JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2010, 58, 7811–7819 7811
DOI:10.1021/jf100409v

Near Infrared Spectroscopy for High-Throughput Characterization of Shea Tree (*Vitellaria paradoxa*) Nut Fat Profiles

Fabrice Davrieux,*,† François Allal,‡ Georges Piombo,§ Bokary Kelly,∥ John B. Okulo,⊥ Massamba Thiam,# Ousmane B. Diallo,▽ and Jean-Marc Bouvet‡

†CIRAD, Performance of Tropical Production and Processing Systems Department, UMR QUALISUD, TA B-95/16, 73, avenue Jean François Breton, 34398 Montpellier Cedex 5, France, ‡CIRAD, Biological Systems Department, Research Unit 39, "Genetic Diversity and Breeding of Forest Tree Species", Campus international de Baillarguet TA A-39/C, 34398 Montpellier Cedex 5, France, §CIRAD, Performance of Tropical Production and Processing Systems Department, UMR IATE, TA B 62/16, 73 rue Jean François Breton, 34398 Montpellier Cedex 5, France, ∥Institut d'Economie Rurale, PRF/CRRA-SIKASSO BP 178/16, Sikasso, Mali, ⊥Department of Forest Biology & Ecosystems Management, Faculty of Forestry and Nature Conservation, Makerere University, PO Box 7062 Kampala, Uganda, #Institut Sénégalais de la Recherche Agronomique, Centre National de la Recherche Forestière, BP 2312 Dakar-Hann, Senegal, and ▽INERA-Centre National de la Recherche Scientifique et Technologique, 03 BP 7041, Ouagadougou, Burkina Faso

The Shea tree (*Vitellaria paradoxa*) is a major tree species in African agroforestry systems. Butter extracted from its nuts offers an opportunity for sustainable development in Sudanian countries and an attractive potential for the food and cosmetics industries. The purpose of this study was to develop near-infrared spectroscopy (NIRS) calibrations to characterize Shea nut fat profiles. Powders prepared from nuts collected from 624 trees in five African countries (Senegal, Mali, Burkina Faso, Ghana and Uganda) were analyzed for moisture content, fat content using solvent extraction, and fatty acid profiles using gas chromatography. Results confirmed the differences between East and West African Shea nut fat composition: eastern nuts had significantly higher fat and oleic acid contents. Near infrared reflectance spectra were recorded for each sample. Ten percent of the samples were randomly selected for validation and the remaining samples used for calibration. For each constituent, calibration equations were developed using modified partial least squares (MPLS) regression. The equation performances were evaluated using the ratio performance to deviation (RPD_p) and R_p^2 parameters, obtained by comparison of the validation set NIR predictions and corresponding laboratory values. Moisture (RPD_p = 4.45; R_p^2 = 0.95) and fat (RPD_p = 5.6; R_p^2 = 0.97) calibrations enabled accurate determination of these traits. NIR models for stearic (RPD_p = 6.26; R_p^2 = 0.98) and oleic (RPD_p = 7.91; R_p^2 = 0.99) acids were highly efficient and enabled sharp characterization of these two major Shea butter fatty acids. This study demonstrated the ability of near-infrared spectroscopy for high-throughput phenotyping of Shea nuts.

KEYWORDS: *Vitellaria paradoxa*; NIRS; fatty acids; high-throughput characterization; shea butter; MPLS regression

INTRODUCTION

The Shea tree (*Vitellaria paradoxa* C. F. Gaertn, Sapotaceae family), called karité in French, is a major tree species of African agroforestry systems. The native range of this long-lived (over 200 years) savannah tree species is a large belt 6000 km long from East (Senegal) to West (Uganda) and 600 km wide from north of the equator to south of the Sahara. *Vitellaria paradoxa* is divided into 2 subspecies: subsp. *paradoxa* is distributed throughout most of the range (including West and Central Africa) while subsp. *nilotica* is found in the eastern part of Sub-Saharan Africa. Within that wide distribution, there is an immense ecological diversity, with altitudes ranging from almost 0 to nearly 1500 m, a mean annual temperature of between 22 and 30 °C and a mean annual rainfall of 400 mm to 1400 mm. Shea trees reach a maximum height of 20 m with a trunk diameter ranging generally from 0.3 to 1 m at breast height. Annual flowering and fruiting occurs from December to May with local variations (*1*).

The Shea tree has been exploited by African communities for about 3000 years and offers an opportunity for sustainable development in Sudanian countries (*2*, *3*). Indeed, Shea butter extracted from fruit kernels provides an attractive potential for both the food and cosmetics industries (*1*). The estimated yield of dry kernel is about 600,000 tons per year, and exports have

*Corresponding author. Tel: +33 467 615 893. Fax: +33 467 614 433. E-mail: davrieux@cirad.fr.

Published on Web 06/02/2010

increased over the past decade, reaching 350,000 tons today, mainly to the USA and Europe. Traditional uses for medicine and foods account for 50%, refining for specialty fats 45%, and cosmetics (refined or not) 5% of production (Lovett personal communication).

In order to improve the sustainable management of this species, to launch prebreeding programs and precisely assess the potential of different geographical origins for marketing, it is now becoming crucial to improve our knowledge of the diversity of the chemical content of nuts.

Initial evaluations of Shea nut chemical composition throughout the natural range of *Vitellaria paradoxa* have shown great diversity particularly for fat content and fatty acid composition (*4*,*5*). Those studies revealed fat content values ranging from 22.3% to 52.8%. Relative fatty acid compositions were found to vary, particularly for oleic acid (C18:1), ranging from 37.1% to 62.1%, and stearic acid (C18:0), from 25.6% to 50.2%. It was suggested that the variation observed in these initial investigations was due to the two subspecies, but also to bioclimatic variations related to temperature and annual rainfall.

However, these first evaluations of Shea nut chemical diversity were based on an unbalanced and restricted collection of material. Consequently, the results may not reflect a statistically accurate assessment of actual *Vitellaria paradoxa* variability within its natural range. Indeed, some geographical origins studied were represented by 29 samples, while others by only one.

Fresh assessment of chemical content variability, based on significant and representative sampling, is therefore needed. Classical chemical tools such as solvent extraction and gas chromatography (GC) (*6*, *7*) have shown their efficiency in determining the chemical composition of plant material, but they are time-consuming and expensive. In this context, there is now a definite need to develop a rapid and accurate method.

Near-infrared spectroscopy (NIRS) is a very efficient method for high-throughput screening of plant materials for their chemical characteristics. This indirect method is based on vibrational properties of organic molecule chemical bonds and their interactions with infrared radiation. The NIR absorption spectrum is therefore correlated with a sample's chemical composition (*8*). The NIRS method can therefore be used to predict the value of a chemical trait in a sample, but requires the development of an accurate and robust calibration equation. To that end, this method requires prior characterization of a large number of samples uniformly covering the range of variability of the studied trait (*9*). Compared to conventional time-consuming biochemical methods, NIRS has already proven its usefulness for estimating various parameters in diverse natural and agricultural products (*10*). Indeed, several studies have demonstrated NIRS efficiency in characterizing fat content and fat profiles in fruit from various species such as peanut, hazelnut and sesame (*11*−*14*).

