Journal of Cereal Science 69 (2016) 218–227

Contents lists available at ScienceDirect

Journal of Cereal Science

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

Pericarp thickness of sorghum whole grain is accurately predicted by NIRS and can affect the prediction of other grain quality parameters

Diarah Guindo [a, b], Fabrice Davrieux [c], Niaba Teme [a], Michel Vaksmann [b], Mohamed Doumbia [a], Geneviève Fliedel [c], Denis Bastianelli [d], Jean-Luc Verdeil [b], Christian Mestres [c], Mamoutou Kouressy [a], Brigitte Courtois [b], Jean-Francois Rami [b, *]

[a] Institut d'Economie Rurale (IER), Laboratoire Sol Eau Plante de Sotuba, P.O. Box 262, Bamako, Mali
[b] CIRAD, UMR AGAP, F-34398, Montpellier, France
[c] CIRAD, UMR QUALISUD, F-34398, Montpellier, France
[d] CIRAD, UMR SELMET, F-34398, Montpellier, France

ARTICLE INFO

Article history:
Received 19 October 2015
Received in revised form
25 February 2016
Accepted 8 March 2016
Available online 21 March 2016

Keywords:
Sorghum grain
Pericarp
NIRS

ABSTRACT

The thickness of grain pericarp, the outer layer of the kernel, is an important breeding criterion for sorghum. This cereal is mainly used through traditional processing in family-based food systems in many regions of the world. We investigated in this study how pericarp thickness could be predicted by Near Infrared Reflectance Spectroscopy (NIRS), a fast and non-destructive measurement method that is commonly used to measure physico-chemical parameters of sorghum grains, and how this trait also influences the prediction of those parameters. We showed that, using a classification approach, it was possible to discriminate thick from thin pericarp whole grain samples with a good accuracy and that the proportion of thin and thick grains in mixed samples could also be predicted. In addition, pericarp thickness had a significant effect on the calibration performance for other grain parameters indicating that the pericarp can distort spectral information of whole grain samples. As a practical consequence, we suggest to develop separate whole grain calibration models for thin and thick pericarp samples, combined with a two-steps prediction approach to improve the accuracy of whole grain NIRS calibrations for grain quality parameters in sorghum.

1. Introduction

Sorghum (*Sorghum bicolor* (L.) Moench) is the fifth most important cereal worldwide, after maize, rice, wheat and barley (FAO, 2014). It is particularly adapted to a wide range of environmental conditions including drylands because of its moderate water requirements and its large genetic diversity (Chantereau et al., 2013; Dendy, 1995; Smith and Frederiksen, 2000). Sorghum grain and stover are used in various ways as food or feed in different geographic areas and countries, at both traditional and industrial levels. In Africa and some regions of Asia and India, traditional uses of sorghum grain are mainly for human consumption. It is traditionally processed into thick (tô, bogobe, ugali) or thin (ogi, motogo) porridges, granulated products (couscous, dégué), flat breads (tortilla, kisra, roti) or beverages (beer, gowé, mahewu) (Anglani, 1998; Kleih et al., 2000; Smith and Frederiksen, 2000).

Most of the food products made from sorghum grain require removing the pericarp, either manually or mechanically, a first processing step known as dehulling. Traditional dehulling is performed by pounding washed grains with a wooden mortar and pestle. During pounding, water is added to soften the pericarp and facilitate its removal (Scheuring et al., 1983). Mechanical dehulling is most often a dry process and abrasive-type dehullers are the most used in Africa for sorghum and millet grain (Reichert, 1982). A good dehulling is defined by a complete removal of the pericarp, of the testa layer if present, and of much of the germ, providing high endosperm recovery with minimum breakage of the endosperm (Anglani, 1998; Bello et al., 1990; Fliedel, 1995; Reichert, 1982).

The thickness of the pericarp is an important factor in both traditional and industrial dehulling. When associated with hard endosperm, a thick pericarp is much easier to remove by pounding which is a common practice in Africa; this is one of the reasons why

* Corresponding author.
E-mail address: rami@cirad.fr (J.-F. Rami).

http://dx.doi.org/10.1016/j.jcs.2016.03.008
0733-5210/

in some case farmers prefer to grow this type of grain in Africa (Murty and Kumar, 1995). Conversely, thin pericarp varieties require two times longer to be traditionally dehulled but they are more suitable for dry mechanical dehulling whatever the nature of the endosperm (Earp et al., 2004; Gomez et al., 1997; Scheuring et al., 1983). However, sorghum varieties with a thick pericarp have some disadvantages such as sensitivity to mold, grain weathering in the field during maturation, and rapid deterioration of grains during storage. Varieties with thin pericarp are usually considered to be more tolerant to these constraints (Glueck and Rooney, 1980). The thickness of the pericarp should thus be an important breeding criterion, and a simple and fast measurement method is needed for analyzing the large number of samples usually handled by breeding programs.

The pericarp is the main envelop covering the kernel. It is subdivided into three layers: the epicarp, which is the outermost layer of the pericarp, the mesocarp, and the endocarp, which is the innermost layer, constituted of cross and tube cell layers transporting moisture throughout the kernel. The thickness of the pericarp depends on the thickness of the mesocarp which refers to the number of layers and the presence of starch granules in the mesocarp (Earp et al., 2004; Hoseney et al., 1974; Rooney et al., 1981; Rooney and Murty, 1982). Mesocarp thickness is genetically controlled by a major gene named Z. Sorghum varieties with thin pericarp carry the dominant allele (Z), while sorghum varieties with thick pericarp are homozygous for the recessive allele (z) (Ayyangar et al., 1934).

Two main methods of pericarp measurement have been reported. The first method uses electronic microscopy. Grains are scanned and the amount of starch granules in the mesocarp is investigated to determine pericarp thickness. Three classes (thin, thick or very thick) are generally distinguished by this method which is precise but complex and extremely time consuming (Earp et al., 2004; Scheuring et al., 1983) and thus not adapted to breeding. The second method is a visual appraisal commonly used by breeders. Grains are scraped with a scalpel and pericarp thickness is observed using a magnifying glass (Gomez et al., 1997). This method identify two classes of thickness (thin and thick) and is simple and fast but more error-prone.

Near Infrared Reflectance Spectroscopy (NIRS) is an indirect method that can be applied to a large number of samples and allows measuring many grain parameters through a single assay. As a preliminary step, spectral data of a reference set of samples characterized for their biochemical or physical properties measured through standard methods are used to develop a predictive model. This calibration model can then be routinely applied to the spectra of new samples. Whole grain calibrations are more appropriate for breeding applications because they do not require any grinding step and are thus faster and nondestructive.

