JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2009, 57, 10539–10547 10539
DOI:10.1021/jf902675n

NIR Determination of Major Constituents in Tropical Root and Tuber Crop Flours

VINCENT LEBOT,*,† ANTOINE CHAMPAGNE,‡ ROGER MALAPA,§ AND DAN SHILEY∥

†CIRAD, P.O. Box 946, Port-Vila, Vanuatu, ‡Université de Saint Etienne, 42023 Saint-Etienne, France, §VARTC, P.O. Box 231, Santo, Vanuatu, and ∥ASD, Inc., 2555 55th Street, Boulder, Colorado 80301

Tropical root and tuber crops (cassava, sweet potato, taro, and yam) are staples in developing countries where rapid urbanization is strengthening the demand for flour based foods. Quality control techniques are still under development, and when available, laboratory analyses are too expensive. The objectives of this study were to calibrate Near-infrared spectroscopy (NIRS) for routine analysis of flours and to test its reliability to determine their major constituents. Flours prepared from 472 accessions (traditional varieties and breeding lines) were analyzed for their starch, total sugars, cellulose, total nitrogen, and ash (total minerals) contents. The near-infrared (350−2500 nm) spectra of all samples were measured. Calibration equations with cross and independent validation for all analytical characteristics were computed using the partial least squares method. Models were developed separately for each of the four crop species and by combining data from all spp. to predict values within each of them. The quality of prediction was evaluated on a test set of 94 accessions (20%) by standard error of prediction (SEP) and r^2 parameters between the measured and the predicted values from cross-validation. Starch, sugar, and total nitrogen content could be predicted, respectively, with 87%, 86%, and 93% confidence, whereas ash (minerals) could be predicted with 71%, and cellulose was not predictable (r^2 = 0.31). The statistical parameters obtained for starch, sugars, and total nitrogen are of special interest for flour quality control. These constituents are quantitatively the most important in the chemical composition of flours, and starch content is negatively correlated with sugars and total nitrogen. NIRS is a low cost technique well adapted to the conditions in developing countries and can be used for the high-throughput screening of a great number of samples. Possible applications are discussed.

KEYWORDS: Cassava; flours; major constituents; NIRS; sweet potato; taro; yam

INTRODUCTION

In developing countries, the tropical root and tuber crops, cassava (*Manihot esculenta*), sweet potato (*Ipomoea batatas*), taro (*Colocasia esculenta*), and yam (*Dioscorea* spp.), are the second most important group of crops, just after the cereals. The world production was estimated in 2008 to be around 415 millions tons produced from 35 millions ha (*1*). These species belong to different botanical families but are often grouped together because their biological similarities: they are vegetatively propagated, bulky, and perishable. In many countries, they are grown in home gardens or in mixed cropping systems complementing each other throughout the year to produce a steady supply of energy.

Cassava is cooked in fresh and boiled form, in toasted granules (*gari* and *attiéké* in West Africa or *farinha* in Brazil), chips, flour (*lafun*), and as paste (*fufu*) in Africa. Rapid urbanization is strengthening flour demand, and in many African cities, cassava flour is increasingly mixed with wheat flour to prepare local bread and reduce dependency on imported cereals (*2*). Sweet potato is traditionally processed into dried chips and flour to preserve the product. Sun-drying is the simplest dehydration technique, and the roots are washed, peeled, and sliced before being exposed directly to the sun. Different cultivars have different properties, and farmers experiment before selecting the best one for processing, which is often not the one they prefer to consume fresh and boiled (*3*). Taro corms may be roasted, baked, boiled, steamed, or fried, or they may be processed into chips, flakes, and flour. The hot air-dried chips are ground into flour with a hammer mill and can be stored for one year. This type of flour is a base for baby food and taro-based bread. There is significant variation in the functional properties of the taro flours depending on the variety used (*4*, *5*), and there are various improvements depending on the locations and the means available (*6*). Yam flour is prepared from tubers sun-dried for several days. The dried tuber pieces are then pulverized into flour with electric or mechanical mills. The resulting flour can be stored in bags for months and is quite convenient for the evergrowing cities of West Africa. The flour is stirred over boiling water and cooked for a few minutes in order to obtain a thick viscous *fufu* which resembles the one obtained with pounded boiled yam. The varied texture characteristics of yam flours have been shown to be of industrial interest in the

*Corresponding author. Tel/Fax: +678 25847. E-mail: lebot@cirad.fr.

Published on Web 10/30/2009

Philippines (*7*). Their similarities to other commercial starches or flours are useful for the product development of noodles, snacks, and baby foods. In Taiwan, the incorporation of *Dioscorea alata* flour in bread has been shown to markedly increase the antioxidant capacity of the blended bread and does not interfere with its acceptability (*8*).

The physicochemical characteristics of the tropical root and tuber crop raw material reflect the genetic diversity of the cultivars used in smallholder cropping systems. Chemical compositions in major constituents (starch, total sugars, cellulose, total nitrogen, and minerals) also reflect this diversity (*9*). Unfortunately, in many cases, quality control techniques are still under development, and when available, laboratory analyses are too expensive. The paucity of technical information hinders wider utilization and adds constraints to the development of new processed products needed by urban dwellers. Among the numerous research priorities are the characterization and the genetic improvement of the nutritional properties of the tropical root and tuber crops (*10*), but very few breeding programs have the financial and technical means to screen large progenies for their chemical composition.

Near-infrared spectroscopy (NIRS) has become a widely used method of quality control in the food processing industry (*11*−*14*). It is a rapid, cost-effective, and nondestructive technique, allowing the simultaneous determination of major constituents in a mixture by multivariate data analysis. NIRS has already been used to measure potato (*Solanum tuberosum*) tuber quality, but the analytical performance has been shown to be highly dependent on the cultivar (*15*, *16*). Potato starch and protein parameters have been estimated by NIRS. Starch content was found to be predicted with 90% confidence, while total protein content could be predicted with 62% confidence (*17*). Among the tropical root and tuber crops, NIRS application to sweet potato quality control has been investigated, especially for measuring starch quality and pasting properties. NIRS was reasonably accurate in predicting total starch and protein contents, but the other major constituents were not studied (*18*).

Although cassava, sweet potato, taro, and yam flours are commonly used in developing countries, NIRS potential is clearly under-researched, especially for food processing enterprises and for screening large number of samples in breeding programs aiming at improving the chemical composition of cultivated varieties. The objectives of the present study are to calibrate NIRS for the routine analysis of tropical root and tuber crops and to test its reliability in determining major constituents (starch, total sugars, cellulose, total nitrogen, and total minerals).

