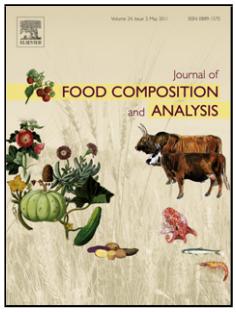
Journal Pre-proof



Discrimination of *Musa* banana genomic and sub-genomic groups based on multi-elemental fingerprints and chemometrics

Kayise Hypercia Maseko, Thierry Regnier, Tonna Anyasi, Belinda Du Plessis, Laura Suzzanne Da Silva, Funso Kutu, Obiro Cuthbert Wokadala

PII: S0889-1575(21)00534-2

DOI: https://doi.org/10.1016/j.jfca.2021.104334

Reference: YJFCA 104334

To appear in: *Journal of Food Composition and Analysis*

Received Date: 14 August 2021

Revised Date: 2 December 2021

Accepted Date: 6 December 2021

Please cite this article as: Maseko KH, Regnier T, Anyasi T, Du Plessis B, Da Silva LS, Kutu F, Wokadala OC, Discrimination of *Musa* banana genomic and sub-genomic groups based on multi-elemental fingerprints and chemometrics, *Journal of Food Composition and Analysis* (2021), doi: https://doi.org/10.1016/j.jfca.2021.104334

Discrimination of *Musa* banana genomic and sub-genomic groups based on multi-elemental fingerprints and chemometrics

Kayise Hypercia Maseko [a, b], Thierry Regnier [b], Tonna Anyasi [a], Belinda Du Plessis [b], Laura Suzzanne Da Silva [b], Funso Kutu [c], Obiro Cuthbert Wokadala [a,c, *]

[a] *Agricultural Research Council – Tropical and Sub-tropical Crops, Agro-Processing and Postharvest Technologies Division, Private Bag X11208, Mbombela, South Africa, 1200;*

[b] *Tshwane University of Technology, Department of Biotechnology and Food Technology, Private Bag X680, Pretoria, South Africa, 0083;*

[c] *University of Mpumalanga, Faculty of Agriculture and Natural sciences, Mbombela, South Africa, 1200*

[*]Corresponding author: University of Mpumalanga, Tel: +27130020368; Email: obiro.wokadala@ump.ac.za

Highlights

- Classification of banana subgenomic and genomic groups was assessed
- Unripe banana flour multi-elemental figure prints successfully facilitated classification
- PCA-SVM was a more effective classification method compared to PCA-LNN and PCA-LDA
- Verified classification accuracy of 100% was attained for genomic and subgenomic groups

ABSTRACT

The potential of unripe banana flour multi-elemental fingerprints for classifying banana genomic and sub-genomic groups was assessed using chemometrics. The elemental concentration of N, P, K, Mg, Ca, Zn, Cu, Mn, Fe, and B in unripe banana flour from 33 banana varieties belonging to four genome groups and 11 sub-genome groups were determined using Flame-atomic Absorption spectrometry and colorimetry. Principal component analysis (PCA) combined with linear discriminant analysis (LDA), support vector machine (SVM), and artificial neural network (ANN) was applied for classification with an 80:20 split between the calibration and verification sets (157 and 39 samples, respectively). The elements K, N, and Mg presented the highest mean concentrations of 1273 mg/100 g, 424 mg/100 g, and 132 mg/100 g, respectively. The classification model verification set samples were successfully classified based on their genome groups (100% accuracy) and sub-genome groups (78.95 to 100% accuracy) for PCA-LDA, PCA-ANN, and PCA-SVM models. The results demonstrate that multi-elemental fingerprinting combined with chemometrics can be employed as an effective and feasible method for classification of *Musa* genomic and sub-genomic groups.

Key words: Unripe banana flour; Elements; Banana sub-genome groups; Banana varieties; Banana genome groups; Principal component analysis; Linear discriminant analysis; Support vector machine; Artificial neural networks.

1. Introduction

Bananas (*Musa* spp.) are important staple foods for people living in tropical and sub-tropical regions of the world (Brown et al., 2017); and have been part of the human diet since the start of history (Nelson, Ploetz, & Kepler, 2006), contributing significantly to food security (Oyeyinka &

Afolayan, 2019). However, the prevalence of various cultivar names and synonyms in their different languages and dialects of a country or region is a common problem faced by banana taxonomists and horticulturists (Karamura, Karamura, & Tinzaara, 2012; Valmayor et al., 2000). Where in most cases the same cultivar (variety) is known by different names in different countries, intermittently the same name is applied to distinct varieties (Valmayor et al., 2000). There are over 1000 different banana varieties cultivated and estimated to exist worldwide, with the genus *Musa* comprising of more than 50 species, some of which have several subspecies, and six naturally occurring genomes (i.e., AA, AAA, AB, AAB, ABB, and ABBB) (Nyombi, 2020; Srivastava & Hu, 2019).