The use of NIRS on sorghum to assess grain quality has been reported, in particular to measure starch content, amylose content, protein content, lipid content, total phenols content, condensed tannin content, grain hardness and endosperm texture (de Alencar Figueiredo et al., 2006; Davrieux et al., 2007; Dykes et al., 2014; Hicks et al., 2002; Hooks et al., 2006; Rami, 1999). These studies were conducted on ground grains, on whole grain, or on both. Calibrations based on ground grains generally performed slightly better than calibrations on whole grain (de Alencar Figueiredo et al., 2006). In all conditions, the most accurate predictions were obtained for protein content, followed by grain hardness or lipid content, while amylose content showed consistently poor calibration performances (de Alencar Figueiredo et al., 2006). However, to our knowledge, no study reported so far the use of NIRS to predict pericarp thickness of sorghum grain. Considering the biochemical nature and the simple genetic determinism of pericarp thickness, it is likely that this trait might be predictable using NIRS. Furthermore, as the Z gene can segregate in breeding populations, bulk of grains harvested in breeding programs may contain a mixture of grains with thin and thick pericarp in varying proportions. It is thus of interest to identify samples showing mixture of thick and thin pericarp grains and even quantify the proportion of each type of grain. Finally, in the case of NIRS calibrations established on whole grains, the pericarp, as the peripheral part of the grain, can possibly alter spectral acquisition of endosperm. It should be useful to know if the endosperm components of a thick pericarp variety could be still accurately represented by the NIRS spectrum of whole grains and consequently, if pericarp thickness could affect the performance of whole grain calibrations for endosperm parameters.

The objective of this study was to assess how pericarp thickness could be accurately predicted by NIRS, both qualitatively and quantitatively and to investigate the effect of pericarp thickness on non-destructive prediction of several biochemical and physical grain parameters by NIRS.

2. Materials and methods

2.1. Plant material

Two sets of material were used for this study: a core collection and a breeding population. The core collection was developed by CIRAD (Deu et al., 2006) and consisted of 278 accessions belonging to the five basic races of cultivated sorghum and five intermediate races. The accessions originated from 39 countries representing sorghum production areas. The methods used to produce seed samples were described by de Alencar Figueiredo et al. (2006).

The breeding population was derived from a cross between two sorghum inbred lines named Lata3 and Tiandougou. Lata3 is a guinea type with a thin pericarp, vitreous endosperm, small size, and translucent grains, while Tiandougou is a caudatum type with a thick pericarp, softer endosperm, large size, and chalky grains. The breeding population included 404 progenies in F_3 generation. The whole population was grown in the field in 2010 in Mali at Sotuba IER research station. Each plot included two rows of 10 F_5 plants obtained through bulk multiplication of F_3 progenies ($F_{3:5}$ families). A total of 403 progenies and 60 replications of the two parents grain samples were harvested and used in this study.

2.2. Sample preparation

Grain samples were cleaned by hand by removing stones, straw, moldy, broken, or insect-damaged kernels, and dust. Flour samples were produced for all the accessions of the core collection and for a random subset of 139 individuals of the breeding population. About 20 g of cleaned grains were ground using a Perten Mill 3100 with 0.8 mm sieve (Laboratory Mill 3100, CEMOTEC 1090, Tecator).

2.3. NIRS instrumentation and measurement

A monochromator Foss NIRS instrument (NIRS 6500) was used to scan whole grain and ground grain samples. A quartz ring cup of 47 mm outside diameter and 36 mm inside diameter was filled with 5 g of whole grains or with 3 g of ground grain. The values of reflectance from 1100 to 2500 nm at 2 nm intervals were collected as log (1/R). A second spectrum was acquired for each sample after refilling the cup. The mean of the two spectra was used for further calculation.

2.4. Pericarp thickness measurement

Pericarp thickness was evaluated on all samples of the breeding

population and on 264 samples of the core collection by using the visual method described by Gomez et al. (1997). Ten kernels were scraped with a scalpel to remove the pericarp. The scraps were observed using a magnifying glass. A thick pericarp comes off in thin flakes, while a thin one usually scrapes off in small fragments or as a powder.

In the breeding population, a sample was considered as segregating for the Z gene when both thick and thin pericarp grains were observed. These samples were designated as "mixed".

2.5. Spectral data analysis

2.5.1. Spectrum pre-treatment

All spectra data underwent scatter correction (SNV and Detrend) and mathematic treatments using the software WinIsi, version 4 (FOSS NIRSystems Inc.). The standard normal variate (SNV) scaled each spectrum to reach a standard deviation of 1.0 that helps reducing particle size effects, and Detrend removed the linear and quadratic curvature of each spectrum. Mathematical pre-treatments were applied to the spectra: smoothing and second order derivation, both calculated over 5 datapoints.

2.5.2. Correction for sample humidity

Grain moisture content of the core collection samples was adjusted to 11.5% by de Alencar Figueiredo et al. (de Alencar Figueiredo et al., 2006). Such conditioning of grains was not conducted on the breeding population samples. As the mean moisture content of the breeding population (8%) was lower than that of the core collection, the main absorption bands corresponding to water (1100–1200 nm, 1350–1550 nm, and 1850–2100 nm) were removed from the spectra in all analyses to reduce any possible bias related to moisture differences of the core collection and breeding population samples sets.

2.5.3. Principal component analysis

A principal component analysis (PCA) was performed using pre-treated spectra of the combination of the two sets of material (core collection and breeding population) without the "mixed" class of the breeding population that was removed to highlight a potential structure of samples based on pericarp thickness. The generalized Mahalanobis distance ($\tilde{H}$) was computed for each spectrum. All samples having $\tilde{H}$ value above 3 were considered as *h*-outliers and removed from the dataset. PCAs were performed using R and the "dudi.pca" function of the "ade4" package (Dray and Dufour, 2007). Score plots representing pericarp thickness classes of individuals were built for whole grain and ground grain spectral data using the "s.class" function of the "ade4" package.

2.6. Calibration of pericarp thickness

2.6.1. Discrimination between thin and thick pericarp

A classification approach was first conducted using Partial Least Squares Discriminant Analysis (PLS-DA) to classify whole grain samples based on the thickness of the pericarp. For this analysis, all whole grain samples of the core collection and the breeding population were used, except the ones previously identified as *h*-outliers and the mixed samples of the breeding population. The thickness of the pericarp was considered as the factor to be predicted with two levels, thin and thick. A first complete model (model 1) was developed on grain samples using all whole grain samples available. In order to validate this model, a training and a validation dataset, representing respectively 75% and 25% of the full dataset were randomly sampled with the constraint to balance thick and thin pericarp classes within the two sets. A training model (model 2) was built on the training dataset and used to predict the validation dataset.

A PLS-DA model was also fitted to the ground samples dataset. For this analysis, all ground samples of the core collection and the breeding population were used, except the ones previously identified as *h*-outliers and the mixed samples of the breeding population. A training and a validation datasets were defined following the same strategy as for whole grain and a PLS-DA model (model 3) was built on the training dataset and used to predict the validation dataset. A fourth model was built using the exact same whole grain samples as the ones used for the ground grain model and using the same training and validation subsets. This model (model 4) was constructed as a reference point to compare prediction accuracy between whole and ground grain models.