MATERIALS AND METHODS

Sample Preparation. Overall, 472 samples were collected from germplasm collections maintained at the Vanuatu Agricultural and Research and Technical Center (VARTC) on Santo Island, Vanuatu (**Table 1**). Local cultivars originated from different islands of Vanuatu, while introduced cultivars originated from different Asian countries. Breeding lines were part of populations obtained via open-pollination in polycross plots corresponding to different cycles of a recurrent selection program. The accessions were selected to represent the morphological variation existing within each species. These accessions were planted together and at the same time in a common plot and were harvested when mature. Depending on genotype, harvest was conducted 3−4 months after plantation for sweet potato accessions and 8−10 months after for cassava, taro and yams. Central transverse sections of the roots and tubers were selected and cut for each accession. Approximately 1−2 kg of fresh weight were manually peeled and sliced into chips and oven-dried at 60 °C for 48 h. Dry matter samples were split into two subsamples: one subsample was used for chemical analysis, and the other subsample was used for NIRS measurements. Samples of 150−250 g, prepared at the Food Processing Laboratory of the Department of Agriculture in Port-Vila, Vanuatu, were sent to France for chemical analyses. Samples of approximately 50 g of dried chips were processed into flour just after oven drying and were ground in a stainless kitchen steel mill (SEB, France) in Vanuatu prior to NIRS analysis.

Table 1. Total Number of Accessions Analyzed for Major Constituents and NIRS

species	total accessions	local cultivars	introduced cultivars	breeding lines
cassava (*Manihot esculenta*)	62	62	0	0
sweet potato (*Ipomoea batatas*)	167	21	6	140
taro (*Colocasia esculenta*)	108	47	16	45
yams (*Dioscorea* spp.)	135	105[a]	20[b]	10[c]
total	472	235	42	195

[a] Including 70 *D. alata*, 5 *D. bulbifera*, 14 *D. esculenta*, 6 *D. nummularia*, and 10 *D. transversa*. [b] Including 14 *D. alata* and 6 *D. cayenensis-rotundata*. [c] Including 10 *D. alata*.

Chemical Analyses. Analyses of major constituents (residual moisture, starch, sugars, cellulose, total nitrogen, and ash) were conducted by Laboratoire d'Analyses Agricoles Teyssier, Bourdeaux, France, according to AFNOR (Association Française, the French standards association) and/or EEC methods (The AFNOR group: http://www.boutique.afnor.org/BGR1AccueilGroupe.aspx/).

Following NF (Norme Française) V 18-109 for dry matter determination, samples were dried again to remove residual moisture (measured as % of total dry weight), and the powder was analyzed on an oven-dried air basis. Moisture was therefore expressed as a measurement of the sample prior to drying. All measurements were then expressed in percentages of dry matter (% DM), and the data were adjusted by the residual moisture following oven drying. Starch was quantified using Ewer's protocol (NF ISO 10-520) corresponding to hydrolysis in HCl, filtration, and polarimetric measurement. Total sugars were quantified through the colorimetric method of Luff Schoorl (CEE 98\54\CE). Crude cellulose was measured by the Weende method (NF V 03-040), which corresponds to nonsoluble organic residues obtained by sulfuric acid and alkaline treatments. Total nitrogen content was calculated using the Kjeldahl method (NF V 18-100). Estimation of total mineral content was obtained from ashes produced at 550 °C (NF V 18-101). All analyses were performed in duplicate with an accepted mean coefficient of variation of ±3% for starch, sugars, cellulose, and residual moisture, and ±2% for proteins (total N) and ashes (minerals).

NIRS Measurements and Data Pretreatment. Flour sample granules size was homogenized using four sieves with decreasing diameters until flour granules pass through the 106 μm sieve. An ASD LabSpecPro spectrophotometer from Analytical Spectral Devices Inc. (ASD Inc., Boulder, Colorado, USA) fitted with a muglight or high intensity source probe (HISP) (ASD Inc.) was used for the measurement of all spectra over the wavelength range of 350−2500 nm. On average, 6 g of homogenized flour was placed in an individual cell for the HISP and compacted with a tea spoon to eliminate air voids within the sample. Each spectrum was obtained by averaging three different cells (repetitions) per sample with 25 scans for each. A reference reading (baseline) was taken when starting a session and another every 30 min. All of the spectra were recorded in diffuse reflectance as $\log(1/R)$ with respect to a Labsphere's Spectralon material reflectance standard (Labsphere, Inc.), which is a lambertian reflective PTFE (thermoplastic resin) with high overall reflectance. Overall, 472 spectra were recorded and converted to absorbance (**Figure 1**) using the Indico software (ASD Inc.).

Data Analysis and Model Development. The spectra and reference data were mathematically modeled using GRAMS/AI version 8.0 with PLSPlus/IQ spectroscopy software (Thermo Electron Corporation). Using the values obtained with chemical analyses as the analyte value, a separate calibration was made for each of the five major constituents of the dry matter: starch, total sugars, cellulose, total nitrogen, and ash (minerals). Calibration of residual moisture was not attempted because spectra were recorded in Vanuatu, just after oven drying the samples,

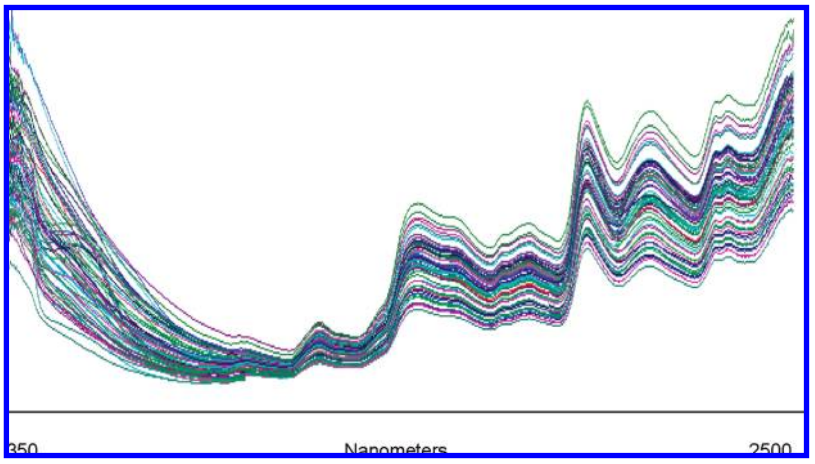

Figure 1. NIRS spectra corresponding to 472 flour samples over 350−2500 nm wavelength (absorbance).

while residual moisture was measured in France on hygroscopic dry raw material.

Partial least-squares (PLS) regression technique was used to develop a predictive model of the near-infrared part of the spectra (1000−2500 nm). The aim of PLS regression was to get as much concentration information as possible into the first PLS factors (*19*). The optimum number of PLS factors used for prediction was determined by full cross-validation and PRESS (prediction residual error sum of squares). Full cross-validation is leave-one-out, and $n - 1$ is the calibration set, and one sample is predicted. This is repeated until all of the samples are predicted. Additionally, light scattering effects due to particle size differences were corrected by multiplicative scatter correction (MSC), a common method squaring the effects by adjusting the spectra based on ranges of wavelengths supposed to carry no specific chemical information. The data was mean-centered and the average spectrum calculated from all of the calibration spectra and then subtracted from every calibration spectrum. Mean centering enhances the subtle differences between the spectra. PLS models were developed for each species separately, and a model combining all crop species (all spp., **Table 3**) together was also created.

