JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2006,** *54,* 8501−8509 **8501**

Development of NIRS Equations for Food Grain Quality Traits through Exploitation of a Core Collection of Cultivated Sorghum

L. F. DE ALENCAR FIGUEIREDO,*,†,‡ F. DAVRIEUX,§ G. FLIEDEL,|| J. F. RAMI,† J. CHANTEREAU,⊥ M. DEU,† B. COURTOIS,† AND C. MESTRES||

Cirad, UMR PIA, TA 40/03, Avenue Agropolis, Montpellier, F-34398 France, Universidade Católica de Brasília, SGAN 916, Mod. B. Asa Norte, Brasília, DF, Brazil, 70790-160, Cirad, UPR Qualité produits pérennes, Montpellier, F-34398 France, Cirad, UPR Qualité des aliments, Montpellier, F-34398 France, and Cirad, UPR Agrobiodiversité en savane, Montpellier, F34398 France

A sorghum core collection representing a wide range of genetic diversity and used in the framework of a sorghum breeding and genetics program was evaluated by near-infrared reflectance spectroscopy (NIRS) to predict food grain quality traits: amylose content (AM), protein content (PR), lipid content (LI), endosperm texture (ET), and hardness (HD). A total of 278 sorghum samples were scanned as whole and ground grain to develop calibration equations. Laboratory analyses were performed on NIRS sample subsets that preserved the core collection racial distribution. Principal component analysis performed on NIRS spectra evidenced a level of structure following known sorghum races, which underlined the importance of using a wide range of genetic diversity. Performances of calibration equations were evaluated by the coefficient of determination, bias, standard error of laboratory (SEL), and ratio of performance deviation (RPD). Ground grain spectra gave better calibration equations than whole grain. PR equation (RPD of 5.7) can be used for quality control. ET, LI, and HD equations (RPD of 2.9, 2.6, and 2.6, respectively) can be used for screening steps. Even with a small SEL in whole sample analysis, a RPD of 1.8 for AM confirmed that this variable is not easy to predict with NIRS.

KEYWORDS: Sorghum; core collection; amylose content; protein content; lipid content; endosperm texture; grain hardness; near-infrared reflectance spectroscopy (NIRS); partial least-square regression

INTRODUCTION

Sorghum (*Sorghum bicolor* (L.) Moench) is a major food cereal in Asian and African countries; therefore, grain quality is an essential trait for farmers and consumers and is becoming a criterion of increasing importance for breeding programs.

Grain quality demand varies depending on the type of final product but is essentially determined by grain biochemical and physical characteristics. Amylose (AM), protein (PR), and lipid (LI) contents influence rheological and sensory properties of some traditional sorghum dishes. The consistency of thick porridge such as tô in western Africa or ugali in eastern Africa is significatively and positively correlated with AM but negatively correlated with PR and LI (*1*). Cooked couscous firmness correlates positively with apparent AM (*2*) while waxy sorghums (with no amylose) produce sticky masa and tortillas with unacceptable rollability (*3*). Endosperm texture (ET) and hardness (HD) affect grain mold resistance, grain storage ability, milling behavior, flour particle size, and cooking properties (*4*−*6*). Hard grains produce flours containing a high proportion of coarse particles with low ash content and high level of damaged starch and yield a high proportion of desirable sorghum couscous granules (*2*). Sorghum cultivars with soft endosperms give good quality injera with soft texture (*7*).

When a large number of samples require analysis, as is usual in breeding programs, standard laboratory methods for measuring these quality traits appear cumbersome, time-consuming, and expensive. The use of near-infrared reflectance spectroscopy (NIRS) seems a more appropriate method, being faster and nondestructive when applied on whole grains. The technique is based on the vibrational properties of molecules and their interactions with light. Spectral data are correlated with biochemical component contents obtained through standard methods (*8*). However, NIRS is an indirect method that needs a large number of samples, covering a broad variability for each trait with a distribution more or less uniform between extreme values, to obtain an accurate calibration equation (*8*).

* Corresponding author. Tel.: +33 4 67 61 55 48. Fax: +33 4 67 61 56 05. E-mail: alencar@cirad.fr.
† Cirad, UMR PIA.
‡ Universidade Católica de Brasília.
§ Cirad, UPR Qualité produits pérennes.
|| Cirad, UPR Qualité des aliments.
⊥ Cirad, UPR Agrobiodiversité en savane.

10.1021/jf061054g CCC: $33.50
Published on Web 10/11/2006

Few studies have reported the use of NIRS on sorghum, with most performed on ground grains. Rami (*9*) evaluated grain biochemical (AM, PR, and LI) and physical characteristics (ET and HD) from ground grains of two recombinant inbred line populations. Fontaine et al. (*10*) evaluated PR and amino acid contents from ground grains of 167 samples. Hicks et al. (*11*) compared whole and ground grain NIRS calibrations for starch content, LI, PR, and protein digestibility for 23 sorghum lines and hybrids planted in two locations with four replications. Rami (*9*) and Fontaine et al. (*10*) reported equations of good performance for PR and HD, with low SECV (standard error of cross validation), but calibration equations were not found sufficiently efficient for prediction of AM and LI (*9*−*11*).

In these previous studies, accessions were limited in number (*11*) or were very specific, such as mapping populations (*9*). The use of a sorghum core collection, which is a set of accessions that best represents the species genetic diversity (*12*), should bring a natural and continuous variability among traits of interest and improve the quality of NIRS prediction equations. A sorghum core collection of 210 accessions was developed from International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) germplasm banks, based on criteria of racial classification, geographic origin, photoperiod sensitivity, and molecular genetic diversity (*13*). The accessions composing the core collection are cultivated varieties and represent a broad diversity of grain size, shape, and color, as well as various consumers' usages and tastes in terms of grain quality. The measurement of their biochemical and physical characteristics, however, had not yet been undertaken.

The objective of this study was to take advantage of the large diversity of the sorghum core collection to develop NIRS calibrations for the most important food grain quality traits (AM, PR, LI, ET, and HD) and use these calibrations to predict the whole core collection to be able, later, to draw on the results for varietal comparisons and genetic analyses in the framework of a sorghum breeding program. Calibration equation performances obtained from whole and ground grain will be compared and the possibility to develop a nondestructive measurement method eliminating the tedious grinding step will be discussed.

MATERIALS AND METHODS

Materials. We analyzed 205 accessions of the core collection, belonging to five basic races, according to the classification proposed by Harlan and de Wet (*14*): Guinea ($n = 60$), Caudatum ($n = 43$), Durra ($n = 29$), Bicolor ($n = 24$), Kafir ($n = 18$), and five intermediate races ($n = 31$). The accessions originated from 39 countries representing sorghum production areas.