The "caret" R package (Kuhn, 2008) was used to build all models. The number of components in each model was determined using the results of a k-fold cross-validation with 10 folds and 10 repeats. The Receiver Operating Characteristic (ROC) metric was used as a measure of performance of the model and the number of components that maximized ROC was kept in the model (Kuhn, 2015). The quality of the chosen models was assessed through the accuracy computed from the confusion matrix as obtained through prediction of the validation datasets.

2.6.2. Classification and quantitative prediction of mixed samples in the breeding population

Mixed samples in the breeding population were observed when a given individual was heterozygous for the Z gene at the F_3 generation. As the grain samples were harvested from a bulk of $F_{3:5}$ plants, a heterozygous F_3 plant yielded a mixture of grains with thin and thick pericarp. As the Z allele (thin) is dominant, the expected proportions of thin (Z-) and thick (zz) pericarp in such a mix are $\frac{3}{4}$ and $\frac{1}{4}$, respectively. As the bulk was constructed from a limited number of plants (10), a large variance was also anticipated around the theoretical proportions.

Two approaches were developed to predict the class of mixed samples in the breeding population. In the first approach, a PLS-DA model was fitted to the whole grain samples of the breeding population including the three classes of pericarp observed in the population: thin, thick, and mixed. A training and a validation datasets were randomly sampled following the same procedure as previously described. The optimal number of components of the model was selected using the results of a k-fold cross-validation with 10 folds and 10 repeats and using the accuracy metric. The quality of the final model was assessed through the accuracy computed from the confusion matrix as obtained through prediction of the validation dataset.

In the second approach, a quantitative PLS calibration was developed using the same combined dataset of thin and thick pericarp samples from the core collection and the breeding population with thin pericarp being coded as 1 and thick pericarp as 0. The model was fitted using the "plsr" function of the "pls" R package (Mevik and Wehrens, 2007). An iterative process aiming at removing possible outliers from the model was implemented: at each iteration, a PLS model was fitted using all samples and the number of components was determined as the one minimizing the Root Mean Square Error of Cross Validation (RMSECV). Student (t) test was then used to identify t-outlier samples. Outlier detection was based on the standardized residuals with a cutoff of 2.5. Three passes of t-outliers elimination were conducted. The breeding population samples used in the first approach were then predicted using this quantitative model. The distribution of predicted values was compared to the classes obtained using the first approach. In order to validate the quantitative prediction, 30

mixed samples were randomly sampled, and 40 kernels per sample were examined to quantify the proportion of thin and thick pericarp in each mixture. The correlation between this value and the predicted value was reported. Finally, predicted values were clustered in three groups using the R implementation of the fuzzy c-means algorithm (Bezdek, 1981) with initial values for cluster centers being 0, 0.75, and 1 to convert quantitative values into three putative classes being thick, mixed, and thin respectively.

2.7. Validation of pericarp measurement and prediction using microscopy

Fifteen samples from the breeding populations were selected in order to have balanced classes of pericarp thickness and a representative range of NIRS predictions as obtained from the quantitative model previously described (Supplementary Table 2). Grains were soaked overnight in a 50% water-glycerol solution, then glued on a vibratome plate and cut into 20 μm thick transverse sections in the middle of the grain, using an Hm650v vibratome (Thermo Scientific Microm, speed 18 mm/s, frequency 70 Hz, amplitude 0.6 mm). The grains used for microscopy were different from the ones used for the visual measure. Five grains were used for the samples previously noted as thin or thick by the visual method, while 10 grains were used for those noted as mixed. Several sections of each grain were mounted on a microscope slide and examined using a Leica DM 4500 B light microscope (Leica, Germany, objectives Leica 20×). Four images from 4 different sections were taken for each grain. On each image, the structure of the mesocarp was used to visually qualify the grain as having a thin or a thick pericarp, and 4 measures of the thickness of the pericarp were performed from the outer border of the aleurone layer to the outer border of the epicarp cell layer and converted to μm (1 pixel = 0.3655 μm) using the ImageJ software (Fig. 4a and b).

2.8. Effect of pericarp thickness on the calibration of other parameters

In a previous study, de Alencar Figueiredo et al. (2006) developed NIRS calibrations using the core collection for 5 grain parameters: endosperm texture (ET), grain hardness (HD), amylose content (AM), protein content (PR), and lipid content (LI). Using the classification model 1 that we developed for pericarp thickness, all the samples of the core collection were predicted as having a thick or a thin pericarp. A sampling procedure has been iterated 100 times for each parameter. At each iteration, a training and a validation dataset, representing respectively 75% and 25% of the number of samples having reference values, were randomly sampled with the constraint to have balanced numbers of thin and thick pericarp samples in both sets. A prediction model was developed for each training dataset and tested using the corresponding validation set separately on thin and thick pericarp samples. The standard error of prediction (SEP), bias, standard error of prediction corrected of the bias (SEPc), and the slope were computed for each validation. The average values of these 4 parameters, over the 100 random samples, were compared between thin and thick samples using a student *t*-test. SEP was calculated as $\sqrt{\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{n}}$, the bias as $\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)}{n}$, and SEPc as $\sqrt{\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i-\text{bias})^2}{n}}$, where n is the number of samples in the validation set, y_i is the measured reference analysis value and $\widehat{y}_i$ is the value predicted by the training model. All calibration models used the PLS regression method and were fitted using the "plsr" function of the "pls" package (Mevik and Wehrens, 2007) with the R software.

3. Results

3.1. Pericarp thickness measurement

The relative proportions of thin and thick pericarp samples in the core collection and in the breeding population are reported in Supplementary Table 1. In the core collection, the proportion of accessions with thick pericarp grains (62%) was almost twice that of accessions with thin pericarp grains (33%). This indicates that thick pericarp is a predominant trait in *Sorghum bicolor* since the core collection is a representative sample set of the diversity of sorghum.

The breeding population included, as expected, a large number of segregating families having mixed pericarp thickness in addition to families belonging to the thin and thick classes. The numbers of individuals within thick, mixed and thin classes were 133, 144, and 126 respectively. These figures significantly deviated from the expected proportions of $\frac{3}{8}:\frac{1}{4}:\frac{3}{8}$ for a F_3 generation with a highly significant Chi^2 test ($P = 3.9 \times 10^{-6}$).

3.2. Principal component analysis (PCA) on NIRS spectra

A first PCA was performed on the NIRS spectra of whole grains of the joint dataset (core collection and breeding population) excluding the mixed class of the breeding population. The computation of generalized Mahalanobis distance ($\tilde{H}$) for each spectrum allowed identification of 3 outliers having a $\tilde{H}$ value above 3. These 3 samples were removed from the dataset. The first 15 principal components explained 99.1% of the variance of whole grain NIRS spectra, with the two first axes accounting for 52.8% and 26%, respectively. The score plot of all spectra for the first two principal components (Fig. 1a) displays a strong structuration between classes of pericarp thickness on the first principal component, and between core collection and breeding population groups of samples on the second principal component.