In order to assess the performance of the calibration, samples were separated into two sets: the calibration and the prediction sets. The calibration set contained 378 samples, and the prediction set was created by randomly selecting 20% of the accessions within each crop species (over a total of 472 accessions) and included 12 cassava accessions, 33 sweet potato, 22 taro, and 27 yams (total of 94 samples) representing subtest sets for each crop species (**Table 3**).

As part of the model process, a principal component analysis (PCA) was used to check for gross spectral outliers. The Mahalanobis distance of each spectrum to the mean spectrum of the group was calculated. The removal of spectral outliers was based on Mahalanobis distance $H > 3$ from the average spectrum of the file, and outlier samples were removed because of their heterogeneity. Spectra and concentration outliers were removed, and PLS was run again until the highest r^2_{cv} (determination coefficient for cross-validation) and the smallest standard error of cross-validation (SECV) were obtained. At that point, factor loadings were used to determine which wavelengths were important to correlate with concentrations in order to narrow down the spectral region. The PLS analysis was then conducted again on this new region in order to obtain for each constituent equations with higher explanation of the total variability in the calibration values without increasing the number of PLS factors used.

The calibration statistical parameters used to evaluate models performances include the standard error of calibration (SEC), the determination coefficient for cross-validation (r^2_{cv}), the standard error of cross-validation (SECV), the determination coefficient for prediction (r^2_{pred}), and the standard error of prediction (SEP). SEC and SEP were calculated using an Excel spreadsheet by squaring the differences of the actual minus the predicted concentrations for each sample in the calibration (SEC) and test (SEP) sets. These values were then summed, and the sum was divided by the number of samples (n). The square root of this value was used for SEC and SEP. SEC describes the calibration set, and SEP describes the test set composed of samples not included in the calibration set. In addition to the coefficients of determination for cross-validation (r^2_{cv}) and prediction (r^2_{pred}), the ratio of performance to deviation (RPD = SD/SECV) was also used to evaluate performances of the developed models (with SD as the standard deviation of the original chemical data in the calibration set) (*20*−*22*).

RESULTS AND DISCUSSION

Variability of the Major Constituents and Composition. Chemical analyses results are presented in **Table 2**. Overall, the sampling approach based on the selection of morphological variation within each crop species (**Table 1**) resulted in the assemblage of 472 highly variable accessions based on their major constituents values. Except for starch (CV of 10.49%), the four other constituents exhibited remarkable variation with respectively, 73.8% for total sugars, 38.3% for cellulose, 49.84% for total nitrogen, and 30.51% for ashes (**Table 2**). This variation appears somewhat lower than the one observed in the international cassava collection (*23*, *24*) but is comparable to the one reported for sweet potato (*3*), and is greater than those previously reported for taro (*25*, *26*) and yams (*27*−*30*). Correlation coefficients computed independently within each of the four crop species (not presented here) revealed the same trends as those of the whole sample composed of 472 accessions. Starch is negatively but significantly correlated with sugars, cellulose, total nitrogen, and ashes.

NIRS Calibration. NIRS spectra corresponding to 472 flour samples over 350−2500 nm wavelengths are presented in **Figure 1**. PLS regression was applied to develop predictive models for the five major constituents using the chemical values of 378 samples and the near-infrared part of the spectra (1000−2500 nm) (**Table 3**). The r^2 values give an indication of the percentage variation in the Y variable that is accounted for by the X variable. Therefore, r^2 values above 0.50 indicate that over 50% of the variation in Y is attributable to the variation in X. Higher r^2 values improve discrimination. It is generally accepted that models with an r^2 of 0.66−0.81 can be used for screening and that approximate quantitative predictions, models with r^2 values between 0.83 and 0.90, can be used for many applications, while models with values of 0.92−0.96 are suitable for most applications including quality control and those above 0.98 for all applications. Model prediction accuracy was evaluated with RPD: values below 1.5 are considered unusable, those between 1.5 and 2.0 can be used for rough predictions, those between 2.0 and 2.5 allow approximate quantitative predictions, while values above 2.5 and 3.0 are, respectively, considered to be good and excellent predictive models (*23*).

Table 2. Major Constituents Analyzed in 472 Samples of Tropical Root and Tuber Crops

species	*n*	moisture % TDW	starch % DM	sugars % DM	cellulose % DM	total nitrogen % DM	ashes % DM
cassava	*62*						
min		8.45	79.84	1.52	1.73	1.33	1.24
max		11.16	91.21	10.12	6.98	5.61	3.53
mean		9.97	86.49	4.43	3.37	2.59	2.45
std error		9.97	2.68	1.60	1.09	0.82	0.45
CV %		*5.86*	*3.1*	*36.08*	*32.27*	*31.71*	*18.26*
sweet potato	*167*						
min		6.6	53.3	1.49	2.39	2.67	2.06
max		12.04	83.83	25.29	12.56	10.2	8.22
mean		9.61	69.15	10.17	4.11	5.92	3.5
std error		0.58	5.85	4.83	1.46	1.23	0.87
CV %		*11.23*	*8.47*	*47.49*	*35.39*	*20.73*	*24.76*
taro	*108*						
min		8.8	60.77	0.9	1.4	2.3	1.47
max		14.05	88.2	18.58	7.3	14.79	8.13
mean		11.2	78.01	5.17	3.4	5.45	4.08
std error		1.26	5.62	3.00	1.02	1.99	0.99
CV %		*11.41*	*7.21*	*58.01*	*30.11*	*36.44*	*24.28*
yam	*135*						
min		5.6	58.78	0.4	0.1	4.4	1.58
max		12.7	90.4	18.3	6.3	21	8.1
mean		10.7	77.14	3.62	2.68	10.39	4.36
std error		1.1	6.09	3.56	0.98	3.13	1.20
CV %		*10.1*	*7.9*	*98.19*	*36.61*	*30.16*	*27.56*
total	*472*						
min		5.6	53.3	0.4	0.1	1.33	1.24
max		14.05	91.21	25.29	12.56	21	8.22
mean		10.29	75.74	6.4	3.44	6.66	3.74
std error		1.25	7.95	4.72	1.32	3.32	1.14
CV %		*12.17*	*10.49*	*73.8*	*38.3*	*49.84*	*30.51*

Regarding starch, the SECV values observed for each root and tuber crop species were close to the SECs, which means fair and robust fitting. These values were good estimations of the accuracy of the equation as they were close to the standard error of prediction (SEP) obtained on the validation samples (**Table 3**). In fact, when combining the four crops together, the combined model (all spp.) values were 2.41 for SECV, 2.70 for SEC, and 2.74 for SEP. The equation for cassava explained the least amount of variation (82%) within the reference values (chemical results) with a SECV equal to 1.10%, whereas the equation of yam presented an r^2_{cv} equal to 0.88 with a SECV error equal to 2.03% and r^2_{pred} of 0.91. The r^2_{cv} for the equation of all spp. combined was equal to 0.91 with an r^2_{pred} of 0.87. Deviations of single samples are visualized in a scatter plot between measured and predicted values (**Figure 2**). In terms of predictive performance, the equations for starch are good with RPD parameters above 2 for yam and all spp. combined.

