víctor M. Fernández-Cabanás:** Validation, Formal analysis, Resources, Writing – review & editing. **Guillermo Rodríguez-Gutiérrez:** Investigation, Resources, Writing – review & editing. **África Fernández-Prior:** Investigation. **Ana Morales-Sillero:** Conceptualization, Investigation, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Carmen Reginfo for her technical support in lab analyses.

References

Afoakwa, O. E., Quao, J., Takrama, J., Budu, A., & Saalia, F. (2013). Chemical composition and physical quality characteristics of Ghanaian cocoa beans as affected by pulp pre-conditioning and fermentation. *Journal of Food Science & Technology, 50*, 1097–1105. https://doi.org/10.1007/s13197-011-0446-5

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (−)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *Journal of Near Infrared Spectroscopy, 20*, 307–315. https://doi.org/10.1255/jnirs.990

Azam, S., Hadi, N., Khan, N. U., & Hadi, S. M. (2003). Antioxidant and prooxidant properties of caffeine, theobromine and xanthine. *Medical Science Monitor, 9*, BR325–R330.

Barbin, D. F., Maciel, L. F., Bazoni, C. H. V., Ribeiro, M. D. S., Carvalho, R. D. S., Bispo, E. D. S., Miranda, M. D. P. S., & Hirooka, E. Y. (2018). Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *Journal of Food Science & Technology, 55*, 2457–2466. https://doi.org/10.1007/s13197-018-3163-5

Barnes, R., Dhanoa, M., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy, 43*, 772–777. https://doi.org/10.1366/0003702894202201

Bermúdez-Oria, A., Rodríguez-Gutiérrez, G., Alaiz, M., Vioque, J., Girón-Calle, J., & Fernández-Bolaños, J. (2019). Polyphenols associated to pectic polysaccharides account for most of the antiproliferative and antioxidant activities in olive extracts. *Journal of Functional Food, 62*, Article 103230. https://doi.org/10.1016/j.jff.2019.103530

Biscuit, Cake, Chocolate and Confectionery Alliance (BCCCA). (1996). *Cocoa beans. Chocolate manufacturers quality requirements* (4th ed., p. 10) London, UK.

Bonvehi, J. S., & Coll, F. V. (1997). Evaluation of bitterness and astringency of polyphenolic compounds in cocoa powder. *Food Chemistry, 60*, 365–370. https://doi.org/10.1016/S0308-8146(96)00353-6

Caligiani, A., Acquotti, D., Cirlini, M., & Palla, G. (2010). H NMR study of fermented cocoa (*Theobroma cacao* L.) beans. *Journal of Agricultural and Food Chemistry, 58*, 12105–12111. https://doi.org/10.1021/jf102985w

Caporaso, N., Whitworth, M. B., Fowler, M. S., & Fisk, I. D. (2018). Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry, 258*, 343–351. https://doi.org/10.1016/j.foodchem.2018.03.039

Carrillo, L. C., Londoño-Londoño, J., & Gil, A. (2014). Comparison of polyphenol, methylxanthines and antioxidant activity in Theobroma cacao beans from different cocoa-growing areas in Colombia. *Food Research International, 60*, 273–280. https://doi.org/10.1016/j.foodres.2013.06.019

Chaves-Lopez, C., Serio, A., Grande-Tovar, C. D., Cuervo-Mulet, R., Delgaso-Ospina, J., & Paparella, A. (2014). Traditional fermented foods and beverages from a microbiological and nutritional perspective: The Colombian heritage. *Comprehensive Reviews in Food Science and Food Safety, 13*, 1031–1048. https://doi.org/10.1111/1541-4337.12098

Cozzolino, D. (2015). Infrared spectroscopy as a versatile analytical tool for the quantitative determination of antioxidants in agricultural products, foods and plants. *Antioxidants, 4*, 482–497. https://doi.org/10.3390/antiox4030482

Davies, A. M. C., Franklin, J. G., Grant, A., Griffiths, N. M., Shepherd, R., & Fenwick, G. R. (1991). Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans. *Vibrational Spectroscopy, 2*, 161–172.

FAOSTAT. (2020). http://www.fao.org/faostat/en/#data/QC.

Hashimoto, J. C., Lima, J. C., Celeghini, R. M. S., Nogueira, A. B., Efraim, P., Poppi, R. J., & Pallone, J. A. L. (2018). Quality control of commercial cocoa beans (*Theobroma cacao* L.) by near-infrared spectroscopy. *Food Anal. Methods, 11*, 1510–1517. https://doi.org/10.1007/s12161-017-1137-2

Hernández-Hernández, C., Fernández-Cabanás, V. M., Rodríguez-Gutiérrez, G., Bermúdez Oria, A., & Morales-Sillero, A. (2021). Viability of near infrared spectroscopy for the fast analysis of bioactive compounds in intact cocoa bean husk. *Food Control, 120*, Article 107526. https://doi.org/10.1016/j.foodcont.2020.107526

Hernández-Hernández, C., Morales-Sillero, A., Fernández-Bolaños, J., Bermúdez Oria, A., Azpeitia-Morales, A., & Rodríguez-Gutiérrez, G. (2019). Cocoa bean husk: Industrial source of antioxidant phenolic extract. *Journal of the Science of Food and Agriculture, 99*, 325–333. https://doi.org/10.1002/jsfa.9191

Hernández-Hernández, C., Viera-Alcaide, I., Morales-Sillero, A. M., Fernández-Bolaños, J., & Rodríguez-Gutiérrez, G. (2018). Bioactive compounds in Mexican genotypes of cocoa cotyledon and husk. *Food Chemistry, 240*, 831–839. https://doi.org/10.1016/j.foodchem.2017.08.018

International Organization for Standardization (ISO). (2016). *ISO 16634-2:2016: Food products - determination of the total nitrogen content by combustion according to the Dumas principle and calculation of the crude protein content - Part 2: Cereals, pulses and milled cereal products*.