In this paper, we demonstrate the efficiency of near-infrared spectroscopy in characterizing fat profiles for nuts from Shea trees of different geographical origin and we propose calibration equations. This technique provides producing countries with an opportunity for the high-throughput characterization of large numbers of Shea nut samples. Moreover, use of the NIRS technique will be helpful in phenotype/genotype association studies and breeding programs.

MATERIALS AND METHODS

Materials. The sampling strategy was designed to ensure maximum coverage of the range of variation in fat composition. As suggested by previous studies (*4*,*5*), variation was assumed to be linked to the difference between subspp. *paradoxa* and *nilotica*, and to environmental gradients. Consequently, within each subspecies area, samples were collected according to north−south clines in order to increase environmental variations (rainfall and temperature decreases from North to South).

Table 1. Description of the 17 Harvested Sites

				N^a	
site index	site	GPS	altitude (m)	2007	2008
		Senegal			
1	Kenioto	12.57N; 12.16W	161		12
2	Samecouta	12.60N; 12.13W	126		18
3	Saraya	12.83N; 11.75W	180		10
		Mali			
4	Nafégué	10.51N; 5.98W	344	35	40
5	Mperesso	12.28N; 5.33W	340	34	37
6	Daelan	13.25N; 4.99W	282	34	37
7	Tori	13.61N; 3.72W	377		35
8	Sassambourou	14.31N; 3.51W	392		35
		Burkina			
9	Titao	13.72N; 2.16W	336		18
10	Guibare	13.07N; 1.61W	303		21
		Ghana			
11	Kawampe	8.43N; 1.56W	125		35
12	Tolon	9.43N; 1.00W	154		35
13	Kulbia	10.83N; 0.96W	206		34
		Uganda			
14	Katakwi	1.82N; 33.99E	1100	34	29
15	Pader	2.80N; 33.31E	1031	31	25
16	Moyo	3.62N; 31.64E	863		16
17	Uleppi-Arua	3.02N; 30.90E	1200		19

[a] *N*: number of sampled trees. Total: 624.

Samples were collected under uniform conditions over two years (2007 and 2008) in four West African countries, Senegal, Mali, Burkina Faso and Ghana for subsp. *Paradoxa*, and one East African country, Uganda, for subsp. *nilotica*. Within each country, different sites were sampled based on a rainfall and temperature gradient. A total of 624 trees (GPS located) were sampled at 17 sites: Senegal (Kenioto, Samecouta and Saraya), Mali (Nafégué, Mperesso, Daelan, Tori and Sassambourou), Burkina Faso (Titao and Guibare), Ghana (Kawampe, Tolon and Kulbia) and Uganda (Katakwi, Pader, Moyo and Uleppi-Arua). A sampling description is given in **Table 1**, and the distribution of sites is illustrated in **Figure 1**. On average, 30 mature fruits (ready to fall to the ground) were collected per tree and pooled. Postharvest treatment, including depulping and drying (3 days at 60 °C), was carried out on site. The dried nuts were sent to the CIRAD laboratory in Montpellier (France). On receipt, nuts were oven-dried for two days at 60 °C in order to stabilize moisture content, and then they were stored at room temperature prior to analysis.

Sample Preparation. Unshelled Shea nuts were first ground in a "Vorwerk Thermomix Robot". Raw powders were frozen at −20 °C and reground in a "SEB Valentin blender" in order to obtain a final particle size between 0.5 and 0.8 mm. The final powder samples were stored at −20 °C.

Laboratory Analyses. For each sample, moisture content (MC) was assessed by gravimetric analysis after drying at 103 °C in an oven (Chopin) for 16 h. Fat content (FC) was solvent-extracted (petroleum ether) from powders using a semiautomatic Soxtec 2050 extractor (FOSS Analytical, A/S Slangerupgade 69, DK-3400 Hillerød, Denmark) according to the manufacturers' instructions. In a preliminary study (data not shown), we compared our extraction procedure with conventional fat extraction method (Soxhlet with hexane as the solvent over 16 h) and observed no significant differences. After gravimetric quantification, extracted oils were stored at −20 °C for further chemical analyses. Fatty acid (FA) profiles were obtained according to the protocol described by Tchobo (*6*): after esterification of the oil using sodium methylate, FA profiles were determined by gas chromatography using a Thermo Focus (Thermo Fisher Scientific, 81 Wyman Street, Waltham, MA 02454) GC with a

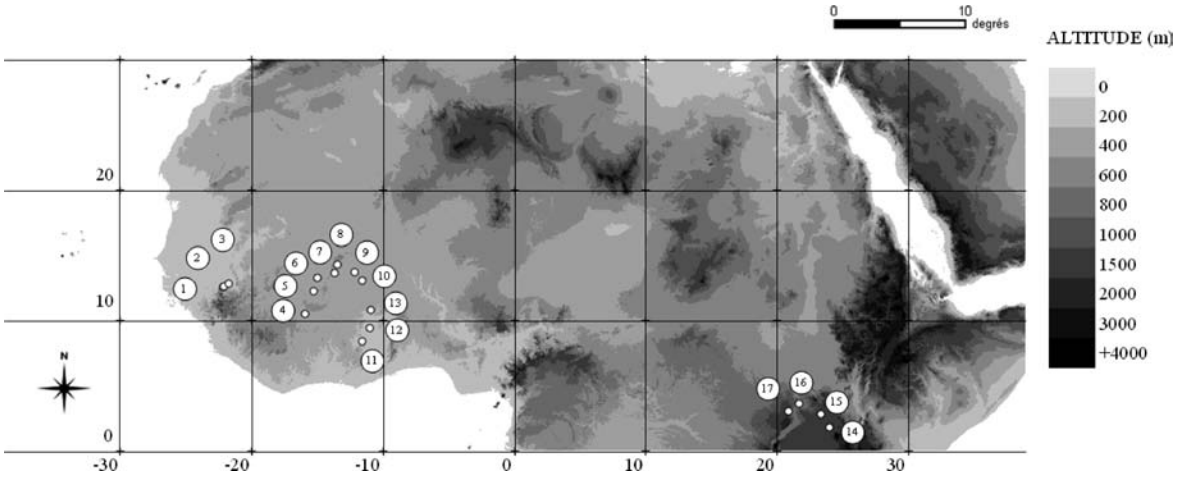

Figure 1. Location of the 17 harvested sites.

CP SIL 88 (highly substituted cyanopropyl phase) column (Varian, Inc., 3120 Hansen Way, Palo Alto, CA 94304-1030). The chromatographic conditions were as follows: the injector temperature was 250 °C and that of the FID detector 270 °C. The oven temperature settings were as follows: 150 °C to 225 at 5 °C/min, and then held at 225 °C for 2 min. Vector (He) gas flow was 1 mL/min with a splitting ratio of 1:80. The injection volume was 1 μL. For all laboratory measurements the standard error of the laboratory (SEL) was estimated as the standard deviation from 10 replicates of a standard Shea sample.

For each parameter, comparisons between sites and subspecies were carried out using one-way analyses of variance (ANOVA) with Student−Newman−Keuls SNK multiple pair comparison tests at the 5% level (XLstat software, Addinsoft, 40 rue Damrémont, 75018 Paris, France). Outlier samples were detected using the Grubbs test (*15*) procedure at the 5% level (Statgraphics Centurion XV, Statpoint Technologies, Inc., 560 Broadview Avenue, Suite 201, Warranton, VA 20186).