The total sugars model developed for all spp. combined presented very similar SECV, SEC, and SEP values, 1.64, 1.64, and 1.66, respectively, indicating very robust fitting. Deviations of single samples are visualized in a scatter plot between measured and predicted values (**Figure 3**). Surprisingly, these values were more variable for models developed for cassava (0.79−0.92−0.58), sweet potato (1.77−2.15−1.53), taro (1.42−1.69−1.21), and yam (0.93−1.21−1.37). Total sugars were predicted with 91% of confidence for sweet potato but only 64% for taro.

Total nitrogen SECV values obtained with the model developed with all spp. combined were close to SEC and SEP values (0.73−0.88−0.77), indicating good and robust prediction with 93% of confidence in prediction. Deviations of single samples are visualized in a scatter plot between measured and predicted values (**Figure 4**). The r^2_{cv} and r^2_{pred} were high for all four species models, with the second ranging from 0.81 for taro to 0.96 for cassava.

Surprisingly and although minerals have a poor relationship with NIRS, they could be predicted with 71% of confidence in all spp. combined, with 82% in taro and up to 90% of confidence in cassava as shown by their respective r^2_{pred} values (**Figure 5**).

Cellulose was not satisfactorily predicted. No PLS term was obtained for cassava, limiting further calculations. The determination coefficients for calibration and prediction (r^2_{cv} and r^2_{pred} values) were very low and were all under 0.4, indicating poor prediction confidence (**Figure 6**).

The confrontation of the NIRS spectra and the chemical value allowed for the establishment of equations of calibration for the prediction of starch, sugars, and total nitrogens. The statistical parameters obtained are of special interest for flour quality control. These constituents are quantitatively the most important ones for determining the chemical composition of flour, sugars and total nitrogen being negatively correlated with starch. Their respective r^2_{pred} values (0.87, 0.86, and 0.93) are high and allow for good estimates of their contents in root and tuber crop flours. However, determination coefficients for the prediction sets (r^2_{pred}) cannot reflect the whole situation because the range of sample values in the prediction test set affects the coefficient value. SEP is, therefore, a better overall indicator.

RPDs are between 2 and 2.5 in eight models allowing for approximate quantitative predictions to be made, and values above 2.5, considered to be good models, are observed for starch in all spp., sugars in yam, and total nitrogen in sweet potato. The model for total nitrogen in all spp. combined with a RPD value above 3 can be considered as an excellent predictive model (*23*). The number of terms is relatively low if we consider a general recommendation of 1 factor for every 10 samples in a model (**Table 3**).

When comparing the performances of the combined model (all spp.) versus models developed for individual crop species, it is

Table 3. Statistical Parameters of the Calibration and Prediction Sets[a]

		calibration set									prediction set					
											each sp. predict[b]				all predict[c]	
model	region (nm)	*n*	mean % DM	SD	SEL ±	out-liers $H > 3$	PLS terms	r^2_{cv}	SECV	SEC	*n*	r^2_{pred}	SEP	RPD	r^2_{pred}	SEP
						Starch										
cassava	1100−2300	*50*	86.01	2.11	2.58	0	8	*0.83*	1.10	1.32	*12*	*0.82*	1.44	1.92	*0.56*	2.07
sweet potato	1000−2400	*133*	68.97	4.53	2.07	0	9	*0.80*	2.60	3.18	*34*	*0.83*	2.55	1.74	*0.81*	2.75
taro	1200−2200	*87*	78.62	4.47	2.36	4	8	*0.78*	2.51	2.95	*21*	*0.86*	2.22	1.78	*0.79*	2.98
yam	1000−2400	*108*	77.77	4.53	2.33	1	11	*0.88*	2.03	2.70	*27*	*0.91*	1.55	2.23	*0.74*	2.79
all spp.	1000−2200	*378*	76.51	6.57	2.30	9	14	*0.91*	2.41	2.70	*94*	*0.87*	2.74	2.73	*0.87*	2.74
						Sugars										
cassava	1000−2200	*50*	4.65	1.15	0.14	2	6	*0.70*	0.79	0.92	*12*	*0.77*	0.58	1.46	*0.86*	0.70
sweet potato	1000−2400	*133*	10.32	3.83	0.31	0	8	*0.86*	1.77	2.15	*34*	*0.91*	1.53	2.16	*0.89*	1.87
taro	1200−2400	*87*	5.33	2.19	0.16	4	6	*0.65*	1.42	1.69	*21*	*0.64*	1.21	1.54	*0.58*	1.51
yam	1200−2400	*108*	3.24	2.60	0.10	6	9	*0.90*	0.93	1.21	*27*	*0.72*	1.37	2.80	*0.55*	1.97
all spp.	1200−2400	*378*	5.74	3.78	0.17	8	10	*0.88*	1.64	1.64	*94*	*0.86*	1.66	2.30	*0.86*	1.66
						Cellulose										
cassava	1000−2500	*50*	3.48	0.81	0.10		0				*12*				*0.01*	1.38
sweet potato	1200−2400	*133*	4.63	0.99	0.14	0	3	*0.14*	1.35	1.87	*34*	*0.25*	0.18	0.73	*0.08*	1.15
taro	1200−2200	*87*	3.18	0.74	0.10	5	2	*0.23*	0.80	0.98	*21*	*0.36*	0.04	0.93	*0.58*	0.85
yam	1000−2400	*108*	2.39	0.75	0.07	3	2	*0.39*	0.69	0.69	*27*	*0.28*	0.81	1.09	*0.34*	0.79
all spp.	1000−2400	*378*	3.34	0.94	0.10	9	8	*0.29*	0.98	1.13	*94*	*0.31*	1.03	0.96	*0.31*	1.03
						Total Nitrogen										
cassava	1200−2400	*50*	2.40	0.62	0.05	4	7	*0.88*	0.29	0.28	*12*	*0.96*	0.14	2.14	*0.95*	0.36
sweet potato	1200−2400	*133*	5.78	0.92	0.12	4	9	*0.88*	0.41	0.61	*34*	*0.81*	0.65	2.24	*0.70*	0.82
taro	1000−2400	*87*	5.68	1.40	0.11	8	9	*0.89*	0.48	0.57	*21*	*0.94*	0.30	2.92	*0.89*	0.67
yam	1200−2400	*108*	10.12	2.38	0.20	3	6	*0.89*	1.03	1.26	*27*	*0.88*	0.78	2.31	*0.85*	0.89
all spp.	1200−2400	*378*	6.53	2.56	0.13	11	11	*0.95*	0.73	0.88	*94*	*0.93*	0.77	3.51	*0.93*	0.77
						Ash										
cassava	1000−2400	*50*	2.58	0.34	0.05	2	6	*0.85*	0.17	0.20	*12*	*0.90*	2.62	2.00	*0.94*	0.15
sweet potato	1200−2400	*133*	3.33	0.54	0.07	3	5	*0.40*	0.55	0.70	*34*	*0.43*	0.51	0.98	*0.26*	0.61
taro	1200−2400	*87*	4.22	0.72	0.08	5	6	*0.84*	0.35	0.46	*21*	*0.82*	0.62	2.06	*0.68*	0.74
yam	1300−2400	*108*	4.38	0.95	0.09	1	7	*0.48*	0.87	1.38	*27*	*0.61*	0.80	1.09	*0.60*	0.90
all spp.	1000−2400	*378*	3.89	0.88	0.08	7	11	*0.70*	0.57	0.70	*94*	*0.71*	0.70	1.54	*0.71*	0.70

