Going bananas : From risky businesses to latest authentication technologies

Acta Horticulturae

Wang, Z.; Erasmus, S.W.; Ruth, S.M.

https://doi.org/10.17660/ActaHortic.2023.1367.36

Going bananas: from risky businesses to latest authentication technologies

Z. Wang, S.W. Erasmus and S.M. van Ruth

Wageningen University and Research, Food Quality and Design group, PO Box 17, 6700 AA, Wageningen, The Netherlands.

Abstract

Food fraud is common in many supply chains. Apart from paper trails, the implementation of control measures is key to combat those fraudulent activities, also in the banana supply chains. A first step in this process is to bridge the knowledge gap between origin and bananas, between the geographical/production origin factors on the one hand and the chemical and physical signatures of bananas on the other hand. There is particularly a need for rapid, low-cost methodologies that can be used to control fraud. In this study, spectral features of Cavendish bananas of different geographical origins and production systems were deciphered using near infrared (NIR) spectroscopy. Bananas from six countries, i.e., Columbia, Costa Rica, Dominica Republic, Ecuador, Panama, and Peru, were assessed. The NIR spectral data were correlated with proximate compositional data to interpret the spectral differences observed. Bananas from Dominican Republic and Ecuadorian farms showed distinct NIR signatures in comparison to the bananas of other origins. These differences appeared to be correlated with their (higher) starch contents and other compositional characteristics. Organic bananas revealed more exceptional, differential proximate compositions than their conventional counterparts, but these differences were not clearly reflected in the NIR signatures. In conclusion, spectral and compositional signatures of bananas reflect certain origin aspects, and these findings can be used to develop authentication tools to assist in the control of the integrity of value chains.

Keywords: banana, food fraud, geographical origin, NIRS, organic production, spectroscopy

INTRODUCTION

The Cavendish banana (*Musa* spp., AAA group) is the most common commercial banana cultivar in global trade. According to the Food and Agriculture Organization (FAO) of the United Nations, approximately 50 billion t of Cavendish bananas (subgroup of the AAA banana cultivar group) are produced globally every year with Latin America and the Caribbean being the largest global exporters (FAO, 2021). There is increasingly interest in bananas from sustainable food systems associated with production in particular geographical locations and in organic production systems (Santeramo and Lamonaca, 2020). Furthermore, geographical location is an important proxy for food safety risks. The growth of the organic sector and supply chain disruptions result in recurrent imbalances between supply and demand. Together with the management restrictions and added value, this results in a breeding ground for illicit, fraudulent practices in the form of non-compliance of production regulations at production sites or by substitution of organic produce with conventional produce later in the chain. Fruits are the most common products reported with organic produce fraud incidents in the period 2004-2021 (22%; Decernis Food Fraud Database, 2022).

Prevention of these illegal operations requires an upgrade of existing approaches via a pathway tailored to the banana supply chain's character. Implementation of control measures, including analytical testing, can help to combat fraudulent activities in the chain. A first step in this process is to bridge the knowledge gap between the environment/management system and bananas, and how these environmental/management system factors influence the chemical and physical characteristics of bananas. Some studies have focused on compositional differences between bananas of different origin (e.g., Forster et al., 2002). However, few

studies have focused on the chemical composition or physical characteristics of bananas specifically in view of potential authentication or the underlying mechanisms of distinctive markers. As one of the very few, the authors previously investigated differential stable isotope and elemental composition of bananas from various Costa Rican farms (Wang et al., 2020a). Although successful, a drawback is that the applied techniques are costly and require advanced laboratories. A step down the ladder of costly, time-consuming, complex technology, one finds vibrational spectroscopy techniques. Spectroscopic methods, such as near-infrared (NIR) spectroscopy, can quickly measure samples at a low cost but high speed. For example, Fourier-transform NIR (FT-NIR) spectroscopy has been applied to many food products for the authentication of geographical origin or production system (e.g., organic agriculture) (Liu et al., 2018), but not yet applied to bananas.

Therefore, in the current study, we examined the spectral features of Cavendish bananas of different geographical origins and production systems in South America using NIR spectroscopy. The spectral data were correlated with proximate compositional data to interpret the spectral differences observed. The findings aim to assist in establishing a solid identification to banana production system or location and substantiated analytical signatures to promote fair competition in banana supply chains.