NIR Spectrum Acquisition. A NIRS 6500 monochromator (Foss NIRSystems, Silver Spring, MD) was used to scan reflectance from 400 to 2500 at 2 nm intervals, using ring cups (50 mm in diameter) with about 3 g of fine Shea nut powder. Data were saved as the average of 32 scans and stored as log(1/R), where R was the reflectance at each wavelength and 1 the reflectance of a standard ceramic reference. Spectra were acquired randomly, each sample being measured twice, and the average spectrum was stored. Statistical analyses were performed using Win-ISI II software (Infrasoft International, Port Matilda, PA), JUMP 7.01 (SAS Institute Inc., Cary, NC), Statgraphics Centurion XV (StatPoint Inc., Warrenton, VA) and XLstat software (Addinsoft, Paris, France).

Spectrum Pretreatment. Spectra were mathematically corrected for light scattering by using the standard normal variate and detrend correction (*16*, *17*). The second derivative was calculated on five data points and smoothed using Savitzky and Golay polynomial smoothing on five data points (*18*).

Principal Component Analysis. Prior to calibration development, a principal component analysis (PCA) was used to extract relevant information from the spectral matrix (n = 624). The generalized Mahalanobis distance (H) was calculated on the extracted PCs for each sample. This statistical distance is useful for defining boundaries of the population and a similarity index between spectra (*19*).

This enabled us to discard 22 outlier samples with a Mahalanobis distance $H > 3$. These samples had been tagged as moldy on arrival at the laboratory.

NIR Calibration Development. In order to assess the performance of the predictive equations, the remaining 602 PCA samples were split into a calibration subset (cal) and a validation subset (val). The validation set was created by randomly selecting 10% of the 602 samples including 2 samples from Burkina, 8 from Ghana, 27 from Mali, 1 from Senegal and 22 from Uganda. The calibration set comprised the remaining 542 samples.

Calibration equations for the parameters were constructed with the calibration subset (542 samples) using the modified partial least-squares regression (mPLS) (*20*) algorithm of WinISI software (*9*).

Calibration statistics included the following parameters: standard deviation (SD), coefficient of determination (R^2), standard error of calibration (SEC), and standard error of cross-validation (SECV). Cross-validation was used during calibration development in order to select the optimum number of latent variables and to minimize overfitting of the equations (*21*, *22*). For SECV estimation, 25% of the samples (randomly selected) were predicted using a calibration model developed with the other 75%. SECV estimation was repeated four times and the average calculated. In addition to R^2, the ratio of performance to deviation (RPD = SD/SECV) was used to evaluate the general quality of the fit obtained for each equation. Unlike SEC and SECV, RPD is independent of parameter units and can therefore be compared between parameters (*23*). The Student (t) test was used to identify t-outlier samples during calibration development. Outlier detection was based on the standardized residuals (= error/SECV) with a cutoff of 2.5. Two passes of outlier elimination were used (*22*).

The standard error of prediction (SEP), corresponding to the standard deviation of residuals, was estimated by predicting the validation subset using a model developed on the calibration subset. The ratio performance to deviation of prediction (RPD_p) was also calculated as $RPD_p = SD_{val}/SEP$ (where SD_{val} was the standard deviation of validation samples). The quality of the fits between wet chemistry values (moisture, fat and FA relative percentages) and NIRS-predicted values was evaluated from the linear regression slope, the R^2 and the bias.

RESULTS AND DISCUSSION

Chemical Analysis. The 602 samples were analyzed in the laboratory for their moisture and fat content, and 599 extracted butter fatty acid profiles were determined by GC (**Table 2**).

Moisture content ranged from 2.25% to 8.37%, with an average value of 4.48%. Data dispersion was rather small (SD = 0.89%). Low moisture content values avoided lipase activity and free fatty acid formation (*24*). Five samples presented extreme values (Grubbs test; level 5%) with a moisture content ranging between 7.49% and 8.37%. No particular defect was noticed on receipt of these samples. Their corresponding NIR spectra were not atypical (H values lower than 3), and the calibration process did not highlight these samples as outliers. This relatively high moisture content had no incidence on calibration development and was probably the consequence of incomplete drying.

In order to avoid bias in spectral pattern interpretation, an analysis of variance (ANOVA) was performed to ensure that sample origins had no effect on moisture content values. The ANOVA concluded that there was no significant difference (level α = 5%) between countries for average moisture content values.

Fat content was equal to 49.66% on average (dry matter basis) with a SD of 5.03% corresponding to relatively low dispersion

Table 2. Descriptive Statistics for Laboratory Results[a]

	MC	fat[b]	palmitic[c] (C16:0)	stearic[c] (C18:0)	oleic[c] (C18:1 n-9)	vaccenic[c] (C18:1 n-7)	linoleic[c] (C18:2 n-6)	linolenic[c] (C18:3 n-6)	arachidic[c] (C20:0)
N	602	602	599	599	599	565[d]	599	509[d]	599
min	2.25	26.96	2.71	23.98	38.76	0.08	3.22	0.06	0.46
max	8.37	59.66	8.44	49.30	62.34	0.90	13.00	0.69	1.83
av	4.48	49.66	4.29	38.13	48.58	0.41	7.13	0.31	1.20
SD	0.89	5.03	0.72	5.89	5.46	0.16	1.04	0.13	0.20

[a] MC: moisture content. *N*: number of samples. SD: standard deviation. [b] Fat expressed as a % of dry matter. [c] Relative fatty acid composition (FA %). [d] GC-detected values.

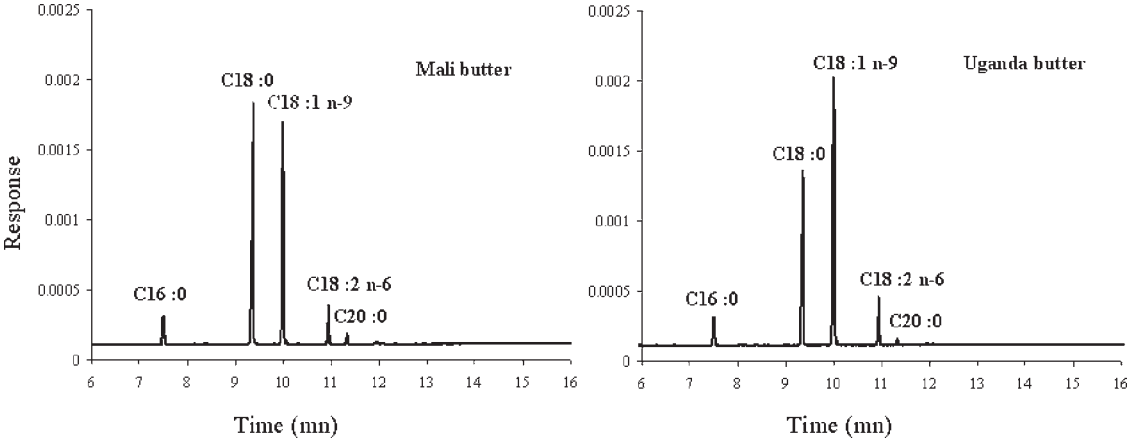

Figure 2. Fatty acid GC chromatograms of two samples from Mali and Uganda.