[a] r^2_{cv} = determination coefficient of calibration; SD = standard deviation of the chemical data; SEL = standard error of the reference method; SECV = standard error of cross-validation; SEC = standard error of calibration; r^2_{pred} = determination coefficient of prediction; SEP = standard error of prediction; RPD = ratio of performance to deviation (RPD = SD/SECV). [b] Models developed for each sp. are used to predict each sp. test sets. [c] Models developed for all spp. combined are used to predict each sp. test sets.

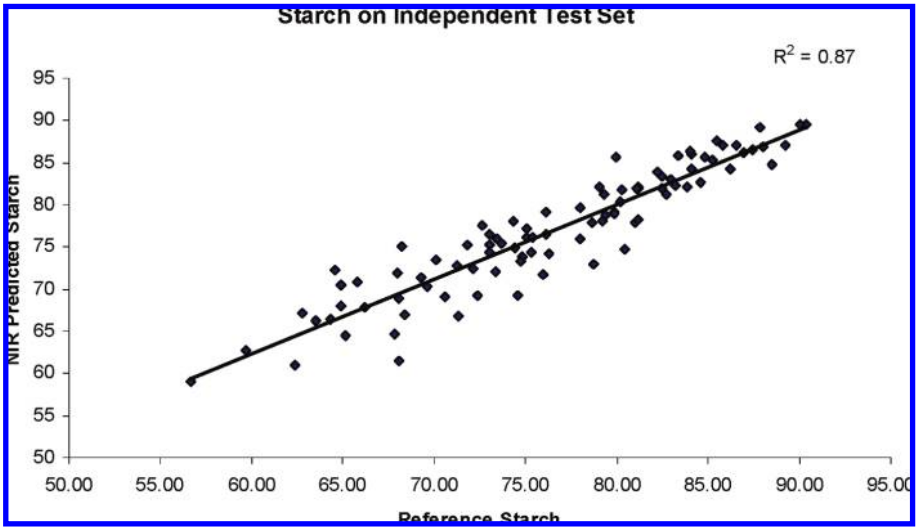

Figure 2. Starch prediction comparison on an independent test set.

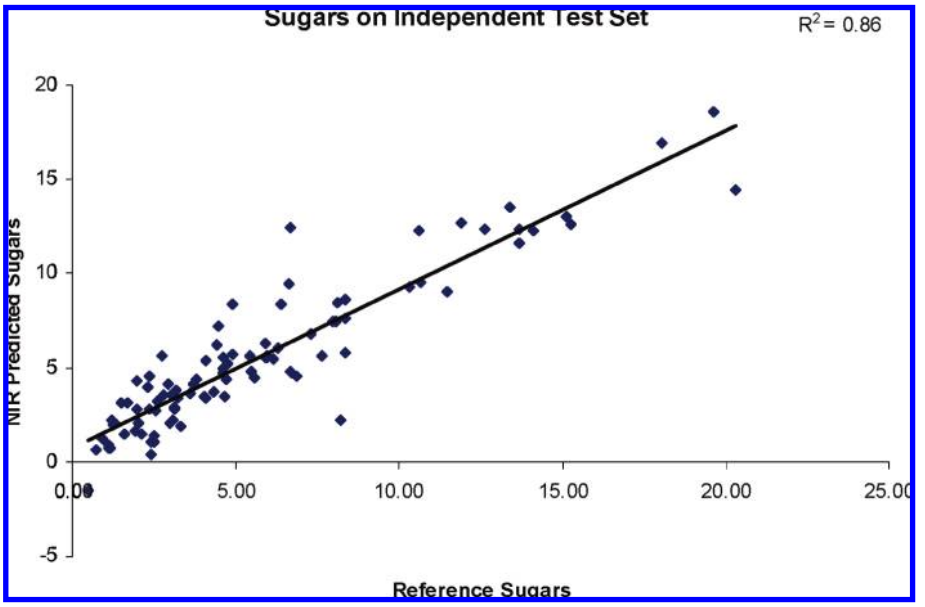

Figure 3. Sugar prediction comparison on an independent test set.

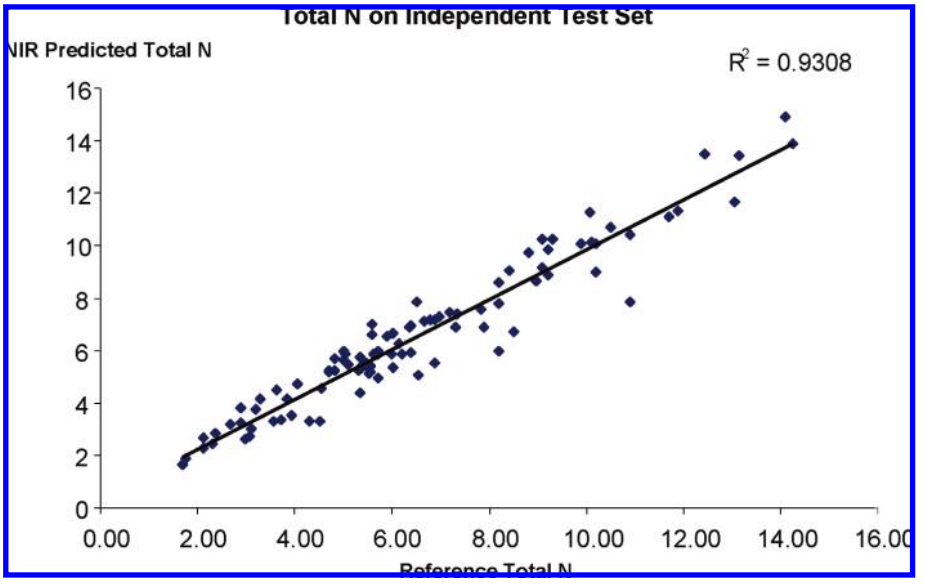

Figure 4. Total nitrogen prediction comparison on an independent test set.