MATERIALS AND METHODS

Banana samples

Cavendish bananas were harvested from ten farms in six different countries, i.e., Colombia, Costa Rica, Dominican Republic, Ecuador, Panama, and Peru. The farms were coded and details about codes and geographical/production origin information are listed in Table 1.

The growing conditions of each farm were also collected relying on the support of local farms, Global Positioning System (GPS) coordinates, and public databases CRU TS4.04 (Climatic Research Unit (CRU) Time-Series (TS) version 4.04). Bananas were harvested at similar ripeness. Briefly, 12 banana bunches were randomly selected from each farm. Then, all the bunches were packaged in clean polyethene bags and send to Wageningen University at 11-13°C. On arrival, two bananas from the top position of each bunch were selected for the current study. The banana pulp and peel were separated for each of these fruits and the pulp sliced. In the current study we will focus on the pulp because of space restrictions, but the peel was analysed as well. The pulp of two bananas from the same bunch were pooled as one sample, freeze-dried, and subsequently ground into a fine powder. Overall, the sample set included 120 banana pulp samples (10 farms × 12 samples).

Table 1. Overview of the banana samples and related growing conditions.

Country	Farm	No. of samples	Production system	Altitude (m)	Monthly mean temperature (°C)	Annual rainfall (mm year^{-1})
Colombia	CO1	12	Conventional	66	23.2	1837
Costa Rica	CR1	12	Conventional	726	23.4	2857
	CR2	12	Conventional	47	24.4	5014
	CR3	12	Conventional	24	26.3	4378
Dominican Republic	DR1	12	Organic	65	26.7	925
	DR2	12	Organic	27	26.7	925
Ecuador	EC1	12	Organic	32	22.9	1511
	EC2	12	Conventional	22	26.5	843
Panama	PA1	12	Conventional	20	19.7	3679
Peru	PE1	12	Organic	40	24.1	200

NIR spectroscopy analysis

The NIR spectra were acquired using a benchtop NIR Flex N-500 spectrometer (Büchi

Labortechnik AG, Flawil, Switzerland) in transmission mode. Approximately 5 g of banana pulp powder was transferred to a glass petri dish. All the samples were scanned across the range of 10000 to 4000 cm^{-1} (1000-2500 nm) with a resolution of 4 cm^{-1}. An internal and external reference were measured periodically for calibration of the spectrometer. Each sample was scanned in duplicate with 24 scans per spectrum. The absorbance scale values of each sample were transformed from NIR transmittance spectra using NIR Ware software (Büchi Labortechnik AG, Flawil, Switzerland) before being subjected to further data analysis.

Analysis of the proximate composition

The raw data from a former study were used (Wang et al., 2020b) to recalculate features to be correlated with the NIR spectral data. The moisture content of the fresh banana pulp samples was calculated by recording the weight difference before and after freeze-drying, which is expressed as g 100 g^{-1} (wet weight). The starch, total dietary fibre, protein, and β-carotene contents were expressed based on dry weight. All methods are detailed in the former publication of the authors (Wang et al., 2020b).

Data analysis

The NIR spectra were recorded as wavelength along the horizontal axis and absorbance values along the vertical axis for subsequent data visualization and chemometrics analysis. The raw spectral data (RAW) were firstly subjected to different data transformations including smoothing, multiplicative scatter correction (MSC), Savitzky-Golay filtering (SG), and standard normal variate transformation (SNV) to minimise the background noise and balance the weights of the variables (Fernández-Cabanás et al., 2007). A combination of pre-processing methods (RAW-SG-SNV transformed) was eventually selected for principal component analysis (PCA). PCA was conducted to explore the data distribution in the different data sets using the R packages *FactoMineR* and *Factoextra* (Kassambara and Mundt, 2017; Lê et al., 2008).

The means and standard deviations of the proximate compositional data were calculated per farm, and subsequently averaged for each component. Analysis of variance (ANOVA) was applied to compare compositional values of bananas of different origins ($p<0.05$). To explore similarities in the general composition of the bananas, the means of the proximate compositional data of the bananas from individual farms were subjected to PCA without further data pre-processing. In a next step, the NIR spectral data were correlated with the proximate compositional data. Correlation coefficients (r) were determined using Kendall methods. All the correlation analyses were performed in R 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS AND DISCUSSION

NIR spectra

All the samples had similar spectral features, resulting in overlap in many spectral regions across the entire spectral range (Figure 1a). Most of the differences between the farms were found in absorbance at wavelengths 1200, 1475, 1670, 1870, 2100, and 2280 nm. To minimise the influences from baseline variations and reveal more distinct differences among all spectra, the raw spectra was pre-processed using SNV. The SNV transformed spectra further confirmed the differences between the samples mainly occurring in the above-mentioned wavelengths (Figure 1b).