Most samples ($n = 199$) were harvested from an irrigated trial conducted during the 2002−2003 dry season at CERAAS (Centre d'Etude Régional pour l'Amélioration de l'Adaptation à la Sécheresse) in Senegal. A dry season trial was chosen to ensure the most homogeneous conditions during grain filling, enhancing the expression of genetic differences in grain quality between varieties. The varieties were sown in an augmented design with 10 blocks and 27 plots per block, with six varieties used as control in each block. The six control varieties were Argence, Vacarès, DK18, IRAT 204, BF201, and B3052. With the exception of IRAT 204, all are breeding materials from temperate areas. Eight local varieties were added to complete the experimental design. A total of 266 samples were collected because three varieties did not germinate and one control was discarded. Twelve varieties, including amylose and protein mutants, not included in the trial were added to the set. A total of 278 samples were analyzed.

Sample Preparation. Specific operations were conducted to remove sources of bias in varietal comparisons. Grains were manually cleaned to remove glumes, dust, broken grains, green grains, and grains damaged by insects and were calibrated by using 2 min of sifting through a sieve, adapted to the average grain size of each sample (Alpine type 200 LS air jet sifter). Moisture content was measured in 40 samples. The average moisture content was used to calculate the quantity of water to be added to each sample to reach 11.5% moisture content. Samples were stored in individual plastic containers for a minimum of 8 days minimum before analysis. For all biochemistry and NIRS analyses, 20 g of cleaned and conditioned sample was ground in a Perten Mill 3100 with 0.8 mm sieve.

NIRS Instrumentation and Measurement. A monochromator instrument NIRS 6500 (Foss NIRSystems, Silver Spring, MD) was used to scan the reflectance from 400 to 2500 nm at 2 nm intervals, using ring cups (50 mm diameter) with 3.0 g of ground grain and 6.5 g of whole grain, respectively. Data were saved as the average of 32 scans and stored as log(1/reflectance) using a ceramic standard reference spectrum. Spectrum acquisitions were duplicated for whole grain (two filled cells) and the spectrum average was stored. For ground grains, spectra were acquired once.

Statistical analyses were performed using Win-ISI II software (Infrasoft International, Port Matilda, PA).

A principal component analysis (PCA) was run and the generalized Mahalanobis distances (H) were computed for each spectrum. All samples having H value above 3 were considered as outliers.

Choice of the Samples for Laboratory Analyses. A subsample of the core collection was selected for laboratory analyses from grain spectra, based on neighborhood Mahalanobis distances (NH). Two samples were considered neighbors when NH was below 0.75. The sample with most neighbors was selected and its neighbors were eliminated. This process was iterated until all samples had been rejected or selected.

Laboratory Analyses. AM was assessed by measuring the latent heat of melting of the complex formation between amylose and phospholipids. It was measured by differential scanning calorimetry on 10−11 mg of ground grain in the presence of a lysophospholipid solution with a Perkin-Elmer DSC 7 (*15*). PR was calculated from the total organic nitrogen content ($N \times 6.25$) determined by the Kjedahl method using 1.0 g of flour with a Kjeltec 1030 Tecator. LI was determined by using the Soxhlet method after extraction with diethyl ether with a Soxtec HT distiller (Tecator). ET was determined by visual assessment of the relative proportion of vitreous and floury endosperm areas on longitudinal cross sections of 15 grains. Each grain was scored on an international scale from 1 to 5 (± 0.5) where 1 corresponded to a completely vitreous endosperm and 5 to a completely floury endosperm (*16*). HD was evaluated using a particle size index (PSI) method, evaluating the percentage of ground grain that sifted through a 250 μm sieve after 2 min by using an Alpine type 200 LS air jet sifter (*16*). The higher the PSI, the more floury the grain. A randomized part of the samples was measured twice to compute the standard error of laboratory (SEL).

NIRS Calibration Development. Unless specified, spectra were mathematically corrected for light scattering by using the standard normal variate and the detrend correction (*17*). The first and second derivative of log(1/R) were systematically compared to log(1/R) for best calibration performance. It was calculated using a gap of 5 points and 5 point polynomial smoothing.

Calibrations were performed using a modified partial least-square (mPLS) regression of WinISI. Calibration statistics included the following parameters: standard deviation (SD) of the population, coefficient of determination (RSQ) and regression slope, bias, standard error of calibration (SEC), and SECV. For SECV, 25% of the samples was used to validate a calibration model developed with the other 75%. SECV was repeated four times and the average calculated. The ratio of performance deviation (RPD) was calculated as SD/SECV. The Student test (t) was used to identify t-outlier samples.

RESULTS

PCA. We have analyzed a core collection representative of the genetic diversity of cultivated sorghums to predict the main food grain quality traits using NIRS. PCA spectra from whole

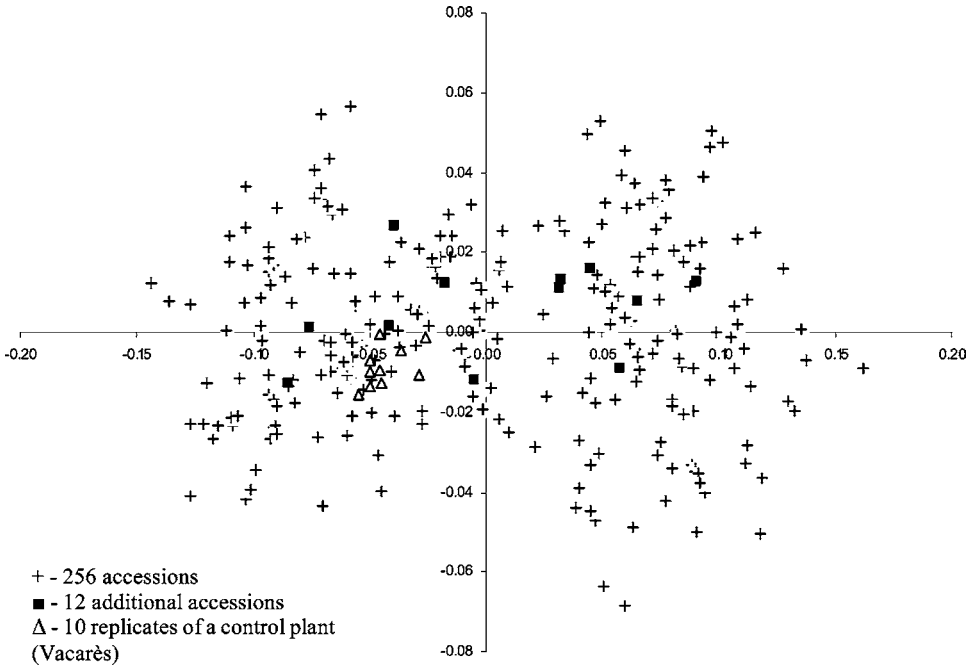

Figure 1. Scatter plot scores of accessions for the first two principal components from ground grain spectra. The 10 replicates of a control variety (Δ) and 12 additional accessions (■) are individualized.