The ploidy arrangement of edible bananas differs, with hybrids resulting from two diploid species *M. acuminata* (AA) and *M. Balbasiana* (BB); containing diploid, triploid, and tetraploid hybrids (El-Khishin, Belatus, El-Hamid, & Radwan, 2009). The hybrids are identified by the letters A and B that denote genome composition and ploidy nature in relation to parental origin (Sulaiman, Yusoff, Eldeen, Seow, Sajak, & Ooi, 2011). As a result, the varieties are divided into genomic groups based on the number of chromosomes (AA, AB, AAA, AAB, ABB, AAAA, AAAB, AABB, and ABBB) (Pereira & Maraschin, 2015).

Genome and transcriptome sequencing have been used to distinguish the genetic diversities and relationships of *Musa* varieties for in their breeding programs (Borborah et al., 2020; Ghag & Ganapathi, 2017; Igwe, Ihearahu, Osano, Acquaah, & Ude, 2021). Other studies have attempted to distinguish between varieties, genome groups, and sub-genome groups of *Musa* bananas based on phenotypic characteristics. Morphometric and marker-based classifications have been recently applied with some success based on morphological descriptors, albeit with a challenge of overlaps (Borborah et al., 2020; Brisibe, Ubi, & Ekanem, 2021; Cruz-Cárdenas, Youssef, & Escobedo-Graciamedrano, 2017).

Multivariate multi-elemental fingerprint analysis has recently been successfully used in distinguishing various crops based on crop varieties/cultivars of pistachios (Esteki, Heydari, Simal-Gandara, Shahsavari, & Mohammadlou, 2021), tea (Chen, Yu, Xu, Chen, & Shi, 2009), and cowpeas (Pérez-Rodríguez, Gaiad, Hidalgo, Avanza, & Pellerano, 2019). It is therefore possible to distinguish between bananas grown in different agronomic practices and regions based on the mineral profile (Forster, Rodríguez, Martín, & Romero, 2002). Multi-elemental fingerprint analysis has been shown to be effective in the identification, authentication, and classification of edible plant products based on their geographical origins and botanical classification (Benabdelkamel et al., 2012; Cheajesadagul, Arnaudguilhem, Shiowatana, Siripinyanond, & Szpunar, 2013; Esteki et al., 2021; Suarez-Tapia, Kucheryavskiy, Christensen, Thomsen, & Rasmussen, 2017). Multi-elemental fingerprints were shown to have capacity to facilitate the distinguishing between cultivars of pistachio nuts (Esteki et al., 2021), cowpeas (Pérez-Rodríguez et al., 2019), cocoa beans (Kruszewski & Obiedziński, 2018) and tea (Chen et al., 2009). Multi-elemental fingerprints have been suggested not to be effective in distinguishing between varieties when crops from different agronomic systems and different growing sites are used (Suarez-Tapia et al., 2017). However, the study by Alkarkhi, Ramli, and Easa (2009), based on ripe banana flour, indicated that it is possible to distinguish bananas varieties of the same genomic group (AAA) grown at different geographic sites. The application of multi-elemental fingerprints on different foods for classification and authentication purposes has utilized Flame Atomic Emission spectroscopy (FAAS) (Fechner, Hidalgo, Ruiz Díaz, Gil, & Pellerano, 2020; Kment, Mihaljevič, Ettler, Šebek, Strnad, & Rohlová, 2005; Szymczycha-Madeja, Welna, & Pohl, 2015), inductively coupled plasma optical emission spectrometry (ICP-OES) (Esteki et al., 2021; Kruszewski et al., 2018), inductively coupled plasma mass spectrometry (ICP-MS) (Pérez-Rodríguez et al., 2019; Ruggiero, Fontanella, Amalfitano, Beone, & Adamo, 2021), and energy dispersive X-ray fluorescence spectrometry (EDXRF) (Herreros-Chavez, Oueghlani, Morales-Rubio, Cervera, & de la Guardia, 2019; Paltridge et al., 2012; Winkler, Rauwolf, Sterba, Wobrauschek, Streli, &

Turyanskaya, 2020). The ICP-MS and ICP-OES methods have an advantage of a high sensitivity when compared to FAAS, however all require tedious sample preparation and marching matrices standards (Drivelos & Georgiou, 2012). The FAAS is relatively cheap and has low running costs which can enable rapid adoption of protocols. The main advantage of EDXRF is its rapid and non-destructive nature (Paltridge et al., 2012), with the major disadvantage being low sensitivity and high matrix effects (Mir-Marqués, Martínez-García, Garrigues, Cervera, & de la Guardia, 2016).