Absorption in several wavelength ranges have been related to particular molecular bonds. For instance, absorption at 1200 and 2280 nm could be related to the C-H bonds of celluloses (Xu et al., 2021). The characteristic bands at 1475 nm could be assigned to the O-H overtone of glucose (Niu et al., 2012). Continuous multiple weaker absorption peaks from 1470 to 1860 nm have been related to the CH_2 vibrations in the presence of carbohydrates in banana pulp (Senna et al., 2014). Moreover, the bands near 1870 nm could be linked to O-H and C=O stretching due to the presence of organic acids (Weyer and Lo, 2006). Finally, the bands between 2100 to 2280 nm could relate to C-H, C=O, and C-O stretching by

polysaccharides, such as starch (Workman and Weyer, 2012).

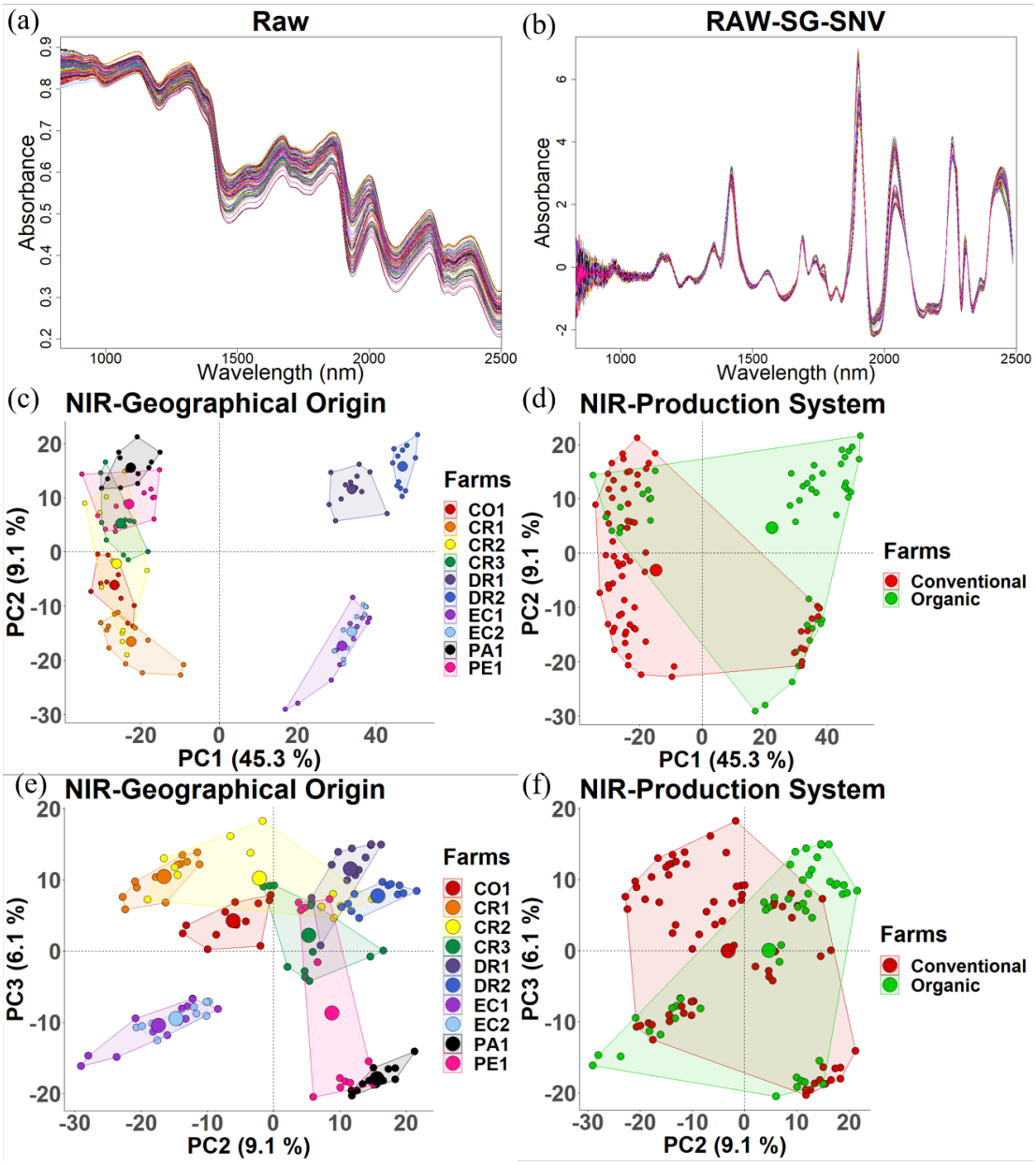

Figure 1. The raw near-infrared (NIR) spectral data (RAW) (a), Savitzky-Golay filtering and standard normal variate transformation of the raw spectral data (RAW-SG-SNV) (b) and principal component analysis (PCA) scores plots based on the RAW-SG-SNV transformed NIR spectral data of the banana samples from different geographical origins (c/e) and production system origins (d/f) displaying PC1 and PC2 (c/d) and PC2 and PC3 (e/f). CO: Colombia; CR: Costa Rica; DR: Dominican Republic; EC: Ecuador; PA: Panama; PE: Peru; Codes of farms are explained in Table 1; n=10 farms × 12 samples.

Exploration of the NIR spectral patterns

In our study, the NIR spectra of bananas showed distinctive absorption peaks near 1200, 1500, 2200, and 2300 nm, which likely indicate the presence of dietary fibre, starch, and other substances. These results align with previously published studies on bananas (Sripaurya et

al., 2021).

