Characterisation of fonio millet by near infrared hyperspectral imaging

J.A. Fernández-Pierna,[a] Ph. Vermeulen,[a] D. Stilmant,[b] B. Dupuis,[b] P. Dardenne[a] and V. Baeten[a]

[a]*Walloon Agricultural Research Centre (CRA-W), Valorisation of Agricultural Products Department, Henseval Building, Chaussée de Namur 24, 5030 Gembloux, Belgium*

[b]*Walloon Agricultural Research Centre (CRA-W), Agriculture and Natural Environment Department, Rue de Serpont 100, 6800 Libramont, Belgium*

Introduction

Fonio millet (*Digitaria exilis* Stapf) plays an important role in food supply in West Africa. This crop is a small-grained, C4-metabolism cereal with a short life-cycle, and is one of the first cereal crops to have been domesticated by West African farmers. It plays a central role in the emergence of traditional agriculture in the West African savannah, where it is now a staple food and an important part of the diet for several million people. This is true especially for the short-cycle varieties harvested at the end of the rainy season when granaries are empty. The cultivation of fonio millet in Conakry Guinea, Mali and Burkina Faso is indicative of its wide ecological adaptability.[1,2]

The objective of this work is to define a method to investigate whether the NIR hyperspectral imaging method can be used for (1) the discrimination of fonio on the basis of their site of production, at the single kernel level, as well as (2) for the analysis of fonio biochemical composition.

Material and methods

Three experimental sites at Bareng and Kankan in Conakry Guinea, and Cinzana in Mali have been studied, they are representative of the distribution of fonio millet in West Africa. All the samples were measured using a near infrared (NIR) hyperspectral camera.[3] The instrument used was a MatrixNIR Chemical Imaging System (Malvern instruments Ltd). It records sequential images with an InGaAs array detector (240 × 320 pixels) active in the 900–1700 nm range. A total of 76800 spectra are acquired by sampling the area measured. This technology is a powerful approach for remote sensing and agronomical applications as for instance in precision agriculture.[4] The major advantages of the technique are that the detection is not dependent on the expertise of the analyst, and that it is possible to automate all procedures, and to analyse more samples per unit of time than in classical near-infrared (NIR) spectroscopy.

In total, 174 sample sets were available and were measured in duplicate, leading to a total of 348 images. From these samples, 89 were not peeled seeds (42 from Bareng, 12 from Cinzana

and 35 from Kankan), while 85 were peeled seeds (35 from Bareng, 10 from Cinzana and 40 from Kankan). Each image contained between 150 and 200 seeds from the same sample set. The mean spectrum of each grain was calculated by the application of a morphological mask obtained through a process of erosion, in order to determine the contour of each grain. This mask is created by using binary images generated with the Isys software. In order to reduce the number of kernels to the approximately 100 more representative seeds in an image, a histogram (with the particle size distribution) was created using a binary image (not shown). About 100 kernels containing a size distribution close to the mean were kept for each image. A total of 35,833 spectra were kept, and constitute the database used. This large quantity of data obtained with the NIR camera needs the use of Chemometrics in order to get chemically relevant information out of measured data, and to represent and display this information. The unsupervised method, Principal Component Analysis (PCA) was performed on the data, in order to make a preliminary homogeneity study. The scores can highlight clustering, trends and outliers in the sample space in the data matrix. Based on the results obtained in this exploratory analysis, different supervised chemometric models to predict the origin of a fonio sample as Cinzana, Bareng or Kankan were generated, using Partial least squares discriminant analysis (PLS-DA). Moreover, reference values for dry matter, starch, NDF,