Alkarkhi et al. (2009) showed that, based on linear discriminant analysis (LDA), that it was possible to distinguish between two banana varieties in the AAA genomic group based on the mineral fingerprint of ripe banana flour. However, recent research done by Devarajan et al. (2021) based on the mineral profile of ripe banana flour from 100 Indian accessions, concluded based on principle component analysis (PCA) and cluster analysis (CA), that the mineral profile cannot facilitate the distinction between banana cultivars/accessions, genomic groups or sub-genomic groups. Research carried out on banana ripening has been shown that ripening affects the mineral profile of bananas (do Prado Ferreira & Tarley, 2020; Sogo-Temi, Idowu, & Idowu, 2014). At present, no studies have reported on the application of unripe banana flour mineral fingerprints for distinguishing between genome, and sub-genome groups. The aim of the present study therefore was to determine efficacy of unripe banana flour mineral fingerprints in distinguishing between banana sub-genomic (11), and genomic groups (4) which consisted of 33 varieties. A chemometrics approach based on a combination of PCA with linear discriminant analysis (LDA), support vector machine learning (SVM), and artificial neural networks (ANN) was used. The chemometrics approach applied, extracts the underlying unsupervised data structure using PCA, then applies the structure to the classical classification techniques of LDA, SVM and ANN.

2. Materials and methods

2.1 Chemicals and reagents

All chemicals and reagents were of analytical grade. Nitric acid (55 %) and perchloric acid (70 %) for the digestion of samples were purchased from ChemLab supplies (Benrose, Johannesburg, South Africa). Sulphuric acid (98 %) and hydrogen peroxide (30 %) for the extraction of nitrogen were purchased from Reflecta Laboratory supplies (Germiston, Johannesburg, South Africa). The reagents ammonium molybdate were purchased from Laboratory supplies (Edenvale, Johannesburg, South Africa), stannous chloride were purchased from Hach (Northriding, Johannesburg, South Africa), ethylenediamine tetra-acetic acid (EDTA) buffer, azomethine-H, and sodium metabisulphite were purchased from Merck Sigma-Aldrich (Kempton Park, Johannesburg, South Africa). Salicylate were purchased from Crest Chemicals (Midrand, Johannesburg, South Africa), and sodium nitroprusside were purchased from Adcock Ingram Limited (Midrand, Johannesburg, South Africa). All solutions were prepared with deionized water (LabTech, Sepsci water still, Daihan, Korea).

2.2 Unripe banana flour collection and preparation

The 33 *Musa* banana varieties (selected based on their availability at the commencement of the study), were grown on the same gene farm orchard and cultivated under the same agronomic conditions (weather, fertilization, irrigation, soil composition) at the Agricultural Research Council (ARC) gene bank farm in Nelspruit, South Africa. The 33 varieties used for this study consisted of 20 varieties that were pure triploids (AAA genome group), with 18 from the Cavendish sub-genome group; three varieties were pure diploids (AA genome group); five varieties were triploids (AAB genome group); and five varieties were tetraploids (AAAB genome group). The banana varieties were made up 11 sub-genome groups and 4 genomic groups (Table 1).

The unripe (green) bananas were harvested at full maturity, when the fingers were at least three quarters full (3/4 of fruit is rounded). For each harvest, at least two crates (40 - 60 kg) of

banana fingers were used to prepare flour for each banana variety. The harvesting took place twice between the period of Jan 2019 to Dec 2019. Banana flour was prepared from the pulp of the unripe bananas, within 24 h after harvest according to Kongolo et al. (2017). The processing initially began with the harvested bananas being rinsed in tap water, manually peeled, and dipped immediately in sodium metabisulphite solution (4 ppm) for 30 min at ± 25 °C to inhibit oxidation. The banana pulp was then sliced using an HLC-300 vegetable cutter (Newin Machinery, Zhengzhou, Henan, China), into slices of about 1 cm thick. The slices were dried for 15 h at 50 °C, using AD-3000 cabinet dryer (Agri-dryers, Dryers for Africa, Limestone Hill, Western Cape, South Africa). Milling of the dried banana chips was done using a pilot-plant hammer mill (Drosky S1, Alberton, Gauteng, South Africa), with an aperture of 1 mm, and the flour immediately packed in airtight containers. The flour from the two harvests was combined to form the analytical sample (a total of 33 types/varieties of unripe banana flour).

2.3 Determination of elemental content

The elemental composition of the unripe banana flour was determined according to Tsomele et al. (2021) using Agri-Laboratory Association of Southern Africa (AgriLASA) methods as described subsequently.