To compare the spectra of the various samples as a whole, PCA was conducted on the RAW-SG-SNV pre-processed spectral data of the 120 banana samples. With this analysis, the aim was to explore similarities and dissimilarities in the NIR patterns of the samples. The scores and loading plots presenting the first two dimensions of the PCA (explaining 54.4% of the variance) are shown in Figure 1c and d, respectively. The main feature in the plot is the separation of banana samples from farms in the Dominican Republic (DR1, DR2) and Ecuador (EC1, EC2), which are associated with high positive scores on PC1 (explaining 45.3% of the variance), from bananas from all other farms (CO1, CR1, CR2, CR3, PA1, PE1). The latter are associated with high negative scores on PC1. Furthermore, the Dominican Republic and Ecuadorian samples are subsequently separated from each other along the second dimension (PC2, explaining 9.1% of the variance): the Dominican Republic samples present high positive scores and the Ecuadorian samples high negative scores. It appears that there is a certain geographical component that leads to the distinctive patterns of the Dominican Republic and Ecuadorian samples on the one hand, and all other samples on the other hand. The production system seems to be not a consistent, influential factor regarding the separation of the samples given the overlap between conventional and organic as shown in Figure 2d. In the scores plot displaying the second and third dimension (PC2 and PC3; Figure 1e-f), the banana samples of the various origins are more equally spread and grouped per country. In this plot the Ecuadorian bananas cluster together too and stand out once again.

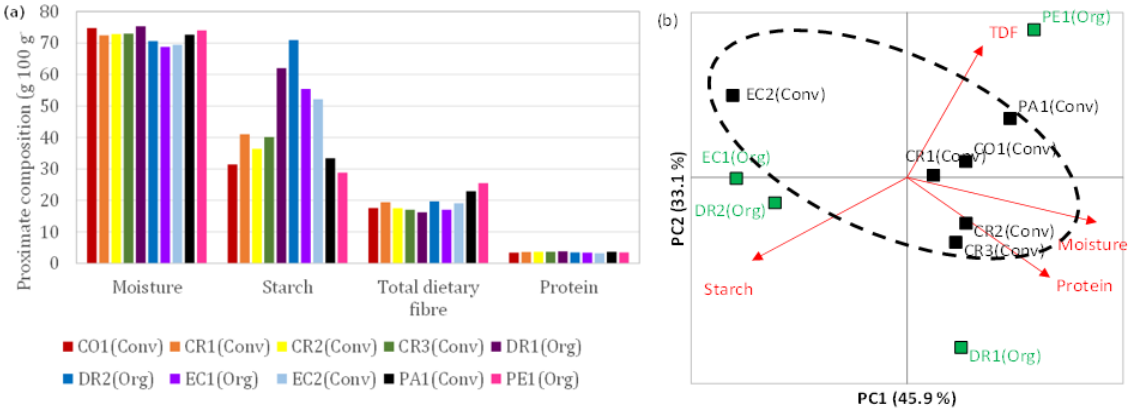

Figure 2. Proximate composition of the banana samples from farms of different geographical and production origins (recalculated and graphically depicted based on data of Wang et al., 2020a, b). a) Moisture contents based on wet weight; starch, total dietary fibre (TDF) and protein contents based on dry weight; b) Principal Component Analysis scores plot of the proximate compositional data. Explanation of the farm codes is provided in Table 1.