interesting to observe, in the case of starch, for example, that since all SEP values were similar to the overall SEP, no crop species was poorly represented by the combined model. Determination coefficients generally improve as the working range increases. Consequently, if more ranges are added when different species are combined in the same model, then they could improve coefficient values. Additionally, when different species values are combined, a larger spectral diversity is described, and therefore, some samples within a particular species might actually be better spectrally described as the spectra of the four species have been added together. When comparing the performance of the models developed using the values of all spp. combined, in most cases the SECV individual was close enough to the SEP combined, indicating fairly robust fitting. Also, comparing the SECV combined with the SEP combined (on individual species) showed that almost all species were well represented by the combined model. This demonstrates that individual species can be combined in a single model without adverse results. These findings show that NIR spectroscopy has the potential to serve as a rapid method for predicting the chemical composition of root and tuber crop flours.

Similar studies have already been conducted on flours made from cereals. NIRS has been shown to be useful for predicting the protein composition of rice flour, the best model giving $r^2 = 0.992$ (SEP = 0.138%) (*31*, *32*). The use of NIRS has also been proposed to monitor the protein content of flour in order to optimize the milling conditions. Control of the blending of flours or supplementation with wheat gluten to achieve a composite flour of a given protein content has been suggested (*33*). The accuracy and precision of NIR for protein, moisture, particle size, color, and starch damage were satisfactory for quality control purposes and rapid flour testing (*34*). For wheat flour, protein content has been predicted with accuracy similar to that of the reference method with an r^2 of 0.99 (*35*). NIRS is also an interesting technique for breeders. NIR analysis has been shown to be sufficiently accurate for the routine screening of large numbers

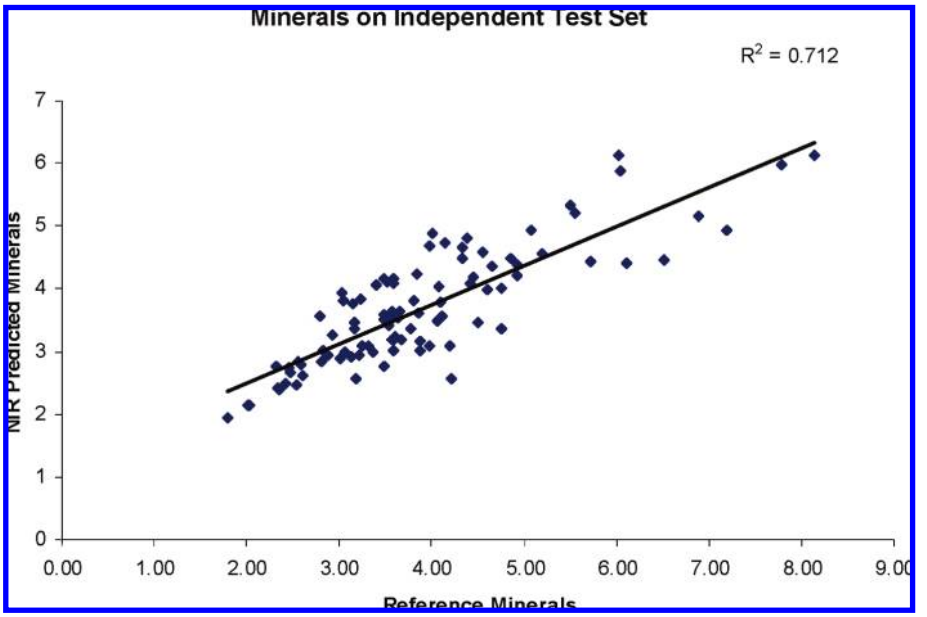

Figure 5. Mineral prediction comparison on an independent test set.

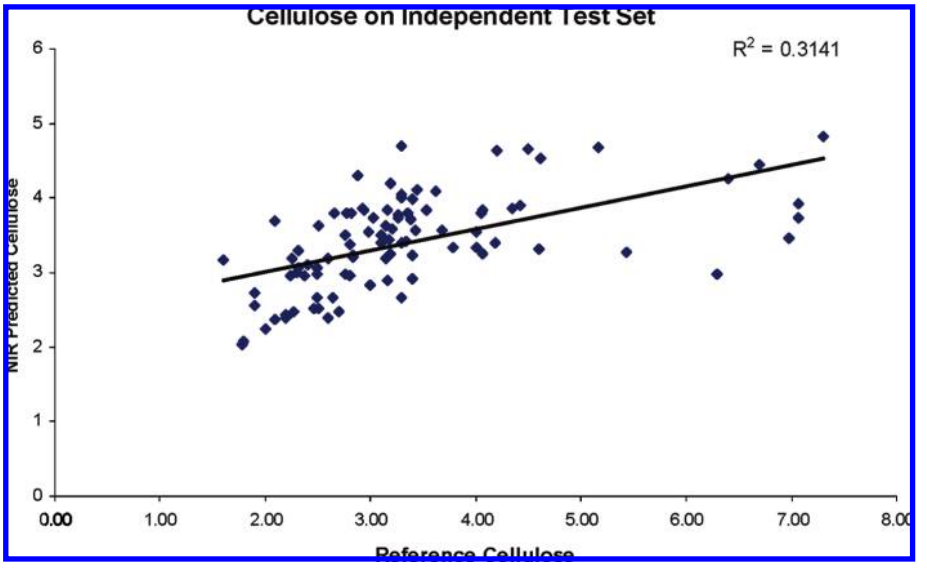

Figure 6. Cellulose prediction comparison on an independent test set.

of samples in early generation selection in rice breeding programs (*36*). Whole NIRS spectra provided a useful tool for describing the global evolution of the chemical composition of the grain of French wheat and for analyzing their evolution (*37*).

The protein content is usually estimated by multiplying the total nitrogen content by a standard conversion factor of 6.25. However, the nitrogen to protein ratio does vary according to the species considered. This factor is different for each crop species and changes with amino acid content and mineral nitrogen and nonprotein nitrogen in flour. Because conversion factors are critical to enabling the simple conversion of total nitrogen values into protein values, it would be necessary to use specific conversion factors for different root crop species (*38*, *39*). This would be more accurate and preferable when attempting to express nitrogen as protein. Unfortunately, although some work has been done on tropical root crop species, such as sweet potato (*40*), the greater yam (*D. alata*) (*41*), and cassava (*42*), not much is known for other yam species and taro. For cassava, however, the work was conducted on only 15 varieties, and conversion factors based on Kjeldahl nitrogen ranged from 2.5 to 3.7. For the present study, we therefore decided to present our results as total nitrogen and not protein. It would be of interest in the future to improve the calibration models on the real protein contents of each species which vary according to amino acids, the principal nutritional viewpoint.