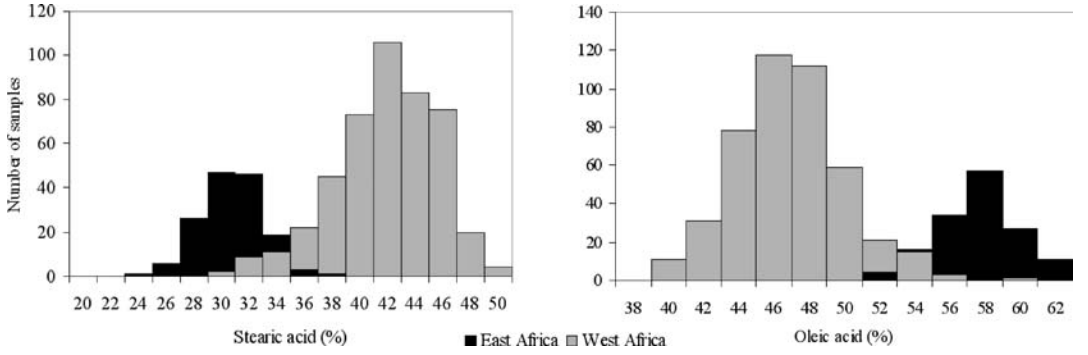

Figure 3. Histograms of relative stearic and oleic acid percentages.

(CV = 10.1%). The values ranged from 29.96% to 59.66%. Five samples were found to be outliers (Grubbs test; 5% level) with a fat content under 34.33%. These samples were suspected of containing immature fruits, and this hypothesis was consistent with the difficulty of distinguishing between mature and immature Shea fruits during harvesting. The average fat content for East African Shea nuts (52.92%) was significantly higher (ANOVA-SNK, at 5% level) than that of West African Shea nuts (48.03%). Senegalese fruits had the lowest fat content (45.18%). Fruits from Burkina Faso (48.54%) were similar to those from Mali (48.53%) and Ghana (49.88%). No significant differences were reported between sites in each country.

This result suggests that variations linked to environmental gradients did not lead to specific fat accumulation. Observed fat contents and differences between East and West African fruits were in accordance with the results reported by Maranz (*4*).

Based on an examination of gas chromatography profiles (**Figure 2**), seven FA (**Table 2**) with relative percentages over 0.05% were adopted for the study, corresponding to three saturated FA (palmitic C16:0, stearic C18:0 and arachidic C20:0), two *cis*-monoenoic FA (oleic C18:1 n-9 and *cis*-vaccenic C18:1 n-7) and two polyenoic FA (linoleic C18:2 n-6 and γ-linolenic C18:3 n-6).

Fatty acid composition mainly consisted of stearic (overall average value: 38.13%) and oleic (overall average value: 48.58%) acids. These results, obtained on a large set of samples representative of the natural range of Shea trees, confirmed previously published results (*5*, *25*).

The gas chromatography profiles from two samples are given in **Figure 2**, one from Mali (West Africa) and one from Uganda (East Africa). These chromatograms confirm the predominance of stearic and oleic acids in FA profiles. In addition, this figure shows a marked dominance of oleic acid in Ugandan butter compared to Malian butter. Indeed, histograms (**Figure 3**) confirmed that East African Shea butters were richer in oleic acid (56.64%) than West African butters (45.91%). Conversely, the relative stearic acid percentage was higher in West Africa (40.91%) than in East Africa (29.72%). Based on the relative proportions of oleic and stearic acid, two groups corresponding to East and West Africa were defined (**Figure 3**).

Within the West African region, average values for oleic acid decreased from Mali (46.54%) to Ghana (44.33%) while stearic acid increased from Mali (40.43%) to Ghana (42.11%).

Palmitic acid was present at an average of 4.29%. The highest values were observed for Senegalese butters (5.61%) and lowest for Malian butters (3.94%). *cis*-Vaccenic and linolenic acids were

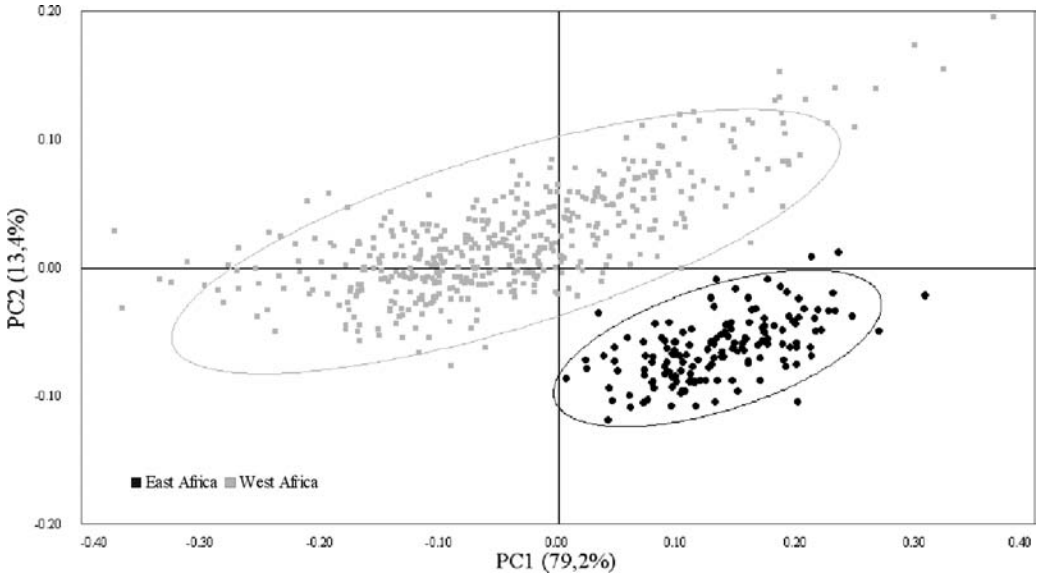

Figure 4. Scatter plot of the 602 sample scores for the first two principal components (95% confidence ellipses).

low (0.41% and 0.31%) and detected in 565 and 509 samples respectively. Linoleic acid was detected in all butters, ranging from 3.22% to 13.0%, with an average of 7.13%. The relative content of linoleic acid was fairly similar for the five countries, with a minimum value (6.59%) for Senegalese butters and a maximum (7.25%) for Malian butters. Arachidic acid ranged from 0.46% to 1.83%, with an average of 1.20%. Arachidic acid was found to be low in East Africa (0.98% in Uganda) and higher in West African countries (1.16% in Senegal, 1.17% in Ghana, 1.26% in Burkina and 1.33% in Mali).

The average ratio of saturated fatty acids (SFA) to unsaturated (USFA) was 0.79. Ugandan Shea butter showed the lowest ratio (0.55) while for West African butters the ratio ranged from 0.85 (Mali) to 0.92 (Burkina).

cis-Vaccenic, linoleic and arachidic acids followed normal distributions (figures not shown). Stearic and oleic acid distributions were bimodal due to West and East African samples (**Figure 3**). Linolenic acid distribution was also bimodal with Malian butters apart from the others.