A second PCA was performed on the NIRS spectra obtained from ground samples. This included all the samples of the core collection and a subset of the breeding population (Supplementary Table 1) from which mixed samples were also removed. The computation of generalized Mahalanobis distance ($\tilde{H}$) for each spectrum allowed identification of 2 outliers having a $\tilde{H}$ value above 3. These 2 samples were removed from the ground samples dataset. The first 15 principal components explained 95.6% of the variance of ground grain NIRS spectra, with the two first axes accounting for 38.7% and 21.9%, respectively. The strong structure due to the thickness of the pericarp mainly explained by the first component on Fig. 1a disappeared on the score plot of ground samples (Fig. 1b). The first principal component was mainly explained by the differentiation between core collection and breeding population when considering ground samples.

In order to assess if the breeding population could be predicted accurately by a calibration developed using the core collection, a PCA was performed on the core collection only and the spectra of the breeding population were projected as supplementary observations. The results showed that for both whole grain and ground grain spectra, the range of spectral variation observed in the breeding population was fairly represented by the core collection (Fig. 1c and d).

3.3. Pericarp thickness calibration

3.3.1. Discrimination between thin and thick pericarp classes

A Partial Least Squares Discriminant Analysis (PLS-DA) was conducted to discriminate thin and thick pericarp based on NIRS spectra. A first complete model was constructed using the full set of

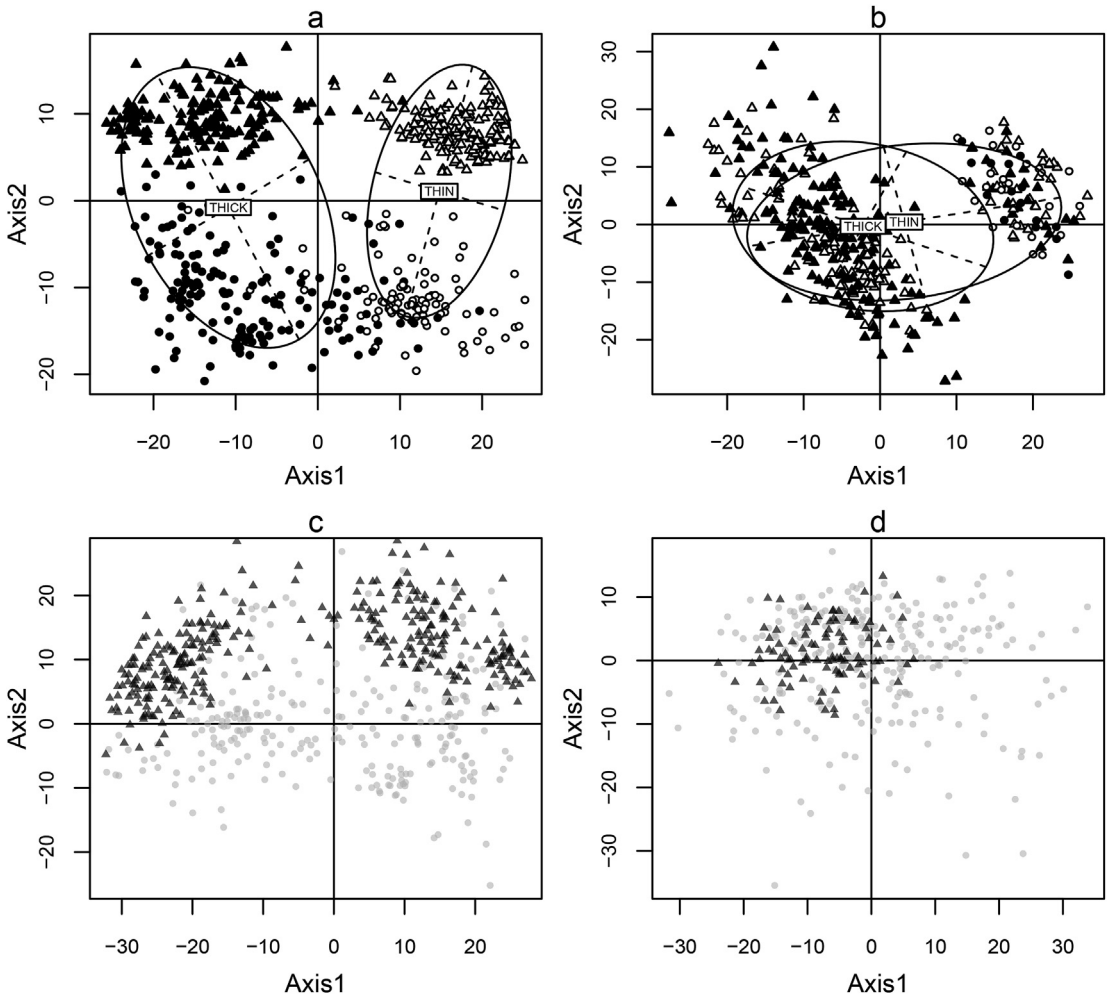

Fig. 1. a. Score plot of core collection + breeding population samples for the first two principal components from whole grain samples spectra. circle: core collection; triangle: breeding population; black: Thick pericarp; white: Thin pericarp. b. Score plot of core collection + breeding population samples for the first two principal components from ground samples spectra. circle: core collection; triangle: breeding population; black: Thick pericarp; white: Thin pericarp. c. Score plot of core collection samples for the first two principal components from whole grain samples spectra, with breeding population samples projected as supplementary observations. circle: core collection; triangle: breeding population. d. Score plot of core collection samples for the first two principal components from ground samples spectra, with breeding population samples projected as supplementary observations. circle: core collection; triangle: breeding population.

whole grain samples (Model 1). An optimal number of 6 components was kept (Supplementary Fig. 1a) and the confusion matrix provided a prediction accuracy of 95.5% (Table 1).

In order to validate this model, a training model (Model 2) was fitted to a random training dataset representing 75% of the full dataset. The cross-validation using the ROC metric (Supplementary Fig. 1b) provided an optimal number of 6 components to be kept in Model 2. The confusion matrix, using this model, computed on the validation dataset provided a prediction accuracy of 95.1% (Table 1).

The same analysis was conducted using the ground samples dataset (Model 3) and compared to a whole grain dataset comprising the exact same samples as the ones constituting the ground samples dataset (Model 4). An optimal number of 9 components was obtained by maximizing ROC metric (Supplementary Fig. 1c) for the model using ground samples, while 5 components where kept for the counterpart whole grain model (Supplementary Fig. 1d). The confusion matrix, using the ground samples model computed on the validation dataset provided a prediction accuracy of 73.6% while the equivalent whole grain model provided a prediction accuracy of 96.6% (Table 1).

Table 1
Confusion matrix for PLS-DA Models 1 to 6. Model 1: whole grain full model, Model 2: whole grain validation model, Model 3: ground samples validation model, Model 4: subset of whole grain samples corresponding to the ones used for the ground samples model, Model 5: whole grain validation model with mixed class, Model 6: whole grain full model with mixed class.