NIRS is a rapid and reliable technique that has found many applications within the fields of food and crop analysis (*43*). The advantages of NIRS are very low operating costs and a lack of production of chemicals (toxic) and/or waste products. NIRS predictions are frequently more reproducible than the chemical analyses used as the reference method. The absence of chemicals involved and the simple spectral collection help to eliminate operator errors and improve the transferability of methods between countries. The major constraint is the need to build a stable and reliable calibration model, which in itself is dependent on robust and accurate laboratory references and a large and diverse sample calibration set. These attributes are making NIRS an interesting technique for developing countries, either for quality control or for breeding programs.

At present, depending on the technical and financial means, there is some variation between the existing breeding programs, but the rationale is the same. Heavy selection pressure is applied very early for resistance to diseases. The selection process is visual in order to minimize the costs and maximize the number of genotypes assessed. The selected clones are then released as new varieties. A new selection cycle begins in which the new varieties are used as parents. The selection process is equivalent to mass recurrent selection. Great numbers have to be screened to achieve some progress. The process is based on the capture of additive effects and is particularly efficient for traits with high heritability when there is a broad genetic base. Unfortunately, chemotypes with attractive properties are often eliminated because of the high selection pressure on other traits. A common difficulty in breeding programs is to satisfy the requirements of both the industry and the fresh market. An industrial variety must have high dry matter and starch contents (for alcohol production and starch extraction). For human fresh consumption, flesh color and good cooking quality are important traits. High carotene and proteins are preferred for the feed industry. Obviously, NIRS could assist breeders in their choice and selection of the best genotypes, on the basis of the chemical composition required for their ideotype depending on the chemotype requested by the market (high starch, amylose, sugars, or protein content).

The results from this study demonstrate that NIRS has good potential for the screening of starch, sugars, and total nitrogen contents of flours made from tropical root and tuber crops. The combined model (all spp.) remains interesting. It is not as accurate as the single species models, but if we consider that unknown genotypes, not previously encountered with the single species model, are added to a collection, it is likely that the combined model would do better because it contains a wider range of diversity. For a germplasm screening tool, it would probably make sense to use the combined model because it represents more spectral variation. It could be of particular interest for genetic resource curators willing to characterize their numerous accessions rapidly. Such a model has a wide prediction range which allows the rapid assessment of unknown genotypes, and it could therefore be used for preliminary screenings.

The species predictive models show good accuracy, but it remains to be seen whether larger sample sets will improve models sufficiently to enable more precise prediction of the concentrations within each crop species (cassava, sweet potato, taro, and yam). Further work should concentrate on validating the results across a wider range of genotypes within the four major species, over different years, and on developing more stable predictive models. These species models could be useful for breeding programs. Other constituents such as amylose, carotenoids, or individual sugars should be investigated.

ABBREVIATIONS USED

r^2_{cv}, determination coefficient of calibration; SD, standard deviation of the chemical data; SEL, standard error of the reference method; SECV, standard error of cross-validation; SEC, standard error of calibration; r^2_{pred}, determination coefficient of prediction; SEP, standard error of prediction; RPD, ratio of performance to deviation (RPD = SD/SECV).

ACKNOWLEDGMENT

We thank S. Stephens for laboratory assistance in preparing samples and taking spectra.

LITERATURE CITED

(1) FAO. www.fao.org.
(2) Westby, A. Cassava Utilization, Storage and Small-Scale Processing. In *Cassava, Biology, Production and Utilization*; Hillocks, R. J., Thresh, J. M., Bellotti, A. C., Eds.; CABI Publishing: Wallingford, U.K., 2002; pp 281−300.
(3) Woolfe, J. A. *Sweet Potato, An Untapped Food Resource*; Cambridge University Press: Cambridge, U.K., 1992; p 643.
(4) Njintang, Y. N.; Mbofung, C. M. F.; Moates, G. K.; Parker, M. L.; Craig, F.; Smith, A. C.; Waldron, W. K. Functional properties of five varieties of taro flour, and relationship to creep recovery and sensory characteristics of *achu* (taro based paste). *J. Food Eng.* **2007**, *82* (2), 114–120.
(5) Njintang, Y. N.; Kesteloot, R.; Blancher, G.; Aboubakar; Scher, J.; Mbofung, C. M. F. Puncture and stress relaxation behaviour of taro paste as affected by moisture content and puncture rate and depth. *J. Food Eng.* **2008**, *86* (3), 349–357.
(6) Huang, A. S.; Komarasamy, K.; He, L. Recent Developments in Taro-Based Food Products in Hawaii. In *Proceedings of the 3rd Taro Symposium*; Guarino, L., Taylor, M., Osborn, T., Eds.; Secretariat of the Pacific Community: Nadi, Fiji; **2004**, pp 226−229.
(7) Salda, V. B.; Ramsden, L.; Sun, M.; Corke, H. Genetic variation in physical properties of flour from selected Asian yams (*Dioscorea* spp.). *Trop. Agric.* (*Trinidad*) **1998**, *75* (2), 212–216.
(8) Hsu, C. L.; Hurang, S. L.; Chen, W.; Weng, Y. M.; Tseng, C. Y. Qualities and antioxidant properties of bread as affected by the incorporation of yam flour in the formulation. *Int. J. Food Sci. Technol.* **2004**, *39*, 231–238.
(9) Bradbury, J. H.; Holloway, W. D. Chemistry of Tropical Root Crops: Significance for Nutrition and Agriculture in the Pacific, 1988; ACIAR Monograph No. 6., Canberra; p 201.
(10) Lebot, V. *Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids*; CABI Publishing: Oxfordshire, U.K., 2009; p 413.
(11) Burns, D. A.; Ciurczak, E. W., Eds. *Handbook of Near-Infrared Analysis*; Marcel Dekker: New York, 1992.
(12) Downey, G. Qualitative analysis in the near-infrared region. *Analyst* **1994**, *119*, 2367–2375.
(13) Bertrand, D.; Dufour, E., Eds. *La Spectroscopie Infrarouge et ses Applications Analytiques*, 2nd ed.; Lavoisier: Paris, France, 2006; p 660.
(14) Huang, H. B.; Yu, H. Y.; Xu, H. R.; Ying, Y. B. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.* **2008**, *87*, 303–313.
(15) Hartmann, R.; Bûning-Pfaue, H. NIR determination of potato constituents. *Potato Res.* **1988**, *41*, 327–334.
(16) Dijk van, C.; Fischer, M.; Holm, J.; Beekhuizen, J. G.; Stolle-Smits, T.; Boeriu, C. Texture of cooked potatoes (*Solanum tuberosum*). 1. Relationships between dry matter content, sensory-perceived texture, and near-infrared spectroscopy. *J. Agric. Food. Chem.* **2002**, *50* (18), 5082–5088.
(17) Haase, N. U. Rapid estimation of potato tuber quality by near-infrared spectroscopy. *Starch* **2006**, *58* (6), 268–273.
(18) Lu, G.; Huang, H.; Zhang, D. Prediction of sweetpotato starch physiochemical quality and pasting properties using near-infrared reflectance spectroscopy. *Food Chem.* **2006**, *94* (4), 632–639.
(19) Martens, M.; Naes, T. Assessment, Validation and Choice of Calibration Method. In *Multivariate Calibration*; Wiley: New York, 1989.
(20) Sinnaeve, G.; Dardenne, P.; Agneessens, P.; Biston, R. The use of near infrared spectroscopy for the analysis of fresh grass silage. *J. Near Infrared Spectrosc.* **1994**, *2*, 79–84.
(21) Williams, P. A Short Course in the Practical Implementation of near-Infrared Spectroscopy for the User. In *Near-Infrared Technology: Getting the Best out of Light*, 5.3 ed.; PDK Projects, Inc.: Nanaimo, Canada, 2003.
(22) Saeys, W.; Mouazen, A. M.; Ramon, H. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosyst. Eng.* **2005**, *91*, 393–402.
(23) Ceballos, H.; Sánchez, T.; Chávez, A. L.; Iglesias, C.; Debouck, D.; Mafla, G.; Tohme, J. Variation in crude protein content in cassava (*Manihot esculenta* Crantz) roots. *J. Food Comp. Anal.* **2006**, *19*, 589–593.