Proximate composition

The means per farm were calculated for the various components, which are graphically represented in Figure 2a.

On average, the bananas comprised 72.4 g ± 2.2 moisture 100 g^{-1} fresh weight, 45.2±14.2 g starch 100 g^{-1} dry weight, 19.2±2.9 g total dietary fibre 100 g^{-1} dry weight, and 3.6 g ± 0.2 protein 100 g^{-1} dry weight. They also consisted of 0.10±0.12 g β-carotene 100 g^{-1} dry weight. The carbohydrates measured (starch and total dietary fibre) determined approximately 65% (w/w) of the dry weight of the bananas, with starch being the predominant component. During ripening starch is converted to sugars, resulting in the softening of texture and sweet taste associated with ripe bananas. The bananas were in the unripe (green) stage. The bananas differed only slightly in moisture and protein contents but showed greater variation in starch and total dietary fibre contents – which could be linked to some slight differences in ripening stage. In previous work by Phillips et al. (2021), similar moisture contents in the

range of 70-80 g 100 g^{-1} fresh weight were reported. In the current study, we determined 12.4 g starch 100 g^{-1} fresh weight. Phillips et al. (2021) reported a similar content for unripe bananas. They also showed that the starch content decreased considerably, to contents below 4 g 100 g^{-1} fresh weight during ripening. The total dietary fibre content measured in the current study amounted to 5.3 g 100 g^{-1} fresh weight. The former authors showed that the content was very much determined by the method used. Our values are between those determined with the two methods that were applied by Phillips et al. (2021). Finally, the protein content of the bananas in the current study was 1.0 g protein 100 g^{-1} fresh weight. Phillips et al. (2021) did not measure the protein content, but they did measure sucrose/fructose/glucose contents. They also showed that the total sugar content increased rapidly with ripening.

The proximate compositional data (moisture, starch, total dietary fibre, and protein contents) were subjected to PCA analysis. A scores plot of the first two dimensions of the PCA is presented in Figure 2b. It appears that a high starch content is associated with relatively lower total dietary fibre contents. Remarkably, the bananas from conventional farms formed an inner circle pattern with the bananas from organic farms distributed outside this circle in multiple directions. This indicates that the bananas from conventional farms were more similar in composition, and the bananas from organic farms presented more differential proximate compositions. Even so, it must be kept in mind that although this study analysed 120 samples, there were only ten different farms involved, and, hence, the practical variation in the current set is limited.

Correlations between NIR spectra and proximate compositional data

The relationship between the spectral features and banana chemical composition was explored. Correlation analyses were carried out and a heatmap of which is shown in Figure 3. The heatmap outlines the correlation between the RAW-SG-SNV pre-processed NIR spectra in the wavelength range 500-2500 nm on the *x*-axis and the compositional data on the *y*-axis.

The starch content of bananas correlated strongly with absorbance at various wavelengths, including those around 2216 nm (r=0.78, $p<0.05$). As the main differences in the NIR spectra as described in the former sections, were seen at 1200, 1475, 1670, 1870, 2100, and 2280 nm (Figure 1b), it provides evidence that the differences in geographical origin are likely caused by differences in the contents of polysaccharides including starch as well as the organic acids in the banana. When comparing to the proximate composition of the bananas from the farms and averaging the contents for the Dominican Republic, Ecuadorian, and all other bananas, respectively, they show a significant difference in starch contents (ANOVA, $p<0.05$). The Dominican Republic samples present a starch content of 66.5±6.3 g 100 g^{-1} dry weight, the Ecuadorian samples a starch content of 53.8±11.0 g 100 g^{-1} dry weight, and all others a starch content of 35.2±4.9 g 100 g^{-1} dry weight. Therefore, the differences in starch contents are likely to have contributed to the spectral differentiation of the groups.