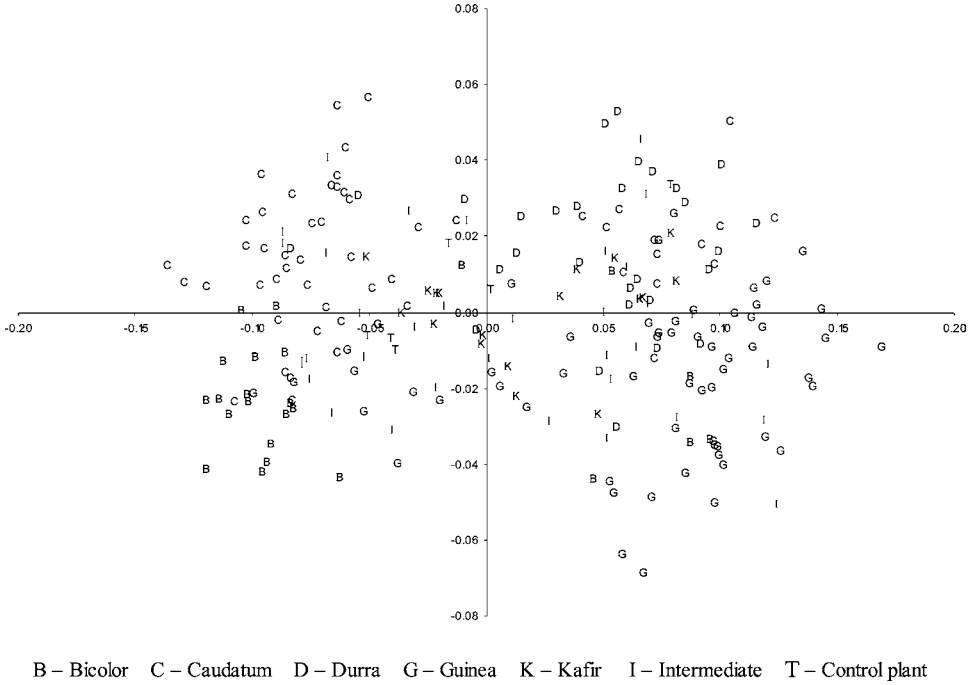

Figure 2. Scatter plot scores of accessions classified by race for the first two principal components from ground grain spectra. One replicate only per control variety is represented on the graph.

and ground grain revealed the importance of the use of a large genetic diversity to develop calibration equations. The first 15 principal components explained 99.7% of the variance of whole grain NIRS spectra, with the two first axes accounting for 71.6% and 18.2%, respectively. For ground grain spectra, the first 24 principal components explained 99.6% of the variance, the two first axes accounting for 39.0% and 34.8%, respectively. The overall variability was well-represented by the first plan of the ground grain PCA, which was used to present graphics in this paper.

Control varieties were replicated 10 times according to the experimental design. **Figure 1** shows the score plots of the 10 replications for the control variety Vacarès. These replications clustered close together (*H* value mean of 0.6), confirming the absence of a block effect in the Senegal trial (*18*). The same pattern was observed for the other five control varieties (data not shown). A comparable grouping was observed for the whole grain spectra.

The 12 extra samples that were not grown in the Senegal trial were not *H* outliers (average *H* value of 1.27 with a minimum of 0.53 and a maximum of 2.38) and appeared scattered in the core collection (**Figure 1**).

Figure 2 shows the importance of using a core collection with a diverse racial representation to sample a large diversity. It shows a tendency for races to gather on the first plan. The accessions belonging to bicolor race clustered on the left bottom

Table 1. Laboratory Data of Sorghum Samples (Three Mutants with No Amylose Excluded)[a]

trait	N	mean	range	SD	CV_{AS}	SEL	CV_{AD}	SD/SEL
AM	275	20.8	15.2–24.6	1.5	7.5	0.50	2.42	3.0
PR	137	13.9	7.3–18.0	2.0	14.4	0.14	1.01	14.3
LI	107	3.82	2.89–6.19	0.62	16.3	0.09	1.60	6.9
ET	275	2.8	1.0–5.0	0.9	32.1	0.13	4.5	6.9
HD	117	13.6	8.2–25.7	3.6	26.5	0.35	2.6	10.3

[a] AM: amylose content in % dry basis; PR: protein content in % dry basis; LI: lipid content in % dry basis; ET: endosperm texture; HD: grain hardness; *N*: number of samples; SD: standard deviation; CV_{AS}: coefficient of variation among samples (%); SEL: standard error of laboratory for duplicated samples; CV_{AD}: coefficient of variation among duplicates (%).

side of the plan. The caudatum accessions were mostly present in the upper quadrants of the graph with two subgroups, one in each quadrant. Most of durra accessions were concentrated in the right upper quadrant. The guinea were relatively well-distributed except in the upper left quadrant. The kafir were grouped in the graph center. The accessions of intermediate races were dispersed in the whole plan. This tendency was confirmed for whole grain spectra (data not shown).

While some trends were also observed on the first PCA plan for grain characteristics such as grain color or presence of a testa (data not shown), there were enough intermediate samples to consider the spectra population as homogeneous.

Sample Selection. Outliers were removed from the calibration when their *H* values for the average spectrum were computed above 3. For ground grain spectra, four samples were excluded. Two of them were mutants with no amylose (*H* of 3.36 and 3.45). To improve homogeneity, the third mutant, with an *H* of 2.86, was also removed, bringing the number of ground grain samples down to 273. For whole grain spectra, in addition to the mutants, another three samples were excluded, reducing the number of whole grain samples to 272.

Once outliers were discarded, a neighborhood clustering based on *NH* values was used to select a subsample representative of the core collection diversity, which was analyzed by laboratory standard methods. Starting from whole grain spectra, a first set of 84 samples was selected. This set was complemented by 26 samples selected from ground grain spectra taking into account ground grain specific diversity. The racial distribution of these 110 samples was similar to that of the core collection with nearly 50% of the accessions selected for each race, showing that NIRS sampling strategy preserved the core collection racial distribution.

The number of 110 samples for laboratory analyses was progressively increased for traits that did not give excellent calibration equations. As a whole, 30 samples selected based on *NH* from ground grain were added for PR and 10 for HD, while all the samples were analyzed for AM and ET.