(24) Chávez, A. L.; Sánchez, T.; Jaramilo, G.; Bedoya, J. M.; Echeverry, J.; Bolaños, E. A.; Ceballos, H.; Iglesias, C. A. Variation of quality traits in cassava roots evaluated in landraces and improved clones. *Euphytica* **2005**, *143*, 125–133.
(25) Wang, J. K. Taro, A Review of *Colocasia esculenta* and Its Potentials; University of Hawaii Press: Honolulu, HI, 1983; p 397.
(26) Aregheore, E. M.; Perera, D. Dry matter, nutrient composition and palatability/acridity of eight exotic cultivars of taro (*Colocasia esculenta*) in Samoa. *Plant Foods Hum. Nutr*. **2003**, *58*, 1–8.
(27) Agbor Egbe, T.; Trèche, S. Evaluation of the chemical composition of Cameroonian yam germplasm. *J. Food Comp. Anal*. **1995**, *8*, 274–283.
(28) Egesi, C.; Asiedu, R.; Egunjobi, J. K.; Bokanga, M. Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L). *J. Sci. Food Agric*. **1998**, *83*, 858–865.
(29) Tamiru, M.; Maass, B. L.; Pawelzik, E. Characterizing diversity in composition and pasting properties of tuber flour in yam germplasm (*Dioscorea* spp.) from Southern Ethiopia. *J. Sci. Food Agric*. **2008**, *88*, 1675–1685.
(30) Baah, F. D.; Maziya-Dixon, B.; Asiedu, R.; Oduro, I.; Ellis, W. O. Nutritional and biochemical composition of *D. alata* (*Dioscorea* spp.) tubers. *J. Food Agric. Environ*. **2009**, *7* (2), 373–378.
(31) Chen, J. Y.; Miao, Y.; Sato, S.; Zhang, H. Near infrared spectroscopy for determination of the protein composition of rice flour. *Food Sci. Technol. Res*. **2008**, *14* (2), 132–138.
(32) Himmelsbach, S.; Barton, F. E.; McClung, A. M.; Champagne, E. T. Protein and apparent amylose contents of milled rice by NIR-FT/Raman spectroscopy. *Cereal Chem*. **2001**, *78* (4), 488–492.
(33) Osborne, B. G. The use of near infrared reflectance spectroscopy in the control of flour protein content. *J. Chem. Technol. Biotechnol*. **2007**, *36*, 364–366.
(34) Osborne, B. G.; Douglas, S.; Fearn, T. The application of near infrared reflectance analysis to rapid flour testing. *Int. J. Food Sci. Technol*. **2007**, *17*, 355–363.
(35) Vàsquez, D.; Williams, P. C.; Watts, B. NIRS Spectroscopy as a Tool for Quality Screening. In *Wheat production in Stressed Environment*; Buck, H. T. et al., Eds.; Springer: New York, 2007; pp 527−533.
(36) Bao, J. S.; CM., Y. Z.; Corke, H. Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy. *J. Food Sci*. **2001**, *66* (7), 936–939.
(37) Roussel, V.; Branlard, B.; Vézine, J. C.; Bertrand, D.; Balfourier, F. NIRS analysis reveals temporal trends in the chemical composition of French bread wheat accessions cultivated between 1800 and 2000. *J. Cereal Sci*. **2005**, *42* (2), 193–203.
(38) Sosulski, F. W.; Imafidon, G. I. Amino acid composition and nitrogen-to-protein conversion factors for animal and plant foods. *J. Agric. Food Chem*. **1990**, *38* (6), 1351–1356.
(39) Mariotti, F.; Tomé, D.; Patureau Mirand, Ph. Converting nitrogen into protein—beyond 6.25 and Jones' factors. *Crit. Rev. Food Sci. Nutr*. **2008**, *48* (2), 177–184.
(40) Purcell, A. E.; Walter, W. M.; Nicholaides, J. I. Nitrogen, potassium, sulfur fertilization, and protein-content of sweet-potato roots. *J. Am. Soc. Hortic. Sci*. **1982**, *107* (3), 425–427.
(41) Go, O.; Ory, R. L.; Graves, E. E. Protein-nitrogen conservation in fresh stored *Dioscorea* yams. *ACS Symp. Ser*. **1986**, *312*, 261–272.
(42) Yeoh, H. H.; Truong, V. D. (1996) Protein contents, amino acid compositions and nitrogen-to-protein conversion factors for cassava roots. *J. Sci. Food Agric*. **1996**, *70* (1), 51–54.
(43) Woodcock, T.; Downey, G.; O'Donnell, C. P. Better quality food and beverages: the role of near infrared spectroscopy. *J. Near Infrared Spectrosc*. **2008**, *16*, 1–29.

Received for review July 31, 2009. Revised manuscript received October 8, 2009. Accepted October 15, 2009. This project was funded by FFEM (Fond Français pour l'Environnement Mondial).