According to the ANOVA-SNK test (level 5%), significant differences were found for stearic and oleic acid content between Uganda (subsp. *nilotica*) and West Africa (subsp. *paradoxa*). Within West Africa, for those FAs, Mali and Ghana were found to be significantly different. Palmitic acid content was similar for Malian and Burkinese butters, and different for other countries. *cis*-Vaccenic acid content was significantly different between Uganda and Mali. Linoleic acid content was similar for Senegal and Burkina, and different for the other countries. Linolenic acid content was similar for Ugandan and Senegalese butters, and for Ghanaian and Burkinese butters, while Malian butters were different from the others. Finally, only butters from Ghana and Senegal displayed a similar relative arachidic acid content.

Spectral Data Matrix: Principal Component Analysis. The first three principal components (PCs) extracted from a PCA carried out on the 602 samples explained 79.2%, 13.2%, and 3.5% of total inertia respectively. Only three samples presented H distances, calculated on 32 PCs, over 3: one from Mali ($H = 3.1$), one from Uganda ($H = 3.2$) and one from Senegal ($H = 3.5$). In terms of fat and moisture content these samples were in the normal range, and no special information was registered for them, therefore based on their relatively low H values they were kept in the database for the rest of the study.

A scatter plot of the first two PCs (**Figure 4**) showed a sharp discrimination between West and East African Shea nuts. Based on previous results (*4*), the observed discrimination was suspected of being due to differences in butter FA profiles between West and East African Shea nuts. This hypothesis was confirmed through the significant correlations (Pearson test, 5% level) calculated between PCs and fatty acids (e.g., correlation between PC1 and stearic acid $r = -0.83$; PC1 and oleic $r = 0.79$). Additionally, we found a significant correlation between PC2 and fat content ($r = -0.86$).

The scatter plot of eigenvector coefficients (loadings) versus wavelength confirmed the origin of the discrimination (**Figure 5**). Indeed, the highest coefficients were associated with the wavelength corresponding to absorption bands of chemical bonds specific to fat, such as the C−H from $-CH_2$ group second overtone stretching band (1214 nm), the $-CH_2$ stretch combination band (1396 nm), the C−H ($-CH_2$ group) first overtone stretching band (1724 nm) and the characteristic −C−H ($-CH_2$ group) stretching and deformation combination band (2308 nm) (*10*, *26*). The profiles for loadings 1 and 2 were similar for major absorption bands, the absorption band due to H−OH second overtone (1916 nm) being taken into account by PC2. Assignment of the major Shea nut NIR absorption bands is summarized in **Table 3**. Moreover, the loading profiles were close to the pure butter spectrum, which accounts for a large share of the Shea nut spectral fingerprint. This was illustrated by comparing the Ugandan Shea nut spectrum and the corresponding extracted butter spectrum (**Figure 6**).

NIRS Calibration. Of the 602 spectra, the 542 samples used for calibration and the 60 randomly selected samples for validation were representative of the variation, in terms of SD and range (**Table 4**) for each constituent. Fat and stearic acid distributions for validation and calibration sets are shown in **Figure 7**. These histograms illustrate the distribution of validation samples with respect to the variation of the population. One validation sample had an extreme value (26.96%) for fat content (lower than the lowest calibration sample whose fat content was 28.44%); keeping this sample in the validation set allowed us to test our fat model in extrapolation. Modified PLS regressions (mPLS) were performed using the calibration set and reference laboratory data. Constituents were then predicted for the validation set using the mPLS models.

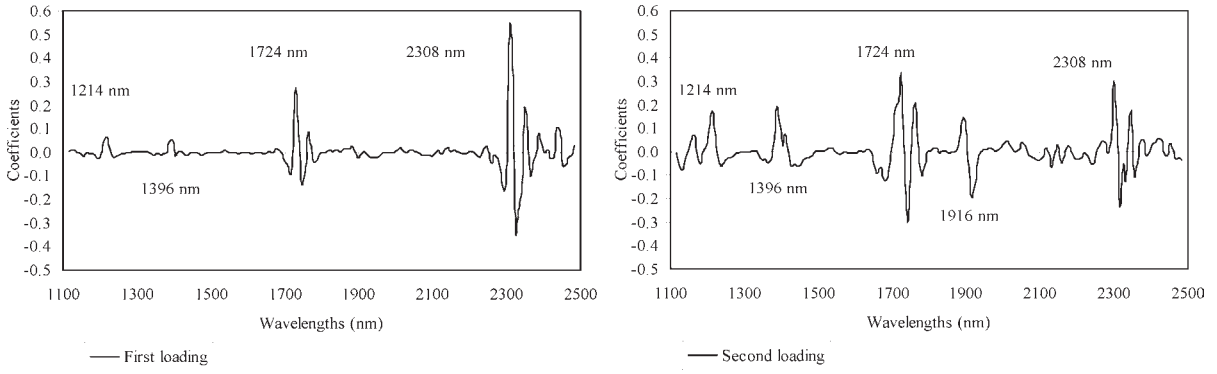

Figure 5. The first two loadings as a function of wavelength.

Table 3. Assignment of Major Shea Nut NIR Absorption Bands

wavelength (nm)	bond type	vibration mode	structure
1160	C−H	stretching 2nd overtone	CH_3
1172	C−H	stretching 2nd overtone	HC=CH
1214	C−H	stretching 2nd overtone	$-CH_2$
1396	$-CH_2$	stretching	$-CH_2$
1450	O−H	stretching 1st overtone	H_2O
1724	C−H	stretching 1st overtone	$-CH_2$
1916	O−H	stretching and deformation combination	H_2O
2100−2200	C=C	stretching and deformation combination	HC=CH
	cis =C−H and C=C	stretching combination	HC=CH
2308	C−H	stretching and deformation combination	$-CH_2$

Model prediction accuracy was evaluated through RPD_p: values below 1.4 corresponded to unusable calibration, RPD_p values between 1.4 and 1.7 corresponded to a calibration usable for rough screening, and values higher than 4 indicated that the calibration could be used for quality control analyses (*10*, *27*). Calibration statistics and validation statistics are reported in **Table 4**.

Moisture Content. Calibration for MC gave both R^2 and R_p^2 of 0.95. For this model, RPD_p was 4.45 and SEP was 0.23% with a regression slope of 0.96 between laboratory and NIR-predicted values (**Figure 8a**). Compared to calibrations obtained in the literature for MC (*28*), this model appeared less efficient. However, this was due to sample preparation (nuts were dried), which led to dissymmetric distribution of MC values: 83% of samples contained less than 4% moisture. In any event, based on model performances, MC NIR calibration can be applied to Shea nuts for routine analysis. To prevent lipid degradation due to hydrolysis of triacylglycerol, commercial nuts must have a moisture content below 9% (*29*), thus our model was efficient enough for its control.

Fat Content. The model developed for fat quantification was efficient: R^2 and R_p^2 were equal to 0.96 and 0.97, and RPD_p to 5.61. The regression slope between laboratory and NIR-predicted values was 1.02 (**Figure 8b**). The estimated SEP was 1.05%, enabling Shea nut fat content determination with an accuracy of ±2.06%. The model obtained was in accordance with results in the literature related to high fat content products (*30*). Our calibration was usable for quality control of Shea nut fat content in producing countries for commercial purposes.

cis-Vaccenic Acid. The calibration developed for *cis*-vaccenic acid was poor with $R^2 = 0.48$. This result was confirmed by the estimation of the R_p^2 on the validation set ($R_p^2 = 0.23$). The RPD_p observed for the *cis*-vaccenic acid model was about 1.2, corresponding to an unusable calibration. This was probably due to the low *cis*-vaccenic acid contents observed (<0.9%) which were close to the NIRS detection limit.