Kim, H., & Keeney, P. G. (1984). (-)-Epicatechin content in fermented and unfermented cocoa beans. *Journal of Food Science, 49*, 1090–1092. https://doi.org/10.1111/j.1365-2621.1984.tb10400.x

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat - determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food Chemistry, 181*, 152–159. https://doi.org/10.1016/j.foodchem.2015.02.084

Kutsanedzie, F. Y. H., Chen, Q., Sun, H., & Cheng, & W. (2017). In situ cocoa beans quality grading by near-infrared-chemodyes systems. *Analytical Methods, 9*, 5455–5463. https://doi.org/10.1039/C7AY01751K

Liendo, R., Padilla, F. C., & Quintana, A. (1997). Characterization of cocoa butter extracted from Criollo cultivars of *Theobroma cacao* L. *Food Research International, 9*, 727–731. https://doi.org/10.1016/S0963-9969(98)00025-8

Marten, G., Shenk, J., & Barton, F. (1989). *Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. U.S. Department of agriculture, agriculture handbook* (Vol. 643, pp. 1–110).

Morales-Sillero, A., Fernández-Cabanás, V. M., Casanova, L., Jiménez, M. R., Suárez, M. P., & Rallo, P. (2011). Feasibility of NIR spectroscopy for non-destructive characterization of table olive traits. *Journal of Food Engineering, 107*, 99–106. https://doi.org/10.1016/j.jfoodeng.2011.05.039

Mrabet, A., Jiménez-Araujo, A., Fernández-Bolaños, J., Rubio-Senent, F., Lama-Muñoz, A., Rodríguez-Gutiérrez, G., & Sindic, M. (2016). Antioxidant phenolic extracts obtained from secondary Tunisian date varieties (*Phoenix dactylifera* L.) by hydrothermal treatments. *Food Chemistry, 196*, 917–924. https://doi.org/10.1016/j.foodchem.2015.10.026

Oracz, J., Zyzelewicz, D., & Nebesny, E. (2015). The content of polyphenolic compounds in cocoa beans (*Theobroma cacao* L.), depending on variety, growing region, and processing operations: A review. *Critical Reviews in Food Science and Nutrition, 55*, 1176–1192. https://doi.org/10.1080/10408398.2012.686934

Peláez, P., Bardón, I., & Camasca, P. (2016). Methylxanthine and catechin content of fresh and fermented cocoa beans, dried cocoa beans, and cocoa liquor. *Science Agropecuaria, 7*, 355–365.

Pissard, A., Baeten, V., Dardenne, P., Dupont, P., & Lateur, M. (2018). Use of NIR spectroscopy on fresh apples to determine the phenolic compounds and dry matter content in peel and flesh. *Biotechnology, Agronomy, Society and Environment, 22*, 3–12. https://doi.org/10.25518/1780-4507.16241

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Estece, E., Talens, P., & Barat, J. M. (2020). Roadmap of cocoa quality and authenticity control in the industry: A review of conventional and alternative methods. *Comprehensive Reviews in Food Science and Food Safety, 19*, 448–478. https://doi.org/10.1111/1541-4337.12522

Shenk, J. S., & Westerhaus, M. O. (1996). Calibration the ISI way. In A. M. C. Davies, & P. C. Williams (Eds.), *Near infrared spectroscopy: The future waves* (pp. 198–202). Chichester, UK: NIR Publications.

Singleton, V. L., & Rossi, J. A., Jr. (1965). Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagent. *American Journal of Enology and Viticulture, 16*, 144–158.

Subhashini, R., Rao, U. S. M., Sumathi, P., & Gunalan, G. (2010). A comparative phytochemical analysis of cocoa and green tea. *Indian Journal Science Technology, 3*, 188–192. https://doi.org/10.17485/ijst/2010/v3i2.3

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Non-destructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture, 124*, 234–242. https://doi.org/10.1016/j.compag.2016.04.012

Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry, 176*, 403–410. https://doi.org/10.1016/j.foodchem.2014.12.042

Teye, E., Huang, X., Takrama, J., & Haiyang, G. (2014). Integrating NIR spectroscopy and electronic tongue together with chemometric analysis for accurate classification of cocoa bean varieties. *Journal of Food Process Engineering, 37*, 560–567. https://doi.org/10.1111/jfpe.12109

Vera, C. J., Vallejo, T. C., Párraga, M. D., Morales, R. W., Macías, V. J., & Ramos, R. R. (2014). Physical-chemical and sensory attributes of the cocoa nacional (*Theobroma cacao* L.) fifteen clone beans in Ecuador. *Revista Ciencia y Tecnología, 7*, 21–34.

Williams, P. C., & Sobering, D. C. (1996). How do we do it: A brief summary of the methods we use in developing near infrared calibration. In A. M. C. Davis, & P. C. Williams (Eds.), *Near infrared spectroscopy: The future waves* (pp. 185–188). Chichester: NIR Publications.

Witham, F. H., Blaydes, D. F., & Devlin, R. M. (1971). *Experiment in plant physiology*. New York, NY: Van Nostrand Reinhold Co.