The heat map reveals a weak correlation between the total dietary fibre contents and absorbance values. For example, a low correlation coefficient for total dietary fibre and absorbance at 1211 nm was determined (r=0.36, $p<0.05$). Similarly, weak correlations between protein contents and absorbances were found. It is also unlikely that these components were differential determinants in the current study since the Dominican Republic and the Ecuadorian samples did not differ significantly from the set of other groups in regard to total dietary fibre or protein contents (ANOVA, $p<0.05$). On the other hand, interestingly, for the carotene content, moderate to strong correlations were observed for absorbance at various wavelengths. Some will be connected directly to β-carotene, but some may also have an indirect relationship.

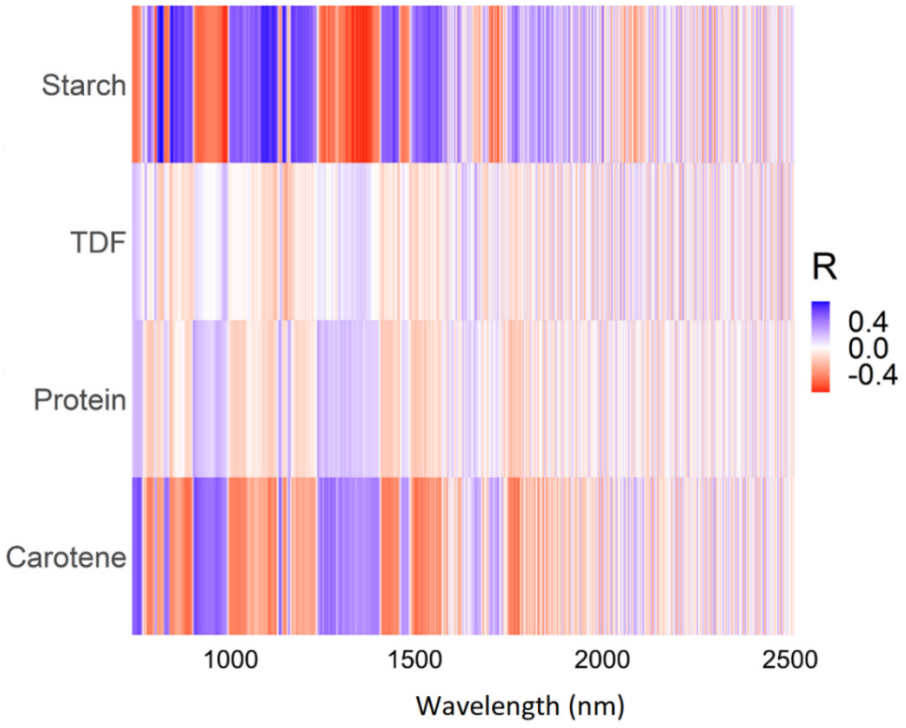

Figure 3. Heat map presenting correlations between near-infrared spectral data of the banana samples along the horizontal axis and their proximate compositional data (starch, total dietary fibre (TDF), protein, and β-carotene contents) along the vertical axis.

CONCLUSIONS

This study revealed that bananas from certain geographical origins, in particular those from the Dominican Republic and Ecuador, could be distinguished from bananas from other geographical origins by their NIR spectral signatures. Most likely, their starch content contributed to their distinction, among other compositional features. The NIR spectral signatures did not reveal consistent differences between bananas produced in organic versus conventional systems. However, the composition of the organic bananas is more exceptional, showing more extreme values than their conventional counterparts and compared to each other. The conventional bananas were more similar in composition as a group. This may be due to more standardized agricultural practices in the conventional production system in comparison to the organic systems, where practices vary widely. To confirm current results, it would be useful to compare bananas from organic and conventional production systems which originate from the same geographical location in future research. Furthermore, established geographical origin authentication technologies, such as stabile isotope analysis and elemental analysis could be explored as confirmatory techniques.

ACKNOWLEDGEMENTS

The authors wish to thank all partners in the 'Going Bananas' project for their assistance in sourcing the banana samples for this study, with special thanks to our colleague Erwin Brouwer from Wageningen Food Safety Research for coordination of the sample collection. Wageningen University and research student Xiaotong Liu is gratefully acknowledged for her contribution to the proximate compositional analyses. Furthermore, authors acknowledge the funding provided by the Dutch Top sector Agri & Food of the project 'Going bananas' (No. AF-16008) and the scholarship of the first author provided by the China Scholarship Council (No. 201706820017).