Laboratory Analyses. Laboratory analyses were duplicated on all the selected samples for PR (140) and a portion for AM (88 out of 278), LI (84 out of 110), and HD (58 out of 120). The number of samples analyzed in the laboratory (*N*), mean, range, standard deviation (SD), and coefficient of variation among samples (CV_{AS}), standard error of laboratory (SEL), and coefficient of variation among duplicates (CV_{AD}) are shown in **Table 1** for the five traits after exclusion of the mutants with no amylose. The repeatability (SEL) was excellent for all biochemical parameters with coefficients of variation among duplicates (CV_{AD}) varying from 1.0% to 2.4%. For physical parameters, CV_{AD} was also good for HD (2.6%) while ET had the highest CV_{AD} (4.5%).

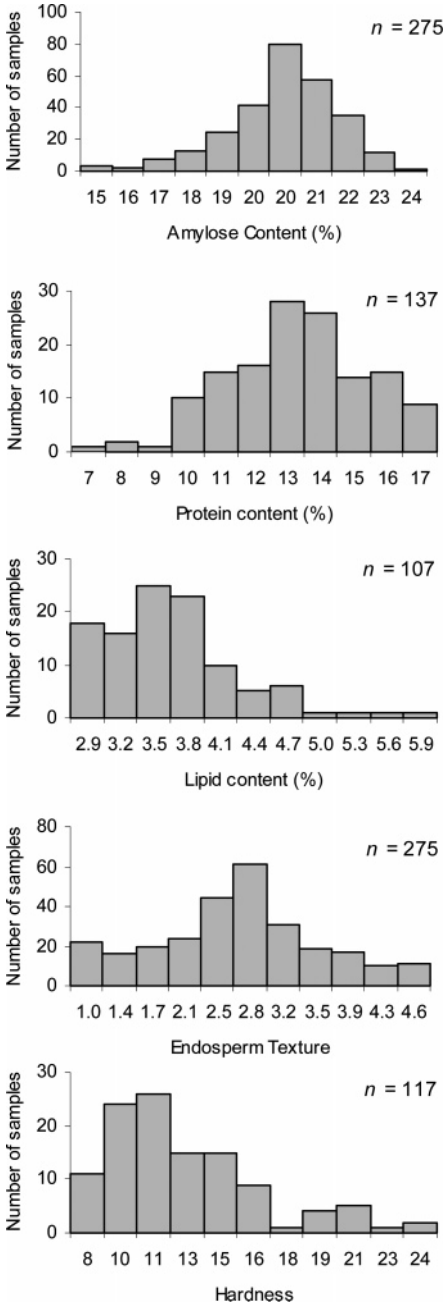

Figure 3. Distribution of amylose content, protein content, lipid content, endosperm texture, and grain hardness from laboratory analyses.

Data distribution was Gaussian only for PR (**Figure 3**). AM had the lowest diversity with values ranging from 15.2 to 24.6 (CV_{AS} of 7.5%, **Table 1**). LI had intermediate diversity (CV_{AS} of 14−16%). These two biochemical traits lacked samples in the extreme values: five samples had amylose content lower than 17% and four had lipid content over 5.2. By contrast, physical traits (HD and ET showed the highest CV_{AS} (around 30%). The whole scale was covered for ET with three cultivars completely floury (ET = 5.0) and three others completely vitreous (ET = 1.0).

Calibration from Whole and Ground Grain Spectra. Calibrations were established from whole and ground grain spectra. During the calibration process, *t*-outliers were removed from the data set (**Tables 2** and **3**, **Figure 4**). For AM, the 10 *t*-outliers comprised some extreme values, hence lowering SD, particularly for ground grain calibration. Few *t*-outliers were evidenced for PR, LI, and HD, but they also comprised some

Table 2. NIRS Calibration Statistics to Physical and Biochemical Characteristics of Sorghum Ground Grains[a]

trait	Nc	t	mean	range	SD	slope	RSQ	bias	SEC	SEC V	RPD	wavelength range	math. corr.
AM	263	10	20.8	15.2–23.6	1.41	0.99	0.75	0.17	0.71	0.77	1.8	1108–249 2	second der, SNVD
PR	131	4	14.0	9.2–18.0	2.0	1.00	0.98	0.00	0.28	0.35	5.7	908–2492	second der, SNVD
LI	101	4	3.78	2.89–5.61	0.56	1.00	0.91	0.00	0.17	0.19	2.9	908–2492	second der, SNVD
ET	263	10	2.9	1.0–5.0	0.93	1.00	0.87	0.00	0.33	0.36	2.6	908–2492	SNV
HD	109	2	13.5	8.2–25.5	3.56	1.00	0.90	0.00	1.14	1.37	2.6	1108–249 2	second der, SNVD

[a] AM: amylose content in % dry basis; PR: protein content in % dry basis; LI: lipid content in % dry basis; ET: endosperm texture; HD: grain hardness; Nc: number of samples used in the calibration; *t*: *t*-outliers; SD: standard deviation; RSQ: coefficient of determination; SEC: standard error of calibration; SECV: standard error of cross validation; RPD: ratio of performance deviation; math. corr.: mathematical correction; der: derivative; SNVD: standard normal variate detrend; SNV: standard normal variate.

Table 3. NIRS Calibration Statistics for Physical and Biochemical Characteristics of Sorghum Grains

trait	Nc	t	mean	range	SD	slope	RSQ	bias	SEC	SECV	RPD	wavelength range	math. corr.
AM	259	13	20.9	17.2–23. 6	1.29	0.99	0.70	0.22	0.71	0.75	1.7	1108–2492	second der, SNVD
PR	131	3	14.0	9.2–18.0	2.01	1.00	0.95	0.00	0.47	0.62	3.2	908–2492	second der, SNVD
LI	102	2	3.8	2.89–6.1 9	0.60	1.00	0.84	–0.02	0.24	0.33	1.8	908–2492	second der, SNVD
ET	260	13	2.9	1.0–5.0	0.93	1.00	0.85	0.01	0.36	0.45	2.1	908–2492	second der
HD	106	5	13.5	8.2–25.5	3.60	0.99	0.88	0.16	1.27	1.71	2.1	1108–2492	second der, SNVD

[a] AM: amylose content in % dry basis; PR: protein content in % dry basis; LI: lipid content in % dry basis; ET: endosperm texture; HD: grain hardness; Nc: number of samples used in the calibration; *t*: *t*-outliers; SD: standard deviation; RSQ: coefficient of determination; SEC: standard error of calibration; SECV: standard error of cross validation; RPD: ratio of performance deviation; math. corr.: mathematical correction; der: derivative; SNVD: standard normal variate detrend.

extreme values (the lowest PR value and the highest LI value for ground grains). Several *t*-outliers were detected for ET but none of them was an extreme value and SD of the calibration population was similar to that of actual values (**Table 1**). The *t*-outliers did not share any specific characteristic in terms of race, grain color, or presence of a testa.