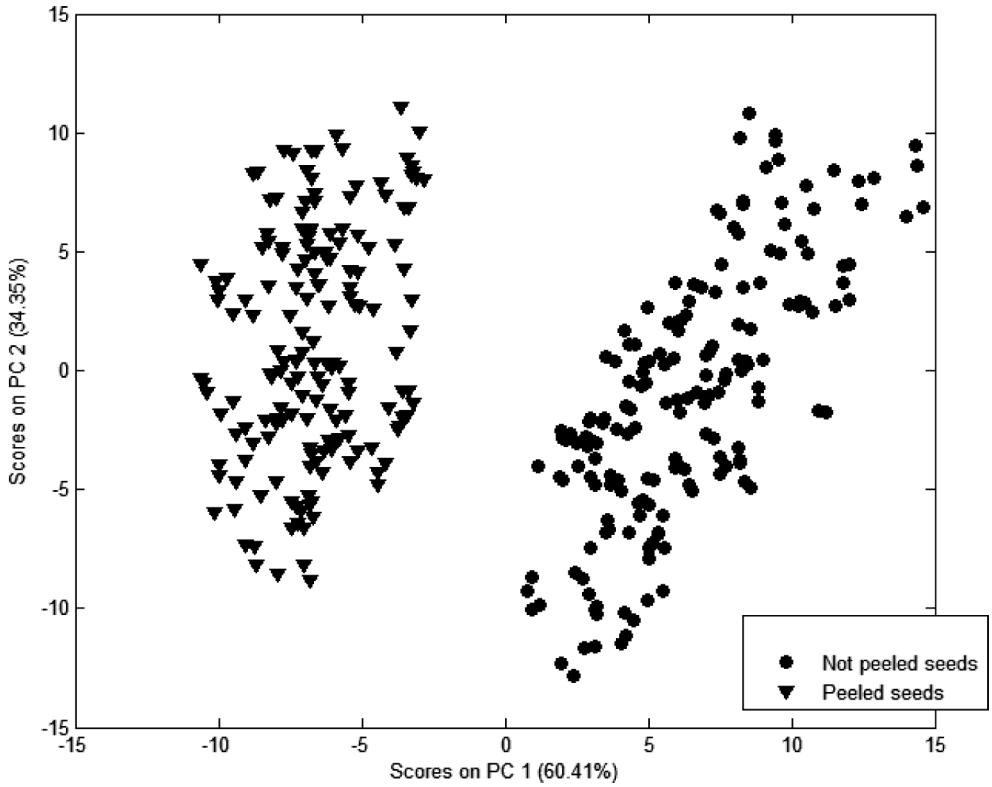

Figure 1. PC1 vs PC2 plot for all the spectra (unpeeled and peeled seeds).

ADF, ADL and ash content for some unpeeled samples were available, allowing the development of regression models, linking these variables to NIR spectra, using PLS.

Results

Unsupervised study

Figure 1 shows the PC1 vs PC2 plot for all the spectra (unpeeled and peeled seeds) indicating clear differences between both kinds of seeds. This gives an indication that both kinds of seeds can be easily discriminated, which justifies the study of each class independently.

Organic composition of only 80 unpeeled samples was available, and was used to construct regression models (see next section). For this reason, it was decided to work, in a first step, only with those unpeeled 80 samples, which corresponded to 8,046 spectra (+/– 100 spectra by grain). Figure 2 shows the PC2 vs. PC3 plot using the mean spectrum for each grain. From this figure it can be concluded that fonio produced in Kankan can be easily discriminated from the ones

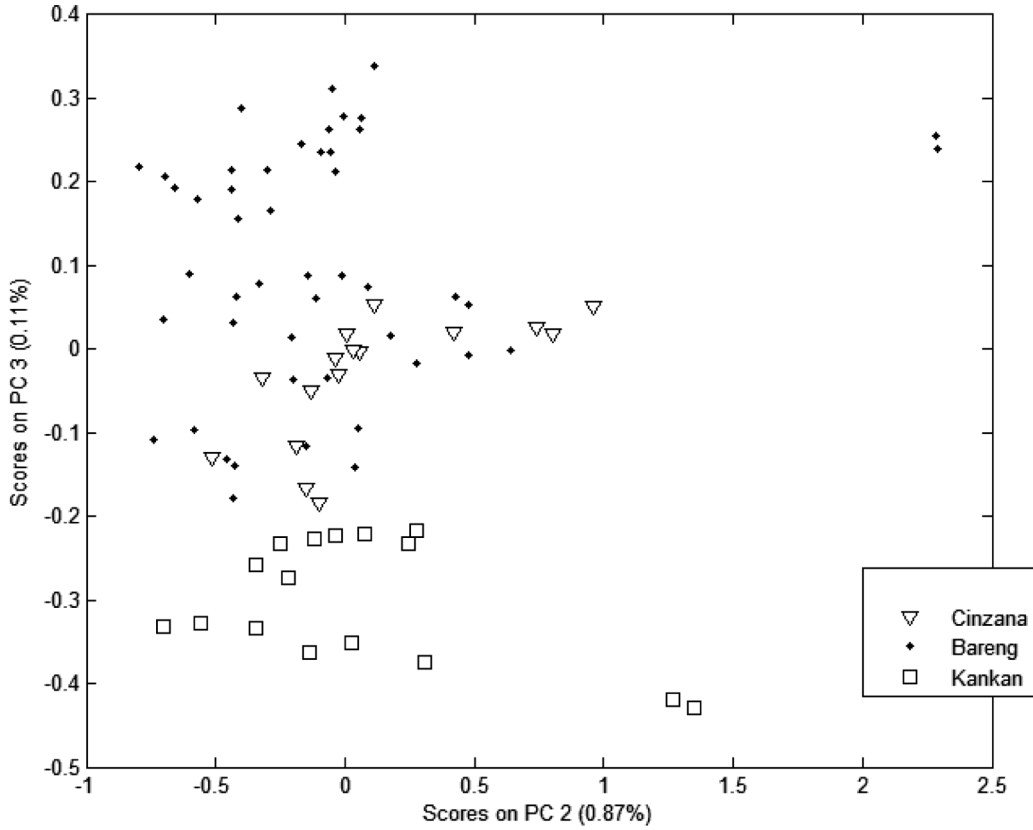

Figure 2. PC2 vs PC3 plot using the mean spectrum for each grain (unpeeled seeds).

produced in Cinzana and Bareng, where the spectra of the fonio produced in these two last sites overlap.