	Model 1		Model 2		Model 3		Model 4		Model 5			Model 6		
	Thick	Thin	Thick	Thin	Thick	Thin	Thick	Thin	Thin	Mixed	Thick	Thin	Mixed	Thick
Thick	314	6	78	2	43	15	49	1	32	11	1	136	36	1
Mixed	–	–	–	–	–	–	–	–	7	15	0	20	88	8
Thin	20	240	5	59	8	21	2	35	0	10	39	0	20	154
Accuracy	95.50%		95.10%		73.60%		96.60%		74.80%			81.60%		

3.3.2. Classification and quantitative prediction of mixed samples

To assess if the class of mixed samples of the breeding population could be classified as the two other classes of thin and thick pericarp samples, a PLS-DA model was fitted to the whole grain samples of the breeding population including the three classes of pericarp observed in the population: thin, thick, and mixed. A first model (Model 5) was fitted to the training dataset. An optimal number of 6 components was obtained using the cross-validation accuracy. The confusion matrix computed on the validation dataset using Model 5 provided a prediction accuracy of 74.8%. A final model was fitted using the whole dataset (Model 6). The confusion matrix computed on the full dataset using Model 6 provided a prediction accuracy of 81.6% (Table 1). These results show that the class of mixed samples was not accurately distinguished from the two other classes using the classification approach.

An alternative quantitative approach was developed to predict pericarp thickness of mixed samples. For this purpose, a PLS calibration model was constructed using the combined dataset of thin and thick pericarp samples from the core collection and the breeding population with the thin class being coded as 1 and the thick class being coded as 0. A total number of 15 components was kept in the model and 48 *t*-outliers (37 from the core collection and 11 from the breeding population) were eliminated during the calibration procedure. The final model showed a R^2 value of 0.95 and a RPD value of 3.67.

This model was used to predict all samples of the breeding

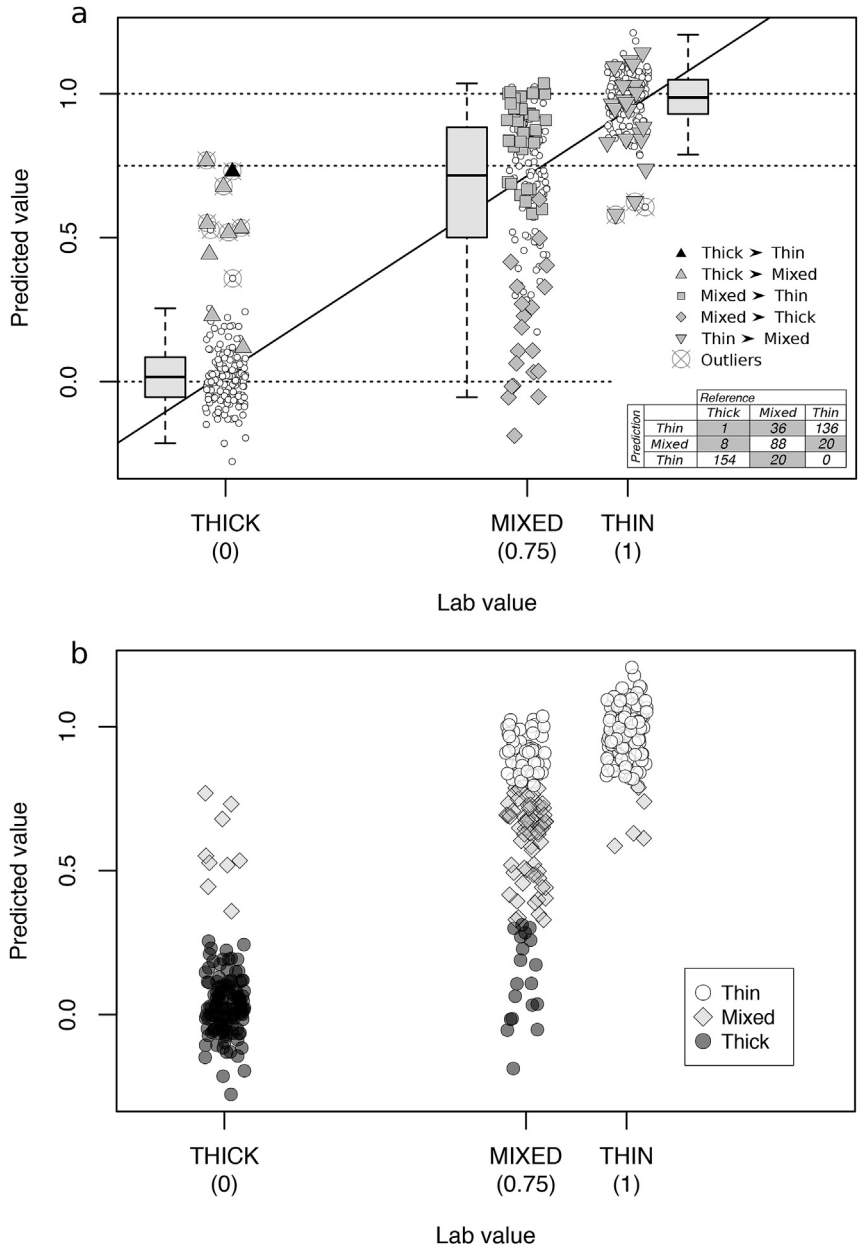

Fig. 2. a. Combined results of classification and quantitative prediction for the breeding population samples. The scatter plot represent the relationship between observed Lab value (0, 0.75 or 1) and predicted value. The x values have been jittered along x-axis to reduce overlaps and improve readability. A boxplot has been added next to each group of dots to illustrate the distribution of predicted value for each class. The stars represent outliers detected during the PLS model construction. The table is the confusion matrix as obtained by the PLS-DA classification. b. Comparison of pericarp classes obtained through fuzzy c-means clustering of predicted values using quantitative model and initial observed (Lab) classes of pericarp thickness.

population including the class of mixed samples. Fig. 2a illustrates the relationship between observed (Lab) values and the quantitative predictions by the model. The median of predicted values for the class of mixed samples was 0.72, which is close to the value of 0.75 expected in a quantitative framework of the inheritance of pericarp thickness considering that the Z (thin) allele is dominant. The predicted value of mixed samples ranged from −0.19 to 1.04.

Fig. 2a also reports the correspondence between the confusion matrix of the classification PLS-DA full model (Table 1) and the quantitative prediction of the breeding population. Among misclassified samples, 6 out of 8 samples that were measured as thick and classified as mixed corresponded to outliers in the PLS model. Symmetrically, 2 out of 20 samples that were measured as thin and classified as mixed corresponded to outliers in the PLS model. The predicted value of 36 mixed samples that were classified as thin by the PLS-DA model ranged from 0.58 to 1.04 with a mean value of 0.86. The predicted value of 20 mixed samples that were classified as thick by the PLS-DA model ranged from −0.19 to 0.63 with a mean value of 0.18.