The RSQ, the bias, and the RPD are the most meaningful statistics for rapid appraisal of calibration performance (*19*). The calibration performance was good for PR, particularly for ground grains (**Table 2** and **Figure 4**): the RPD was 5.7 and the RSQ was 0.98 while the regression line between actual and predicted PR values had a slope of 1.00 without any bias. RPD of 3.2 and RSQ of 0.95 for PR were slightly lower for whole grains whereas SECV almost doubled (from 0.35 to 0.62), indicating a degradation of the calibration performance. Nevertheless, slope and bias did not change. A quite similar figure was observed for LI, HD, and ET calibrations, with lower RPD (over 2.5 for ground grain calibration and very close to 2 for whole grain calibration (**Tables 2** and **3**, respectively) and RSQ values (over 0.84). The bias was null for ground grain calibration and remained low for whole grain calibration. Based on RPD, the calibration performance for AM was poor. It was nevertheless very similar for whole and ground grain (RPD = 1.7 and 1.8, respectively).

The best calibrations were obtained with complete scatter correction (second derivative and standard normal variate detrend (SNVD) correction) for chemical components and HD. For ET, however, scatter correction could be detrimental, and better calibrations were obtained without derivating ground grain spectra and without SNVD correction of whole grain spectra. Spectral information was reduced to the infrared portion for all models: the visible region was detrimental in all cases, increasing noise and decreasing the performance of the models. Full infrared region (908−2492 nm) was used for PR, LI, and ET but the reduction to the 1108−2492 nm portion improved the calibration performances of AM and HD either for whole and ground grain spectra.

DISCUSSION

We have analyzed a large collection of accessions representative of the genetic diversity of cultivated sorghum using NIRS. The quality of standard analyses is a prerequisite for developing NIRS calibration. Williams (*19*) recommended for example a CV of the reference methods (CV_{AD}) lower than 2% for chemical composition but close to 1% for protein content. The CV_{AD} of PR was 1.0% (**Table 1**) and most traits had a CV_{AD} below 2.6%. However, ET showed a lower precision (CV_{AD} of 4.5%) which was linked to the method used, based on a score after visual examination of 15 grain cross sections. It could be improved by using objective assessment of vitreous and floury endosperm areas as used for maize (*20*) and by increasing the number of analyzed grains.

Spectral reproducibility is another prerequisite. It can be estimated by measuring the rms (root mean square) of subsamples. In our study the rms values were below 250 μabs (absorbance) units for whole grain and below 200 μabs for ground grain. This underlined the quality of spectrum data and the proper manipulations during harvest, conditioning, and NIRS acquisitions. The vicinity of control varieties on PCA figures also confirmed the minimal variation due to environment heterogeneity in the trial (*18*).

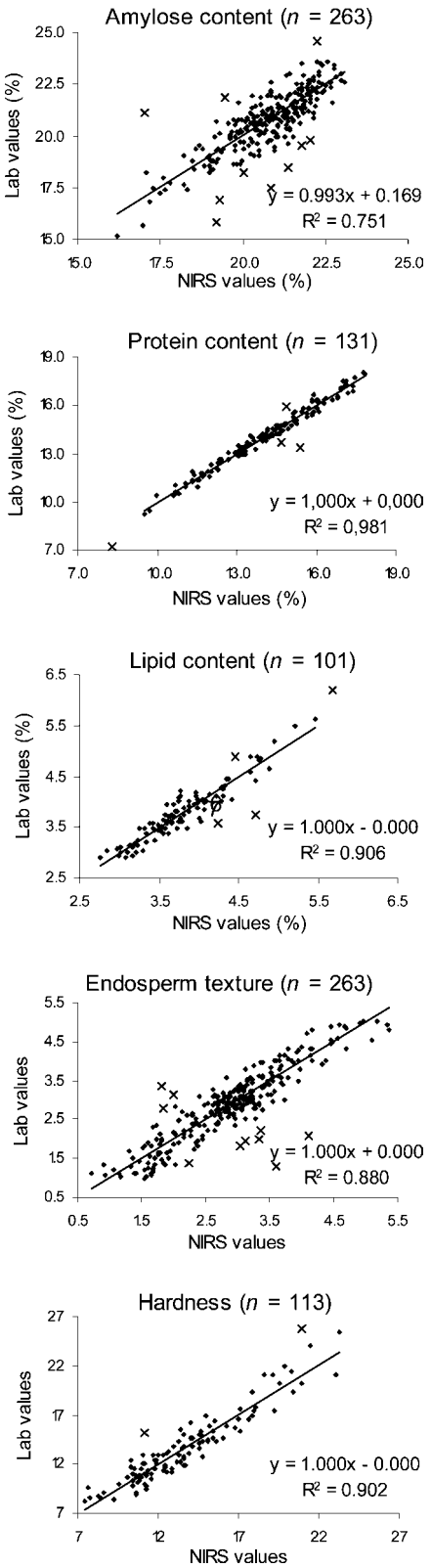

Figure 4. Regression of NIRS predicted data on laboratory data using ground grain spectra, with all *t*-outliers excluded but visible on the figure as (x) for each trait.

PCA figures of NIRS ground grain spectra evidenced a cluster formation based on race basis. Samples of each race were hence necessary to represent the grain genetic diversity in *Sorghum bicolor*. This validated the approach of using a core collection for building a NIRS model. It should be noticed that although the core collection was not built on food grain quality criteria, it included reasonably diverse samples of each race for these traits.

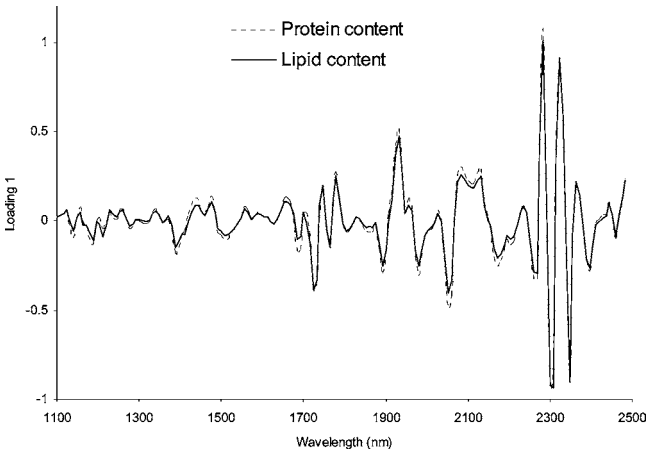

Figure 5. Loading spectrum of PR and LI along all nonvisible light (908–2484 nm) for the first modified partial least-square.