Literature cited

Decernis Food Fraud Database. (2022). https://decernis.com/products/food-fraud-database/ (accessed October 3, 2022).

FAO. (2021). FAOSTAT Statistics Database. http://www.fao.org/faostat/en/#home.

Fernández-Cabanás, V.M., Garrido-Varo, A., Olmo, J.G., De Pedro, E., and Dardenne, P. (2007). Optimisation of the spectral pre-treatments used for Iberian pig fat NIR calibrations. Chemom. Intell. Lab. Syst. *87* (*1*), 104–112 https://doi.org/10.1016/j.chemolab.2006.10.005.

Forster, M.P., Rodríguez Rodríguez, E., and Díaz Romero, C. (2002). Differential characteristics in the chemical composition of bananas from Tenerife (Canary Islands) and Ecuador. J Agric Food Chem *50* (*26*), 7586–7592 https://doi.org/10.1021/jf0257796. PubMed

Kassambara, A., and Mundt, F. (2017). Factoextra: extract and visualize the results of multivariate data analyses. R Package Version *1* (*5*), 337–354.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR : an R package for multivariate analysis. J. Stat. Softw. *25* (*1*), https://doi.org/10.18637/jss.v025.i01.

Liu, N., Parra, H.A., Pustjens, A., Hettinga, K., Mongondry, P., and van Ruth, S.M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. Talanta *184*, 128–135 https://doi.org/10.1016/j.talanta.2018.02.097. PubMed

Niu, X., Zhao, Z., Jia, K., and Li, X. (2012). A feasibility study on quantitative analysis of glucose and fructose in lotus root powder by FT-NIR spectroscopy and chemometrics. Food Chem *133* (*2*), 592–597 https://doi.org/10.1016/j.foodchem.2012.01.064. PubMed

Phillips, K.M., McGinty, R.C., Couture, G., Pehrsson, P.R., McKillop, K., and Fukagawa, N.K. (2021). Dietary fiber, starch, and sugars in bananas at different stages of ripeness in the retail market. PLoS One *16* (*7*), e0253366 https://doi.org/10.1371/journal.pone.0253366. PubMed

Santeramo, F.G., and Lamonaca, E. (2020). Evaluation of geographical label in consumers' decision-making process: a systematic review and meta-analysis. Food Res Int *131*, 108995 https://doi.org/10.1016/j.foodres.2020.108995. PubMed

Senna, A.M., Novack, K.M., and Botaro, V.R. (2014). Synthesis and characterization of hydrogels from cellulose acetate by esterification crosslinking with EDTA dianhydride. Carbohydr. Polym *114*, 260–268 https://doi.org/10.1016/j.carbpol.2014.08.017. PubMed

Sripaurya, T., Sengchuai, K., Booranawong, A., and Chetpattananondh, K. (2021). Gros Michel banana soluble solids content evaluation and maturity classification using a developed portable 6 channel NIR device measurement. Measurement *173*, 108615 https://doi.org/10.1016/j.measurement.2020.108615.

Wang, Z., Erasmus, S.W., Dekker, P., Guo, B., Stoorvogel, J.J., and van Ruth, S.M. (2020a). Linking growing conditions to stable isotope ratios and elemental compositions of Costa Rican bananas (*Musa* spp.). Food Res Int *129*, 108882 https://doi.org/10.1016/j.foodres.2019.108882. PubMed

Wang, Z., Erasmus, S.W., Liu, X., and van Ruth, S.M. (2020b). Study on the relations between hyperspectral images of bananas (*Musa* spp.) from different countries, their compositional traits and growing conditions. Sensors (Basel) *20* (*20*), 5793 https://doi.org/10.3390/s20205793. PubMed

Weyer, L.G., and Lo, S.-C. (2006). Spectra-structure correlations in the near-infrared. In Handbook of Vibrational Spectroscopy, P.R. Griffiths, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.1817–1837.

Workman, J., Jr., and Weyer, L. (2012). Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, 2nd edn (Boca Raton, Florida, USA: CRC Press), pp.326.

Xu, J.L., Lin, X., Hugelier, S., Herrero-Langreo, A., and Gowen, A.A. (2021). Spectral imaging for characterization and detection of plastic substances in branded teabags. J Hazard Mater *418*, 126328 https://doi.org/10.1016/j.jhazmat.2021.126328. PubMed