To verify whether the quantitative prediction of mixed samples through a model developed from the non-mixed samples was accurate, the proportion of thin and thick pericarp grains was measured on a validation set of 30 mixed samples of 40 grains. The relationship between the measured and predicted proportion is illustrated by Fig. 3. A R^2 value of 0.89 was obtained with no significant bias. The SEP of the prediction was 0.1 and the corresponding RPD was 2.99. C-means clustering of predicted values using initial values of cluster centers of 0, 0.75, and 1 led to cluster relative sizes of 35.5%, 20.1%, and 44.4% respectively, which were closer to the expected segregation ratios of the Z gene (37.5%, 25.0%, and 37.5%) than what we observed using visual measurements. Fig. 2b illustrates the comparison between the clustering of predicted values with initial observed values. Samples predicted as mixed using this method included 9 samples initially observed as thick and 5 samples initially observed as thin. Conversely, among the samples initially observed as mixed, 19 were predicted as thick and 58 were predicted as thin. These misclassifications, combined with the over-representation of thin class in c-means based clusters, suggest that the distinction between mixed and thin classes remains challenging.

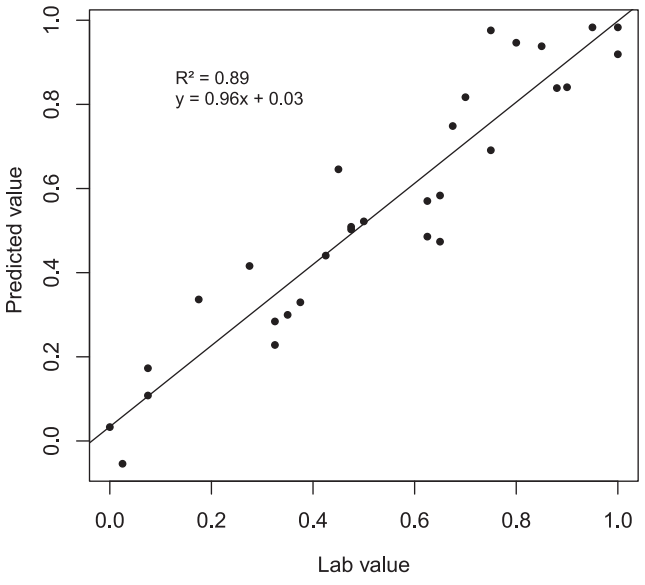

Fig. 3. Relationship between the proportion of thin pericarp sample observed in mixed grain samples and the proportion predicted by the quantitative model.

3.4. Validation of pericarp measurement and prediction using microscopy

Fifteen samples of the breeding population have been analyzed using microscopy to validate the visual measurement of pericarp thickness as well as the NIRS predictions. The thickness of the pericarp was on average 87 μm and 49 μm for the grains identified by microscopy as having thick and thin pericarp, respectively (Fig. 4, Supplementary Table 1). The standard deviation was much higher for thick (31 μm) than for thin (11 μm) pericarps. A discrepancy between visual and microscopy notations was observed for 3 samples (Supplementary Table 1, Fig. 4). In order to compare to the NIRS predictions, those were classified into “Thin”, “Mixed”, “Thick” and “Uncertain” based on c-means clustering with a membership threshold of 0.90. The sample 260 that was visually noted “Thick” was predicted “Mixed” (0.57) by NIRS and was found as having 50% of thick and thin pericarp grains by microscopy. Similarly, the sample 20, visually noted as “Mixed”, and predicted as “Thick” (−0.02) was identified as “Thick” by microscopy. Finally the sample 55 visually noted as “Thin” and predicted as “Uncertain” (0.8) was identified as “Mixed” using microscopy (60% of thin grains). The correlation between the percentage of thin grains as observed by microscopy and the NIRS prediction was 0.93.

3.5. Effect of pericarp thickness on the calibration of other grain parameters

We used the core collection and reference values reported by de Alencar Figueiredo et al. (2006) to assess whether pericarp thickness had an effect on the performance of whole grain calibrations for endosperm texture (ET), grain hardness (HD), amylose content (AM), protein content (PR), and lipid content (LI).

For each parameter, we randomly sampled from the whole core collection dataset, 100 training and 100 validation datasets having balanced numbers of thin and thick pericarp samples and representing 75% and 25% of the number of samples having reference laboratory data, respectively. Because the number of available reference laboratory data was different among parameters, the sizes of the training and validation datasets were different between parameters. It ranged from 66 (HD) to 158 (AM) (Table 2). For each parameter, 100 PLS models were developed using the different sampled training datasets and tested using the corresponding validation set, separately on thin and thick pericarp samples. The average of the Standard Error of Prediction (SEP) and Standard Error of Prediction corrected of bias (SEPc) over the 100 random validation sets were highly significantly lower for thin pericarp than for thick pericarp for all parameters except HD. The bias values were low for all parameters both for thin and thick pericarp (Table 2). Finally, the slope of the regression of predicted value on actual measured values was significantly lower for all parameters as reported on Table 2 and illustrated on Supplementary Fig. 2. These results indicated that using a model developed from a dataset including the two classes of pericarp, predictions were less accurate for thick pericarp samples than for thin pericarp samples.

4. Discussion

The pericarp thickness of sorghum grain is an important trait for breeding as it can interact with traditional and industrial transformation processes and may also influence the properties of the grains regarding sensitivity to mold, maturation, and post-harvest storage. Depending on the targeted end uses, a thick or a thin pericarp can be pursued by breeding programs in complement to other traits. The measurement method classically used is a visual appraisal that is time consuming and error prone. As NIRS is

commonly used for other sorghum grain parameters (de Alencar Figueiredo et al., 2006; Davrieux et al., 2007), pericarp thickness could be simultaneously predicted at no cost.

PCA showed that pericarp thickness was the first source of variation of NIR spectra of whole grain as opposed to what was found for ground samples. This suggests that NIR spectra of whole grain samples is firstly linked to grain structure (pericarp thickness) rather than to biochemical content. We reported an average pericarp thickness of 49 μm and 87 μm for thin and thick pericarp grains, respectively. It is thus likely that a significant part of the absorbance observed in thick pericarp whole grain samples is due to the pericarp. We indeed showed that using PLS based discriminant classification it was possible to distinguish thick from thin pericarp samples with a good accuracy from whole grain NIR spectra. The accuracy of the classification was much lower when applying the same procedure to ground samples, which confirmed the contribution of the absorbance at the surface of the whole grains was important.