Interpretation of PLS factors can provide information on the intrinsic features of the relationship between the optical and reference data (*19*). We attempted to interpret in particular the loadings of the first axis that accounted for the major contribution of the models for PR, LI, HD, and AM. Surprisingly, the loadings of the first factor were very similar for PR and LI (**Figure 5**). This may be due to spectral acquisition and reference data obtained on whole grain (ground or not); lipids and proteins originating from the germ may indeed contribute to a large portion of the variability, being both largely correlated to the size of the germ. This will be further explored by analysis focused on physicochemical relationships within the sorghum core collection. The major bands could be interpreted quite easily by reference to a near-infrared second derivative spectra database (*21*). The sharp doublet centered on 2300 nm was clearly due to lipids. Negative doublets around 2000 and 1700 nm could be assigned to proteins and the positive doublet around 2100 nm to starch. The peak observed around 1900 nm was surely related to water. It should be noticed that the bands assigned to lipids and proteins varied in the same direction as for the pure components whereas an inverse figure was observed for starch. Loadings were hence more correlated with LI and PR bands and less correlated with starch bands. This appeared meaningful, hence confirming the validity of the models for PR and LI.

AM and HD loadings of the first PLS factor were also quite similar (**Figure 6**). They presented in particular the very sharp doublet around 2300 nm, but as a top/down image of LI and PR loadings. Similarly, starch and protein bands centered at 2100 nm, and at 2000 and 1700 nm, respectively, were present and reversed comparing to LI and PR loadings. AM loadings showed in addition to a negative peak around 2430 nm and a negative/positive peak around 1400 nm that both are typical of starch second derivative spectra. Furthermore, it could be noticed that the negative doublet centered at 2100 nm was more intense before this peak, with a weak shoulder after it. This appeared also very meaningful as amylose presents this figure whereas amylopectin shows adversely a main peak above 2100 nm and a shoulder before (*22*).

The figure was quite different for ET. The three first axes of the PLS accounted for the same weight. The loadings on the two first axes appeared mainly linked to light absorption and scattering, particularly in the very near infrared (below 1100

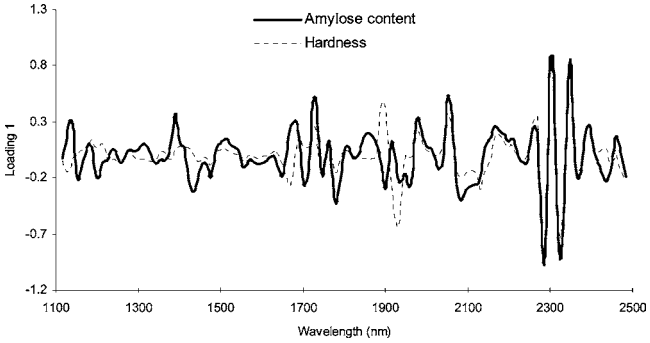

Figure 6. Loading spectrum of AM and HD for the first modified partial least-square.

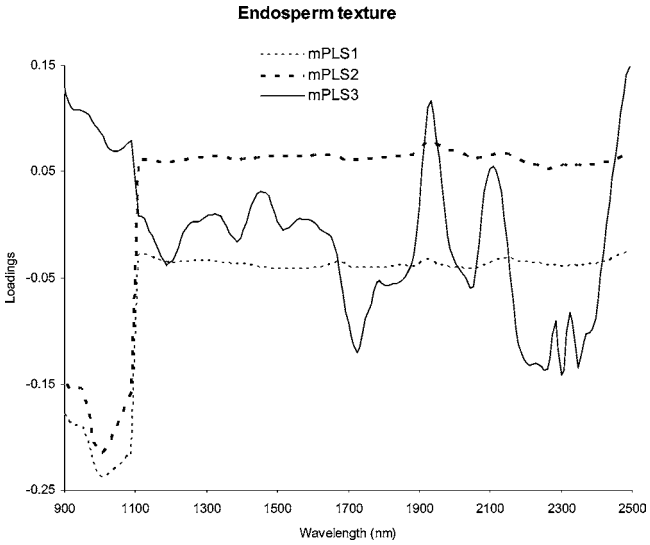

Figure 7. Effect of scattering phenomenon for ET among loading spectra for the three first modified partial least-squares (mPLS).

nm, **Figure 7**). Absorption bands could only be evidenced on the third one, with the bands assigned to lipids centered at 2300 nm, to starch at 2100 nm, and to proteins around 1950 nm. This confirmed that ET was mainly modeled by NIRS through the scattering phenomenon, in accordance with the nature of the trait itself.

Data range for all five traits were similar to some previously reported in studies involving large accession numbers and measured by standard analysis methods (*23*). It was also similar to data obtained in previous studies dealing with NIRS calibration for PR and AM (*9*, *10*) but higher for LI and physical traits (*9*). The narrow AM range, with a continuous distribution close to normal once the mutants were removed, may be due to selection because of domestication of accessions assigned to human consumption for which high AM is required in many traditional dishes (*16*). It is a general rule in chemometrics that to obtain good NIRS calibration, the distribution in the calibration population should be uniformly distributed between the extreme values (*8*). It was more or less the case for ET but not for AM, PR, LI, and HD. Another rule to obtain good calibration is to have a sufficiently reliable reference method which can be measured by the ratio of the SD in the population to the SEL reference method (SD/SEL, *10*). The ratio should be over 5 to 20 (*8*, *10*) but never less than 3. We obtained a satisfactory ratio for most traits except for AM (**Table 1**). In short, our data confirmed that even with a precise reference method (CV_{AD} of 2.4%), the natural genetic diversity of AM was not suitable for developing a good calibration.

Calibration performance is largely appraised by the RPD, which is considered adequate for screening in breeding programs when ranging from 2.5 to 4.9 and adequate for quality control when ranging from 5 to 10 (*24*). We obtained a RPD of over 5 only for PR with ground grain spectra. Our models for PR either from grain or ground grain spectra were more accurate than those obtained by Hicks et al. (*11*) and Rami (*9*). Fontaine et al. (*10*), however, recorded a higher RPD (7.7), probably due to a larger number of samples ($n = 167$). PR is by far the most frequent application of NIRS analysis, particularly on cereals for which high RPDs are found (*10*, *24*, *25*).