Supervised study

Discrimination models

The study includes the construction of individual PLS-DA models for each production site (Cinzana, Bareng or Kankan). For these models, the validation procedure used was the leave-one-out cross-validation (LOOCV). Once the models were constructed they were validated in order to estimate their performance, expressed in terms of the sensitivity and the specificity. The sensitivity is defined as the proportion of actual positives which were correctly identified as such, and the specificity measures the proportion of negatives which were correctly identified as negatives. Table 1 shows a summary of these values for the three classes using LOOCV.

In every case the models corresponded to the 'one *vs*. the rest' models. The results show that all the sites of production were easily discriminated using PLS-DA; for the Cinzana site, 93.8% of the Cinzana samples were correctly classified as Cinzana and 92.2% of the other samples were correctly classified as either Bareng or Kankan. Bareng samples had a percentage of correctly classified of 91.7% and in the case of the Kankan, as expected according to the PCA study, 100% of the samples were correctly classified.

Regression models

The available reference values for the organic composition of the 80 unpeeled seeds samples (comprising 8,046 spectra) are the dry matter, starch, NDF, ADF, ADL and ash contents. The models were constructed using PLS. A summary of all the results is shown in Table 2. In every case

Table 1. Sensitivity and specificity for the three production sites using PLS-DA.

	Cinzana	Bareng	Kankan
Sensitivity	93.8 %	91.7 %	100 %
Specificity	92.2 %	100 %	95.3 %

Table 2. Results for the different fonio components available ($N = 80$).

	Min	Max	SD	R^2	Latent Variables	RMSEC	RMSECV
Starch	61.48	68.78	1.679	0.886	9	0.563	0.751
NDF	13.95	20.18	1.359	0.844	9	0.483	0.647
Dry matter	88.03	90.07	0.581	0.875	11	0.207	0.282
ADF	10.64	14.45	0.915	0.812	12	0.261	0.375
ADL	3.01	5.05	0.502	0.401	9	0.287	0.384
Ash	2.18	3.16	0.239	0.535	8	0.133	0.165

the data were pre-processed by SNV (Standard Normal Variate) that corrects spectra for spectral noise and background effects. As for the previous study, for these models, the validation procedure used was leave-one-out cross-validation (LOOCV), and model performances are expressed through the R^2, the *RMSE* for both the calibration set (*RMSEC*) and the cross-validation set (*RMSECV*). Also the range and the standard deviation (*SD*) of the chemical constituents measured as well as the number of latent variables used to construct each model are indicated.

Conclusion

This study has demonstrated the potential of the NIR hyperspectral imaging technique for the analysis of Fonio millet samples. Results have shown that discrimination of fonio production sites, even of the same varieties, discrimination of sample according to the process stage (peeled and unpeeled) and determination of major fonio constituents by NIR hyperspectral imaging are possible.

Acknowledgements

The authors wish to thank the FP6 INCO programme for funding the STREP project n° 015403 'Upgrading quality and competitiveness of fonio for improved livelihoods in West Africa' (FONIO).

References

1. H. Adoukonou-Sagbadja, C. Wagner, A. Dansi, J. Ahlemeyer, O. Daïnou, K. Akpagana, F. Ordon and W. Friedt, *Theor. Appl. Genet.* **115,** 917 (2007).
2. J. Gigou, D. Stilmant, T.A. Diallo, N. Cissé, M.D. Sanogo, M. Vaksmann and B. Dupuis, *Exp. Agric.* **45,** 401 (2009).
3. V. Baeten, J.A. Fernández Pierna and P. Dardenne, in *Techniques and Applications of Hyperspectral Image Analysis,* Ed by H.F. Grahn and Paul Geladi. John Wiley & Sons, Ltd, Chichester, UK, p. 289 (2007).
4. J.A. Fernández Pierna, V. Baeten, J. Dubois, J. Burger, E.N. Lewis and P. Dardenne, in *Comprehensive Chemometrics,* Vol 4, Ed by S. Brown, R. Tauler and B. Walczak. Elsevier, Oxford, UK, p. 173 (2009).