Detecting grain samples having a mixture of thin and thick pericarp, which are often observed in breeding populations, is challenging and NIRS can provide a quantitative alternative. Indeed, the characterization of mixed samples through visual appraisal requires examining a large number of grains, which is in practice not feasible. In the breeding population that we have analyzed in this study, the visual measurement of pericarp thickness identified 36% of mixed samples which is significantly higher than the 25% of heterozygous individuals expected at the F_3 generation. We checked the segregation of the Z locus in the population using molecular markers (data not shown) and did not observe any segregation distortion, therefore supporting the hypothesis that the visual assessment was error prone. The calibration of pericarp thickness using non-mixed samples allowed to quantitatively predict the proportion of thin and thick grains in mixed samples. The high number of *t*-outliers from the breeding population observed in this calibration (Fig. 2a), suggested that these samples where mixed samples that were improperly assigned to thin or thick classes thus supporting the difficulty in identifying mixed samples visually. The microscopy validation confirmed the NIRS prediction for 3 samples incorrectly characterized by the visual method. Fuzzy c-means clustering of the quantitative values predicted by NIRS provided an alternative method to assign grain samples to one of the three classes. However, because the Z (thin)

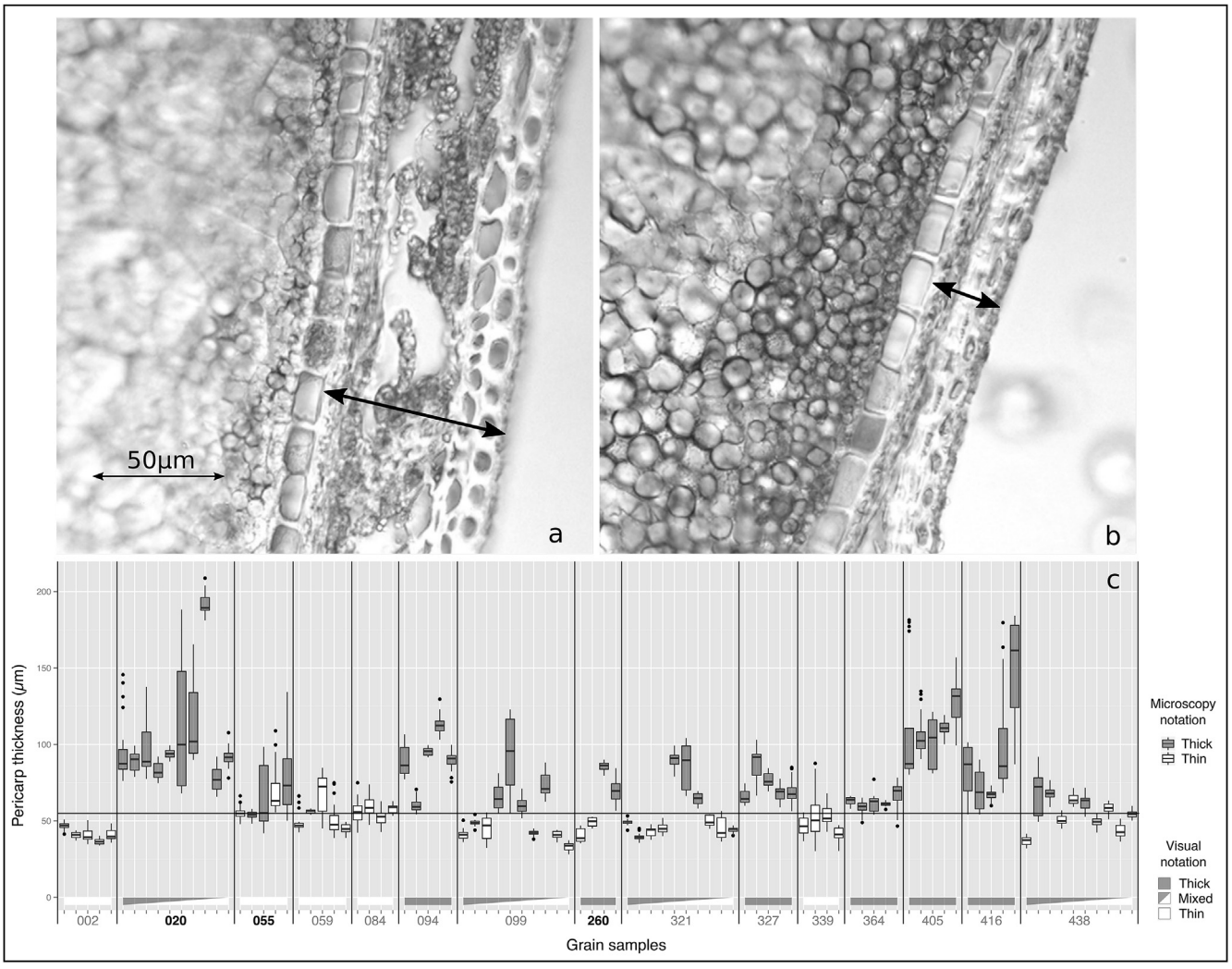

Fig. 4. a. Light microscopy image of a portion of a transverse section of a Thick pericarp grain. b. Light microscopy image of a portion of a transverse section of a Thick pericarp grain. The arrows represent the distance that as been measured by image analysis to estimate pericarp thickness. c. Validation of pericarp thickness by microscopy for 15 grain samples of the breeding population. Each box plot represents the distribution of pericarp thickness of one grain measured over several sections. The different samples are separated by a black vertical line. The color of the boxplot represents the thickness of the pericarp (thin or thick) as visually assessed by the examination of the microscopy sections. The rectangles at the bottom of the figure represent the thickness of the pericarp of each sample as previously determined by visual examination of the grain. The samples names highlighted in bold indicate the 3 samples for which a discrepancy was observed between visual determination and microscopy.

Table 2
Comparison of SEP, BIAS, SEPc and Slope values averaged over 100 random validation sets between thin and thick pericarp classes. Ntr: Number of sample in training sets. Nva: Number of samples in validation sets, Mean: average value of each parameter for thin and thick classes of pericarp.

	Ntr	Nva	Mean		SEP		BIAS		SEPc		Slope	
			Thin	Thick	Thin	Thick	Thin	Thick	Thin	Thick	Thin	Thick
PR	80	28	13.85	13.82	0.77	1.00***	−0.050	0.031	0.73	0.96***	1.00	0.72***
LI	70	24	3.77	3.87	0.36	0.45***	−0.022	0.015	0.35	0.42***	0.79	0.64***
AM	158	54	20.79	20.26	0.84	0.94***	0.046	0.000	0.82	0.91***	0.68	0.66^{ns}
ET	158	54	2.80	2.94	0.51	0.61***	−0.027	−0.010	0.50	0.60***	0.76	0.62***
HD	66	24	12.98	14.03	2.27	2.10*	−0.161	−0.025	2.14	1.99*	0.79	0.73*

***: P < 0.001; **: P < 0.01; *: P < 0.05; ^{ns}: P > 0.05.

allele is dominant, and the proportion of Z- and zz genotypes in mixed samples highly variable, the distribution of predicted values in mixed and thin classes were overlapping making them difficult to distinguish. Depending on the application, a combination of classification-based and quantitative predictions can thus be recommended for breeding.