LI showed the second highest RPD on ground grain spectra. It was higher than that obtained by Rami on sorghum (*9*), and in the same range of those obtained for wheat (*26*) or chickpea and pea (*25*). Prediction accuracy was much lower for the whole grain spectra model, in agreement with Flinn et al. (*25*). Vines et al. (*27*) obtained a much higher RPD (8.5) for LI. However, their work was based on cereal food products (breakfast cereals, snacks, cookies, etc.), having a very large range of LI (0.5−43.2%) and their SECV was 1.1, 5 times higher than ours. Moreover, the comparison of SECV with SEL is helpful: SECV of PR and LI predicted from ground grain spectra (0.35 and 0.19, respectively) was about double of SEL and were in the same order as numerous studies on wheat (SECV for PR ranging from 0.18 to 0.34, *28*). This appears very satisfactory for simple assessment of these traits in view of evaluation without any manipulations except grinding. SECV increased to 0.62 and 0.33 for PR and LI, respectively, when predicted from whole grains, which could be sufficient for rapid screenings of numerous samples that could be sown in addition after analysis.

Physical traits did not register as good performance with a RPD of 2.6 for ground grain spectra and 2.1 for whole grain spectra (**Tables 2** and **3**). Rami (*9*) obtained a RPD of 1.5 for HD predicted from sorghum ground grain spectra. HD has been however mainly studied on wheat (*24*, *26*, *29*) for which higher RPD was found. Our model could be improved by increasing the number of reference determinations for HD (117 reference values) but not for ET for which all 278 samples were analyzed. A better calibration was obtained for ET when absorbance was not corrected from light scattering. ET is indeed a visual character due to light scattering on individual particles and/or on air bubbles. It is therefore consistent that the light scattering phenomenon takes place in the model. On the other hand, it was surprising that the same figure was not observed for HD, which represents the particle size of the flour: the model was better including SNVD correction for both whole and ground grain spectra.

The lowest RPD was obtained for AM (lower than 2), although all the 278 samples were analyzed and 88 duplicated. Working on a mapping population, Rami (*9*) obtained the same RPD (1.82) with a higher SECV (0.99). AM appears indeed difficult to analyze by NIRS. Delwiche et al. (*30*) and Wu et al. (*31*), even using larger populations and AM ranges that comprise amylose mutants, did not get a RPD higher than 3.1 for milled rice. Sohn et al. (*32*) obtained a RPD of 5.3 but with a rice population presenting a SD value of 5.4 and samples with intermediate AM. Their SECV (1.02) was actually higher than ours (0.77). Considering the SECV, our model appeared indeed quite efficient: SECV was only 1.5 higher than SEL (*33*). The drawback of this model is the small AM range (15−24%) of applicability. The three mutants with no amylose were thus predicted with AM between 19.3 and 21.8%. When the calibration was performed without excluding the mutants, a higher RPD was obtained (2.7) with only a small degradation

of the SECV (0.84). In this case, two of the three amylose mutants were correctly predicted (−0.59, 0.39, and 6.97). Delwiche et al. (*22*) developed a model with a SEP of 1.0 that can predict AM of waxy and non-waxy rice. However, such model used as many as 18 PLS factors, against five in our case, a sign of a possible overfitting. Calibration populations including waxy mutants always present a discontinuity without any sample in the 2−14% range. PLS was in this case inappropriate; therefore, Campbell et al. (*34*) using near-infrared transmittance spectroscopy on maize and Delwiche and Graybosch (*35*) using NIRS on wheat proposed to use a factorial discriminant analysis. They were able to correctly predict a large number of mutants with no amylose with few misclassifications. We obtained the largest natural diversity available by using a world core collection but failed to cover the whole range of AM that would be necessary for designing a better model. Although artificial calibrations are seldom possible (*8*), it might be necessary, to improve our model, to build artificial samples by mixing waxy and nonwaxy ground grains, or use progenies obtained by crossing contrasted varieties to cover a larger AM range (*31*). Such approaches should still be validated.

The use of whole grain to develop calibration equations has been studied for some cereals and legumes (*11*, *25*, *29*, *36*) in view of eliminating the tedious grinding step and developing a nondestructive measurement method. We obtained acceptable reproducibility of NIRS spectra acquisition on whole grain, even using spin cells, probably because of the small size and round shape of sorghum grains. This allows working with small amounts of grain which is often the case in breeding programs. However, in agreement with previous studies (*11*, *24*, *25*), whole grain models evidenced a lower accuracy than ground grain models for most analyzed traits, except amylose. This may be due to a higher scattering noise for whole grain spectra. The PCA first axis indeed explained a much higher proportion of variance for whole grains while this axis is generally attributed to scatter variability due to the particle size effect (*8*, *11*, *24*, *29*, *37*) even after spectrum pretreatments (*17*, *37*). The efficiency of calibration models from whole grain could be improved by increasing the number of accessions. When a larger volume of seed per sample is available, the use of a transport module with a larger scanning surface should probably improve NIRS acquisition and calibration performances (*36*).

The performances of the NIRS equations developed in our study for AM, PR, LI, ET, and HD were appreciated using SECV and RPD statistics. While the validation of these equations using independent samples remains to be done, they can already have an application in breeding and technology programs to compare varieties. The quality and potential of our NIRS equations is due to the use of a core collection of cultivated sorghums that covers a broad diversity of consumers' usages, and the accuracy of the developed models. ET, which is a very important sorghum food physical quality trait, was for the first time exploited by NIRS. Whole and ground grain models were developed in view of eliminating the tedious and time-consuming grinding step. PR had the best equation and can be used for quality control in breeding and food technology programs. ET, LI, and HD equations can be used for screening steps. AM was more difficult to model by NIRS, which suggests the use of an artificial reference with intermediate AM samples to improve accuracy and robustness of the model.

LITERATURE CITED

(1) Fliedel, G.; Sabumukama, C.; Songre, L.; Matencio, F. Evaluation of sorghum grain quality for traditional thick porridges. In *Proceedings of the West African Hybrid Sorghum and Pearl Millet Seed Workshop*, US Agency for International Development, 2002, 98, Niamey, Niger; pp 213−225.

(2) Aboubacar, A.; Hamaker, B. R. Physicochemical properties of flours that relate to sorghum couscous quality. *Cereal Chem.* **1999**, *76*, 308−313.

(3) Almeida-Dominguez, H. D.; Serna-Salvidar, S. O.; Rooney, L. W. Properties of new and commercial sorghum hybrids for use in alkaline-cooked foods. *Cereal Chem.* **1991**, *68*, 25−30.

(4) Anglani, C. Sorghum endosperm texture - a review. *Plant Foods Hum. Nutr.* **1998**, *52*, 67−76.

(5) Audilakshmi, S.; Stenhouse, J. W.; Reddy, T. P.; Prasad, M. V. R. Grain mould resistance and associated characters of sorghum genotypes. *Euphytica* **1999**, *107*, 91−103.