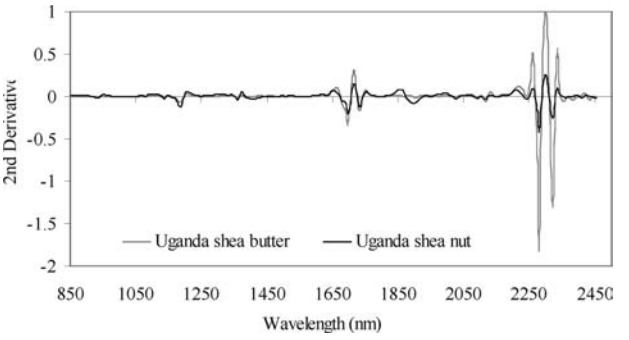

Figure 6. Spectra of Shea nut powder (Uganda) and corresponding extracted butter.

Palmitic Acid. The palmitic acid calibration had a low R_p^2 (0.53) for the validation set compared to R^2 (0.61) of the calibration set. This was due to an overestimation of the R^2 coefficient resulting from extreme values in the calibration set. RPD_p obtained with the model (1.3) also corresponded to an unusable calibration. The lack of correlation between the NIR fingerprint and chemical data resulted from the low variability of palmitic acid content which ranged from 3.31% to 5.62% for 83.8% of samples.

Linoleic Acid. For linoleic acid, the R^2 value for calibration (0.63) was similar to R_p^2 (0.68). This reflected a relatively efficient model for both calibration and prediction. With an RPD_p equal to 1.71, the calibration developed for linoleic acid could therefore be used for rough screening.

Arachidic Acid. For arachidic acid, the model developed was similar to the linoleic acid model with $R^2 = 0.63$, $R_p^2 = 0.57$ and $RPD_p = 1.50$. To conclude, as for linoleic acid, this model could only be applied for very rough screening.

Linolenic Acid. The calibration model for linolenic acid seemed fairly good, as for arachidic and linoleic acids. However, validation revealed the poor performance of the model, with a very low R_p^2 (0.39) and RPD_p (1.30). This result highlighted the importance of using a validation set covering the whole range of values. Indeed, the R^2 observed for linolenic acid was artificially high, due to a bimodal distribution of values for the calibration samples.

Stearic and Oleic Acids. The R^2 values for calibrations were close to 1 for stearic acid (0.96) and oleic acid (0.98). High R^2 coefficients correspond to a very good data fit. These results were confirmed by estimating R_p^2 on the validation set (stearic acid

Table 4. Descriptive Statistics for Calibration Subsets, Validation Subsets and NIR Equations[a]

	MC	fat	palmitic	stearic	oleic	*cis*-vaccenic	linoleic	linolenic	arachidic	ratio[b]
					Calibration $N = 542$					
mean	4.46	49.68	4.24	38.25	48.44	0.40	7.15	0.30	1.20	0.79
range	6.12	31.22	5.73	25.32	23.58	0.82	8.47	0.63	1.37	0.76
SEL	0.1	0.79	0.19	0.98	0.66	0.09	0.34	0.12	0.05	0.02
SD_{cal}	0.84	4.77	0.66	5.68	5.21	0.15	0.90	0.12	0.19	0.17
SEC	0.18	1.00	0.41	1.06	0.74	0.11	0.55	0.07	0.12	0.03
R^2	0.95	0.96	0.61	0.96	0.98	0.48	0.63	0.66	0.63	0.97
SECV	0.20	1.08	0.46	1.14	0.81	0.11	0.62	0.08	0.13	0.04
RPD	4.22	4.43	1.43	4.97	6.47	1.32	1.46	1.54	1.53	4.25
					Validation $N = 60$					
mean	4.44	50.40	4.49	36.69	49.68	0.41	7.28	0.31	1.17	0.76
range	4.65	31.38	3.06	24.16	23.15	0.69	9.11	0.49	0.87	0.79
SD_{val}	1.02	5.88	0.69	7.43	7.11	0.19	1.33	0.13	0.21	0.23
SEP	0.23	1.05	0.53	1.19	0.90	0.16	0.78	0.10	0.14	0.04
R_p^2	0.95	0.97	0.45	0.98	0.99	0.23	0.68	0.39	0.57	0.97
RPD_p	4.45	5.61	1.30	6.26	7.91	1.19	1.71	1.30	1.50	5.72

[a] MC: moisture content. N: number of samples. SEL: standard error of laboratory. SD: standard deviation for calibration subset. SEC: standard error of calibration. R^2: coefficient of multiple determination. SECV: standard error of cross validation. RPD (ratio performance to deviation) = $SD_{cal}/SECV$. SD_{val}: standard deviation for validation subset. SEP: standard error of prediction; R_p^2: coefficient of multiple determination for prediction. RPD_p (ratio performance to deviation for prediction) = SD_{val}/SEP. [b] Ratio saturated to unsaturated fatty acids.

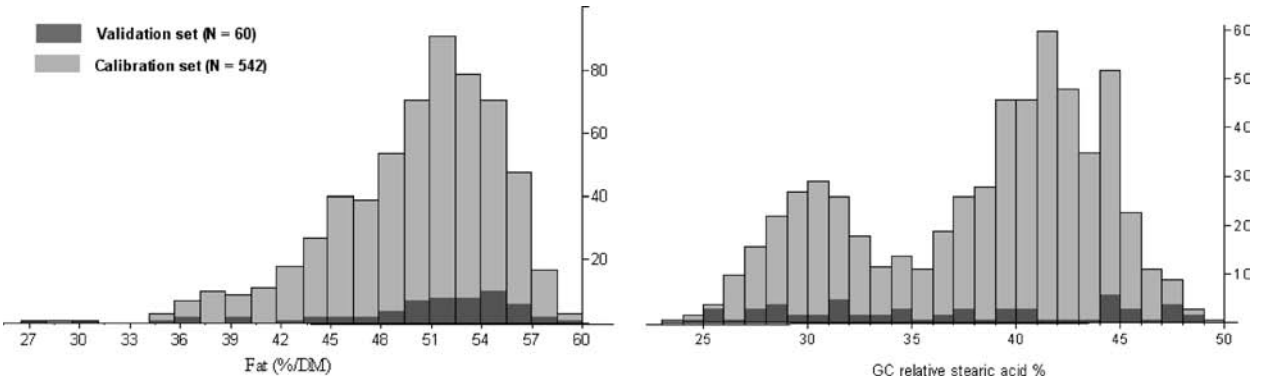

Figure 7. Distributions of fat content and stearic acid % for calibration and validation subsets.