In the core collection, pericarp thickness, as predicted by our classification model, had no significant effect on the variation of grain biochemical and physical parameters indicating that it should be possible to breed for one or the other class of pericarp independently of other breeding targets. Whole grain NIRS calibration for grain quality parameters is preferable because it is non-destructive and does not require grinding of grain which is labor and time consuming. However, as the pericarp greatly influences NIR spectra obtained from sorghum whole grains, we investigated how it could affect the quality of NIRS prediction for grain biochemical (PR, LI, AM) and physical (ET, HD) parameters. The results showed that when using a training dataset including a combination of thick and thin pericarp samples, the error of prediction of the model was higher when it was applied to thick pericarp samples for all parameters except grain hardness. The model developed for this parameter was, however, based on small size training sets (66 samples), and it is likely that the SEP difference between thin and thick samples could not be detected in such condition. Globally, the thick pericarp can distort the spectral information available to predict endosperm parameters. Separate calibration models for each class of pericarp would provide better performance. Indeed, we examined how a calibration for protein content specifically developed with only thick pericarp samples would perform better. The average SEP value obtained from 100 randomly sampled training (N = 73) and validation (N = 10) sets of thick pericarp samples was 0.85. This value, compared to the SEP value of 1.00 obtained from the general calibration model (Table 2), shows that specific calibrations for thick pericarp samples perform significantly better. We thus propose, as a practical approach for using NIRS routinely on whole grain samples in breeding programs to develop separate calibrations for each class of pericarp and to apply a two-steps prediction approach: predict pericarp thickness using the classification model described in this study and apply the appropriate model for each type of grain to predict the other parameters.

Acknowledgments

This study was funded by the Generation Challenge Programme Grant Number G4008.48.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jcs.2016.03.008.

References

de Alencar Figueiredo, L.F., Davrieux, F., Fliedel, G., Rami, J., Chantereau, J., Deu, M., Courtois, B., Mestres, C., 2006. Development of NIRS equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. J. Agric. Food Chem. 54, 8501–8509.
Anglani, C., 1998. Sorghum for human food – a review. Plant Foods Hum. Nutr. 52, 85–95.
Ayyangar, G.N.R., Vijiaraghavan, C., Ayyar, M.S., Rao, V.P., 1934. Inheritance of characters in sorghum – the great millet. VI. Pearly and chalky grains. Indian J. Agric. Sci. 4, 96.
Bello, A.B., Rooney, L.W., Waniska, R.D., 1990. Factors affecting quality of sorghum tô, a thick porridge. Cereal Chem. J. 20–25.
Bezdek, J.C., 1981. Pattern Recognition With Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell, MA, USA.
Chantereau, J., Cruz, J.-F., Ratnadass, A., Trouche, G., 2013. Le Sorgho. Éditions Quae, Versailles, France.
Davrieux, F., de Alencar Figueiredo, L., Fliedel, G., Rami, J., Chantereau, J., Deu, M., Courtois, B., Mestres, C., 2007. Development of NIR equations for food grain quality traits through exploitation of a core collection of cultivated sorghum. NIR News 18, 12. http://dx.doi.org/10.1255/nirn.1047.
Dendy, D.A., 1995. Sorghum and Millets: Chemistry and Technology. Dendy, David AV.
Deu, M., Rattunde, F., Chantereau, J., 2006. A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49, 168–180.
Dray, S., Dufour, A.-B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.
Dykes, L., Hoffmann Jr., L., Portillo-Rodriguez, O., Rooney, W.L., Rooney, L.W., 2014. Prediction of total phenols, condensed tannins, and 3-deoxyanthocyanidins in sorghum grain using near-infrared (NIR) spectroscopy. J. Cereal Sci. 60, 138–142. http://dx.doi.org/10.1016/j.jcs.2014.02.002.
Earp, C.F., McDonough, C.M., Rooney, L.W., 2004. Microscopy of pericarp development in the caryopsis of Sorghum bicolor (L.) Moench. J. Cereal Sci. 39, 21–27. http://dx.doi.org/10.1016/S0733-5210(03)00060-2.
Fliedel, G., 1995. Appraisal of Sorghum Quality for Making Tô. Agriculture et Développement, pp. 34–42.
Food and Agriculture Organization of the United Nations, 2014. FAOSTAT.
Glueck, J.A., Rooney, L.W., 1980. Chemistry and structure of grain in relation to mold resistance. In: Sorghum Diseases, a World Review: Proceedings of the International Workshop on Sorghum Diseases, sponsored jointly by Texas A & M University (USA) and ICRISAT, 11–15 Dec. 1978. International Crops Research Institute, Hyderabad, India, pp. 119–140.
Gomez, M.I., Obilana, A.B., Martin, D.F., Madzvamuse, M., Monyo, E.S., 1997. Manual of Laboratory Procedures for Quality Evaluation of Sorghum and Pearl Millet. International Crops Research Institute for the Semi-Arid Tropics.
Hicks, C., Tuinstra, M.R., Pedersen, J.F., Dowell, F.E., Kofoid, K.D., 2002. Genetic analysis of feed quality and seed weight of sorghum inbred lines and hybrids using analytical methods and NIRS. Euphytica 127, 31–40.
Hooks, T., Pedersen, J.F., Marx, D.B., Vogel, K.P., 2006. Variation in the U.S. photoperiod insensitive sorghum collection for chemical and nutritional traits. Crop Sci. 46, 751.
Hoseney, R.C., Davis, A.B., Harbers, L.H., 1974. Pericarp and endosperm structure of sorghum grain shown by scanning electron microscopy. Cereal Chem. 51, 552–558.
Kleih, U., Ravi, S.B., Rao, B.D., Yoganand, B., 2000. Industrial Utilization of Sorghum in India. Working Paper Series, p. 44.
Kuhn, M., 2015. A Short Introduction to the Caret Package [WWW Document]. https://cran.r-project.org/web/packages/caret/vignettes/caret.pdf.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28, 1–26.
Mevik, B.-H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. J. Stat. Softw. 18, 1–24.
Murty, D.S., Kumar, K.A., 1995. Traditional uses of sorghum and millets. In: Sorghum and Millets: Chemistry and Technology, pp. 185–221. Dendy D.A.V.
Rami, J.F., 1999. Etude des facteurs génétiques impliqués dans la qualité technologique du grain chez le maïs et le sorgho. Université Paris XI.
Reichert, R.D., 1982. Sorghum dry milling. In: Sorghum in the Eighties, Proceedings of the International Symposium on Sorghum, 2–7 November 1981,

pp. 547–564. Patancheru, AP, India.
Rooney, L.W., Miller, F.R., Mertin, J.V., 1981. Variation in the structure and kernel characteristics of sorghum. In: Proceedings of the International Symposium on Sorghum Grain Quality, pp. 143–162. Patancheru, India.
Rooney, L.W., Murty, D.S., 1982. Evaluation of sorghum food quality. In: Sorghum in the Eighties, p. 571. Patancheru, Andhra Pradesh; India.
Scheuring, J.F., Sidibe, S., Rooney, L.W., Earp, C.F., 1983. Sorghum pericarp thickness and its relation to decortication in a wooden mortar and pestle. Cereal Chem. 60, 86–89.
Smith, C.W., Frederiksen, R.A., 2000. Sorghum: Origin, History, Technology, and Production. John Wiley & Sons.