(6) Schober, T. J.; Messerschmidt, M.; Bean, S. R.; Park, S.-H; Arendt, E. K. Gluten-free bread from sorghum: quality differences among hybrids. *Cereal Chem.* **2005**, *82*, 394−404.

(7) Yetneberk, S.; de Kock, H. L.; Rooney L. W.; Taylor, J. R. N. Effects of sorghum cultivar on injera quality. *Cereal Chem.* **2004**, *81*, 314−321.

(8) Pasquini, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **2003**, *14*, 198−219.

(9) Rami, J. F. Etude des facteurs génétiques impliqués dans la qualité technologique du grain chez le maïs et le sorgho. Ph.D. Thesis, Université Paris XI Orsay, Paris, 1999.

(10) Fontaine, J.; Schirmer, B.; Horr, J. Near-infrared reflectance spectroscopy (NIRS) enables the fast and accurate prediction of essential amino acid contents. 2. Results for wheat, barley, corn, triticale, wheat, bran/middlings, rice bran, and sorghum. *J. Agric. Food Chem.* **2002**, *50*, 3902−3911.

(11) Hicks, C.; Tuinstra, M. R.; Pedersen, J. F.; Dowell, F. E.; Kofoid, K. D. Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. *Euphytica* **2002**, *127*, 31−40.

(12) Frankel, O. H.; Brown, A. H. D. Current plant genetic resources: a critical appraisal. In *Genetics, new frontiers*; Oxford & IBH Publishing Co.: New Delhi, India, 1984; Vol. 4, pp 1−11.

(13) Deu, M.; Rattunde, F.; Chantereau, J. A global view of genetic diversity in cultivated sorghums using a core collection. *Genome* **2006**, *49*, 168−180.

(14) Harlan, J. R.; de Wet, J. M. J. A simplified classification of cultivated sorghum. *Crop Sci.* **1972**, *12*, 172−176.

(15) Mestres, C.; Matencio, F.; Pons, B.; Yajid, M.; Fliedel, G. A rapid method for the determination of amylose content by using differential-scanning calorimetry. *Starch/Stärke* **1996**, *48*, 2−6.

(16) Fliedel G. Appraisal of sorghum quality for making tô. *Agric. Dév.* (special issue) **1995**, 34−42.

(17) 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.

(18) Sine, B. Evaluation d'une core collection de sorgho en conditions de stress hydrique pré-floral. M.S. Thesis, Université Cheikh Anta Diop, Dakar, 2003.

(19) Williams, P. C. Implementation of near-infrared technology. In *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd ed.; Williams, P. C., Norris, K. H., Eds.; American Association of Cereal Chemistry: St. Paul, MN, 2001; pp 145−169.

(20) Mestres, C.; Matencio, F. Biochemical basis of kernel milling characteristics and endosperm vitreousness of maize. *J. Cereal Sci.* **1996**, *24*, 283−290.

(21) Williams, P. C.; Norris, K. H. *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd ed.; American Association of Cereal Chemistry: St. Paul, MN, 2001; pp 1−296.

(22) Delwiche, S. R.; Bean, M. M.; Miller, R. E.; Webb, B. D.; Williams, P. C. Apparent amylose content of milled rice by near-infrared reflectance spectrophotometry. *Cereal Chem.* **1995**, *72*, 182−187.

(23) Waniska, R. D.; Rooney, L. W. Structure and chemistry of the sorghum caryopsis. In *Sorghum: origin, history, technology, and production*; Smith, C. W., Frederiksen, R. A., Eds.; John Wiley & Sons: New York, 2000; pp 649−688.
(24) Williams, P. C.; Sobering, D. C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *JNIRS* **1993**, *1*, 25−32.
(25) Flinn, P. C.; Black, R. G.; Iyer, L.; Brouwer, J. B.; Meares, C. Estimating the food processing characteristics of pulses by near infrared spectroscopy, using meal or whole samples. *JNIRS* **1998**, *6*, 213−220.
(26) Garnsworthy, P. C; Wiseman, J.; Fegeros, K. Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy. *J. Agric. Sci.* **2000**, *135*, 409−417.
(27) Vines, L. L.; Kays, S. E.; Koehler, P. E. Near-infrared reflectance model for the rapid prediction of total fat in cereal foods. *J. Agric. Food Chem.* **2005**, *53*, 1550−1555.
(28) Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Application in Food and Beverage Analysis*, 2nd ed.; Longman Scientific & Technical; Harlow, UK, 1993; pp 1−227.
(29) Williams, P. Observations on the use, in prediction of functionality in cereals, of weights derived during development of partial least squares regression. *JNIRS* **1996**, *4*, 175−187.
(30) Delwiche, S. R.; Mckenzie, K. S.; Webb, B. D. Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chem.* **1996**, *73*, 257−263.
(31) Wu, J. G.; Shi, C. H. Prediction of grain weight, brown rice weight and amylose content in single rice grains using near-infrared reflectance spectroscopy. *Field Crops Res.* **2004**, *87*, 13−21.
(32) Sohn, M.; Barton, F. E.; McClung, A. M.; Champagne, E. T. Near-infrared spectroscopy for determination of protein and amylose in rice flour through use of derivates. *Cereal Chem.* **2004**, *81*, 341−344.
(33) van Barneveld, R. J.; Nuttall, J. D.; Flinn, P. C.; Osborne, B. G. Near infrared reflectance measurement of the digestible energy content of cereals for growing pigs. *JNIRS* **1999**, *7*, 1−7.
(34) Campbell, M. R.; Brumm, T. J.; Glover, D. V. Whole grain amylose analysis in maize using near-infrared transmittance spectroscopy. *Cereal Chem.* **1997**, *74*, 300−303.
(35) Delwiche, S. R.; Graybosch, R. A. Identification of waxy wheat by Near-infrared Reflectance Spectroscopy. *J. Cereal Sci.* **2002**, *35*, 29−38.
(36) Pérez-Marin, D. C.; Garrido-Varo, A.; Guerrero-Guinel, J. E.; Gome-Cabrera, A. Near-infrared reflectance spectroscopy (NIRS) for the mandatory labelling of compound feedingstuffs: chemical composition and open-declaration. *Animal Feed Sci. Technol.* **2003**, *116*, 333−349.
(37) Stuth, J.; Jama, A.; Tolleson, D. Direct and indirect means of predicting forage quality through near infrared spectroscopy. *Field Crop Res.* **2003**, *84*, 45−56.

Received for review April 14, 2006. Revised manuscript received August 21, 2006. Accepted August 24, 2006.

JF061054G