0.98 and oleic acid 0.98). The RPD_p values obtained for stearic and oleic models were 6.26 and 7.91 respectively. Stearic and oleic acids displayed bimodal distributions due to eastern and western butter specificities, however the fit was efficient for each subpopulation, resulting in an overall accurate calibration. SEP values observed on the validation set for stearic and oleic acids were 1.19% and 0.90% respectively. The regression slopes between GC values and NIR-predicted values were 1.04 and 1.06 respectively (**Figures 8c** and **8d**). Both models were set up with a similar number of PLS terms (10 and 11), though the oleic acid calibration performed better. Oleic acid contains a double C-bond (C═C) with a specific signature between 2100 and 2200 nm (*30*). The slight difference between the stearic and oleic models was therefore probably due to this specificity resulting in a higher correlation between oleic acid GC values and NIR data. Furthermore, for regression coefficients, higher correlations were found at 1160 nm and 1172 nm, corresponding to double bond-related wavelengths (**Table 3**) and in the 2140−2180 nm region assigned to the combination of cis ═C−H and C═C stretching, $-CH_2$ asymmetric stretching and C═ stretching of the −HC═CH− structure (*31*).

Saturated vs Unsaturated Acids. Calibrations developed for total saturated fatty acids and total unsaturated fatty acids (mono- and polyunsaturated) led to similarly efficient models, with a RPD_p of 5.72 (**Table 4**). This performance was not surprising due to the high quality of individual calibrations obtained for oleic and stearic acids which represented 86% of unsaturated FA and 87% of saturated FA acids in Shea butter.

Industrial and Research Prospects. Obtaining a precise description of Shea nut quality for sustainable development, research and industry is becoming ever more critical. In this study, we proposed NIR models for accurate prediction of Shea nut stearic and oleic acid composition, fat and water content. Transferring these novel tools to producing countries will provide them with the opportunity to control and promote the quality of their production. In addition, for the food and cosmetics industries, this application will enable early selection of products according to their end use. Accurately predicting the saturated:unsaturated FA ratio is relevant for Shea butter use as a cocoa butter equivalent (CBE) in the chocolate industry.

In terms of research, understanding the environmental and genetic basis of variation patterns in the fat composition of Shea nuts is an important issue. Studies being developed (data not published) reveal a specific genetic variation pattern for a gene strongly involved in the variation of the relative percentages of oleic and stearic acids, and by extension the saturated:unsaturated fatty acid ratio (*32*, *33*). To that end, our NIR models

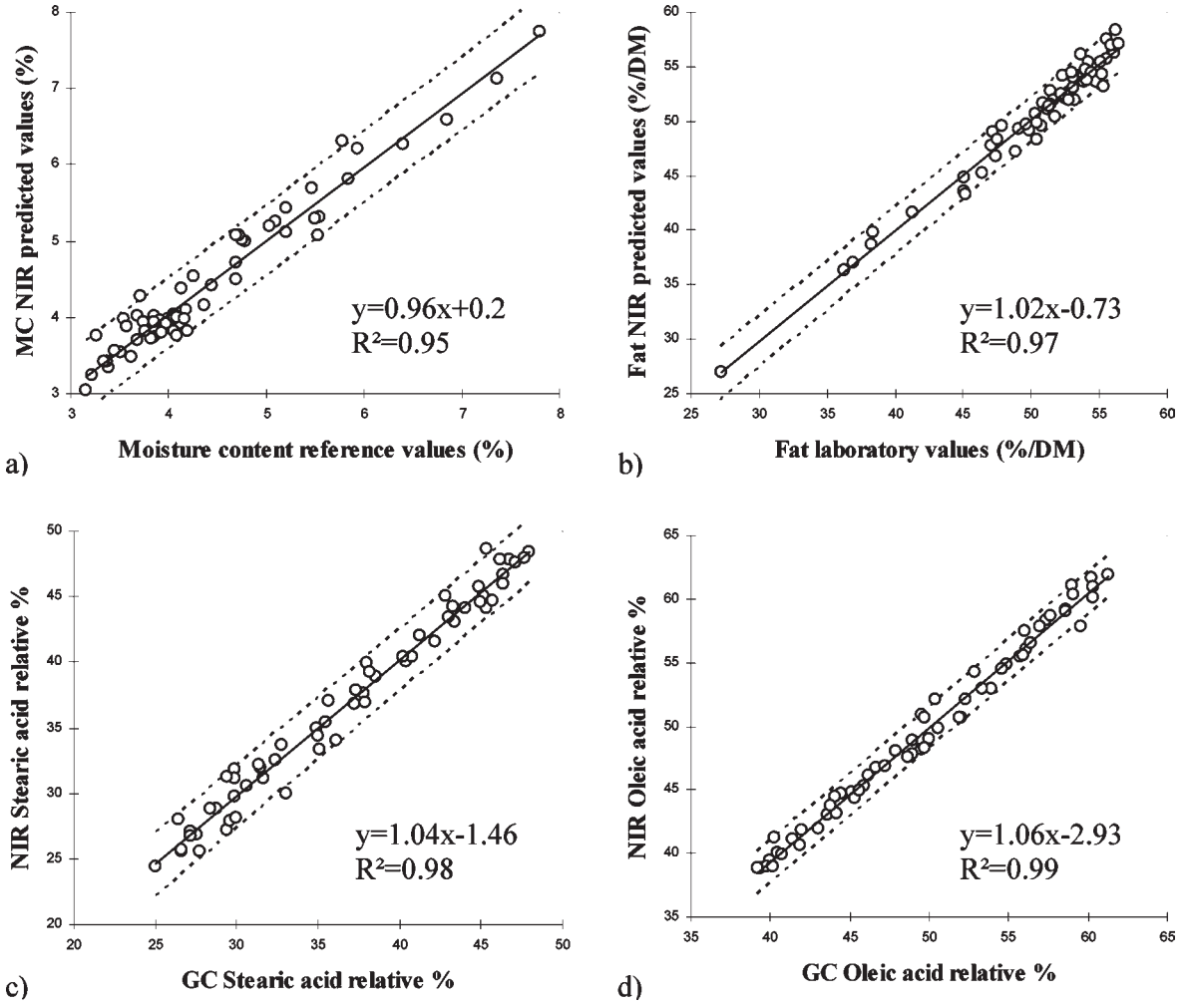

Figure 8. Scatter plots of laboratory values versus NIR-predicted values for the validation subset (95% confidence interval). (**a**) Moisture content as a %. (**b**) Fat content as a % of dry matter (DM). (**c**) Relative stearic acid composition as a %. (**d**) Relative oleic acid composition as a %.

enable us to achieve an accurate, quick and cheap determination of a large number of samples needed for quantitative genetics studies.

In this paper, based on a large collection of samples, we confirmed a differentiation in Shea nut fat composition correlated to geographical origin. The performance of the calibrations obtained indicated that Shea nut moisture and fat contents could be determined with acceptable accuracy using NIR spectroscopy methods. The good calibration performance for relative stearic and oleic acid percentages indicated their suitability for fine determination. NIR calibration enables fine characterization of fat profiles in terms of saturated and unsaturated fatty acids. We demonstrated that NIR allows rapid (less than one minute per sample) nondestructive and reliable determination in one shot of moisture, fat, stearic and oleic acid contents of Shea nuts. Thus our models can be applied for high-throughput characterization of Shea nut quality. Transferring this technology will enable a Shea nut quality control and traceability survey. Moreover, our NIR models make it possible to carry out further quantitative genetics investigations.

ACKNOWLEDGMENT

The laboratory work was conducted at the CIRAD laboratory in Montpellier France, and chemical analyses were carried out by F. Allal, L. Millet and N. Forestier-Chiron. Field activities were conducted by scientists from African countries and from CIRAD, and we thank all our co-workers for their help in fruit collection. We thank Peter Biggins and Guy Self, from CIRAD, for revising the English.

LITERATURE CITED

(1) Hall, J. B.; Aebischer, D. P.; Tomlinson, H. F.; Osei-Amaning, E.; Hindle, J. R. *Vitellaria paradoxa: A Monograph*; School of Agricultural and Forest Sciences, University of Wales: Bangor, U.K.: 1996; p 105.

(2) Neumann, K.; Kahlheber, S.; Uebel, D. Remains of woody plants from Saouga, a medieval West African village. *Veget. Hist. Archaeobot.* **1998**, *7*, 57–77.

(3) Lovett, P.; Haq, N. Evidence for anthropic selection of the Sheanut tree (Vitellaria paradoxa). *Agroforestry systems* **2000**, *48*, 273–288.

(4) Maranz, S.; Wiesman, Z.; Bisgaard, J.; Bianchi, G. Germplasm resources of Vitellaria paradoxa based on variations in fat composition across the species distribution range. *Agrofor. Syst.* **2004**, *60*, 71–76.

(5) Di Vincenzo, D.; Maranz, S.; Serraiocco, A.; Vito, R.; Wiesman, Z.; Bianchi, G. Regional Variation in Shea Butter Lipid and Triterpene Composition in Four African Countries. *J. Agric. Food Chem.* **2005**, *53*, 7473–7479.

(6) Tchobo, F.; Natta, A.; Barea, B.; Barouh, N.; Piombo, G.; Pina, M.; Villeneuve, P.; Soumanou, M.; Sohounhloue, D. Characterization of Pentadesma butyracea sabine Butters of Different Production Regions in Benin. *J. Am. Oil Chem. Soc.* **2007**, *84*, 755–760.

(7) Buchgraber, M.; Ulberth, F.; Anklam, E. Cluster Analysis for the Systematic Grouping of Genuine Cocoa Butter and Cocoa Butter Equivalent Samples Based on Triglyceride Patterns. *J. Agric. Food Chem.* **2004**, *52*, 3855–3860.

(8) Pasquini, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **2003**, *14*, 198–219.
(9) Shenk, J. S.; Westerhaus, M. O. Population Definition, Sample Selection, and Calibration Procedures for Near Infrared Reflectance Spectroscopy. *Crop Sci.* **1991**, *31*, 469–474.
(10) Williams, P.; Norris, K. *Near-Infrared Technology: in the Agricultural and food industries*, 2nd ed.; American Association of Cereal Chemists, Inc: USA, 2001; p 330.
(11) Fox, G.; Cruickshank, A. Near infrared reflectance as a rapid and inexpensive surrogate measure for fatty acid composition and oil content of peanuts (Arachis hypogaea L.). *J. Near Infrared Spectrosc.* **2005**, *13*, 287–291.
(12) Sato, T. New Estimation Method for Fatty Acid Composition in Oil Using Near Infrared Spectroscopy. *Biosci., Biotechnol., Biochem.* **2002**, *66*, 2543–2548.
(13) Sato, T.; Kawano, S.; Iwamoto, M. Near infrared spectral patterns of fatty acid analysis from fats and oils. *J. Am. Oil Chem. Soc.* **1991**, *68*, 827–833.
(14) Sato, T.; Maw, A.; Katsuta, M. NIR reflectance spectroscopic analysis of the FA composition in sesame (Sesamum indicum L.) seeds. *J. Am. Oil Chem. Soc.* **2003**, *80*, 1157–1161.
(15) Grubbs, F. E. Procedures for Detecting Outlying Observations in Samples. *Technometrics* **1969**, *11*, 1–21.
(16) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777.
(17) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Letter: Correction to the description of Standard Normal Variate (SNV) and De-Trend (DT) ransformations in Practical Spectroscopy with Applications in Food and everage Analysis-2nd Edition. *J. Near Infrared Spectrosc.* **1993**, *1*, 185–186.
(18) Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
(19) Shenk, J. S.; Westerhaus, M. O. *Routine Operation, Calibration Development Manual*; Infrasoft International: Silver Spring, MD, 1996.
(20) Wold, H. Non-linear iterative partial least squares (NIPALS) modelling. Some current developments. In *Multivariate Analysis*; Krishnaiah, P. R., Ed.; Academic Press: New York, 1973; Vol. III, pp 387−407.
(21) Wold, S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* **1978**, *20*, 397–405.
(22) Burns, D. A.; Ciurczak, E. W. E. *Handbook of Near-Infrared Analysis*; Marcel Dekker, Inc.: New York, 1992; p 680.
(23) Sinnaeve, G.; D., P.; Agneessens, R.; Biston, R. The use of near infrared spectroscopy for the analysis of fresh grass silage. *J. Near Infrared Spectrosc.* **1994**, *2*, 79–84.
(24) Verger, R.; Rivière, C. Les enzymes lipolytiques: une étude cinétique. *Rev. Fr. Corps Gras.* **1987**, *1*, 7–13.
(25) Lovett, P. N.; Haq, N. Diversity of the Sheanut tree (Vitellaria paradoxa C.F. Gaertn.) in Ghana. *Genet. Resour. Crop Evol.* **2000**, *47*, 293–304.
(26) Osborne, B.; Fearn, T.; Hindle, P. *Practical NIR spectroscopy: with applications in food and beverage analysis*; Longman Scientific & Technical: Harlow, England, 1993; p 227.
(27) Kovalenko, I.; Rippke, G.; Hurburgh, C. Measurement of soybean fatty acids by near-infrared spectroscopy: Linear and nonlinear calibration methods. *J. Am. Oil Chem. Soc.* **2006**, *83*, 421–427.
(28) Guyot, B.; Davrieux, F.; Manez, J. C.; Vincent, J. C. Determination of caffeine and dry matter content by Near Infrared spectrometry. Applications in raw Robusta coffee and roasted coffees. *Café Cacao Thé* **1993**, *37*, 53–64.
(29) ProKarité Regional Technical Committee on Africa regional standard for shea kernel: www.ProKarité.org.
(30) Panford, J.; deMan, J. Determination of oil content of seeds by NIR: Influence of fatty acid composition on wavelength selection. *J. Am. Oil Chem. Soc.* **1990**, *67*, 473–482.
(31) Sohn, M.; Kim, Y.; Vines, L. L.; Kays, S. W. Near infrared analysis of lipid classes in processed cereal products. *J. Near Infrared Spectrosc.* **2009**, *17*, 127–133.
(32) Shanklin, J.; Somerville, C. Stearoyl-acyl-carrier-protein desaturase from higher plants is structurally unrelated to the animal and fungal homologs. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2510–4.
(33) Somerville, C.; Browse, J.; Jaworski, J. G.; Ohlrogge, J. B. Lipids. In *Biochemistry and Molecular Biology of Plants*; Buchanan, B., Jones, R., Gruissem, W., Eds.; American Society of Plant Physiologists: Rockville, MD, 2000; Chapter 10.

Received for review February 1, 2010. Revised manuscript received May 20, 2010. Accepted May 20, 2010. This study was one of the tasks of the INCO INNOVKAR project funded by the European Union for 'Innovative Tools and Techniques for Sustainable Use of the Shea Tree in the Sudano-Sahelian zone'.