Sample pretreatment: The flour samples were dried in a Term-O-Mat oven (Labotec, Randjespark, Midrand Johannesburg, South Africa) at 60 °C for 24 h and then passed through a 1 mm, Model 2 Wiley® mill (Thomas Scientific, Swedesboro, New Jersey, United States of America). The elements were then extracted by digestion on a D60 digestion block (Labcon Laboratory Equipment, Chamdor, Johannesburg, South Africa). The extraction for calcium (Ca), magnesium (Mg), phosphorus (P), potassium (K), iron (Fe), zinc (Zn), copper (Cu), manganese (Mn), and boron (B) involved the digestion of samples (0.5 g) with a mixture of nitric acid (4 mL) and perchloric acid (2 mL). The samples were heated at 100 °C for 2 h, and then at 180 °C for 6 h for complete the decomposition of organic substances. The samples were quantitatively transferred

to a volumetric flask and made up to 25 mL with deionized water. The extraction for nitrogen (N) involved the digestion of the sample (0.25 g) at 300 °C for 6 h with a mixture of sulphuric acid (4 mL) and hydrogen peroxide (3 mL) then quantitatively transferred the sample to a volumetric flask and making up the volume to 100 mL with deionized water.

Element composition analysis: The determinations of the elemental content for Ca, Mg, K, Fe, Zn, Cu, and Mn were performed by Flame-atomic Absorption spectrometry with a Varian AA 200 series spectrometer (SpectraLab Scientific Incorporation, Markham, Ontario, Canada). The FAAS used an acetylene/air flame. The absorbance measurements done at wavelengths of 422.7 nm, 285.2 nm, 766.5 nm, 213.9 nm, 324.7 nm, 279.5 nm, and 248.3 nm for Ca, Mg, K, Zn, Cu, Mn, and Fe, respectively. The slit widths for each element and associated wavelength were set for a specific working range appropriate for the samples (see Appendix Table A1). The analysis for the elements N, P, and B, was performed using colorimetry using an auto-analyzer (Bran and Luebbe Auto Analyzer 3®; Bran+Luebbe GmbH, Norderstedt, Germany). The reagents used were ammonium molybdate and stannous chloride; EDTA buffer and azomethine-H; salicylate and sodium nitroprusside solutions for P, B and N respectively. The absorbance readings were taken at 420 nm (B) and 660 nm (N and P).

For all the analyses (both FAAS and colorimetry), the concentrations of each element were determined from appropriately diluted standard solution curves that were run concurrently with the samples and had regression coefficient values of ≥ 0.9995. The accuracy of the results was ascertained using an AgriLASA certified reference material (dried avocado) and a correction factor applied to the obtained concentrations for each element. The recoveries were in the range 90-110% %. To prevent ionization interferences during Ca, Mg and K analyses, an ionization buffer (2000 ppm strontium chloride) was added to each standard and sample solution. Analysis was done six times for each banana flour sample, and the macro-element and micro-element concentrations reported in mg/100 g, on a dry weight basis (dwb). The nitrogen content was converted to protein content using a conversion factor of 6.25, for discussion purposes. The concentrations of the

elements were discussed as a percentage (%) of the Food and Agriculture Organization of the United Nations/World Health Organization (FAO/WHO) recommended daily allowance (RDA) (FAO/WHO, 2019a), based on 100 g of flour.

2.4 Multivariate data analysis for mineral fingerprint classification

Multivariate analysis was done using Statistica® software version 8 (StatSoft, Tulsa, OK, USA). The multivariate data analysis was done separately (independently) for genomic groups, and sub-genomic groups. These analyses were subsequently referred to as the genomic group based and sub-genomic group based analysis respectively. An initial analysis was done, involving descriptive statistics (means, standard deviations, standard errors of means) and one-way analysis of variance (ANOVA) with Fisher's least significant test (LSD) at a 95 % confidence interval ($p \leq 0.05$), with the genomic and sub-genomic groups as the independent/categorical variables. Principal component analysis (PCA) in combination with linear discriminant analysis (PCA-LDA), artificial neural network (PCA-ANN) and support vector machine (PCA-SVM) was used to identify multi-elemental mineral fingerprints and for classification analysis.

The element content data was separated into training/calibration and verification/test sets with an 80:20 split using MS-Excel random numbers sorting. The training/calibration set consisted of 157 cases while the test set consisted of 39 cases. The training/calibration set comprised of 96 AAA, 24 AAAB, 25 AAB, and 12 AA cases for genomic group based analysis. While the calibration/training set comprised of 86 'Cavendish', 19 'Goldfinger selection', four 'Inarnibal', five 'Lakatan', five 'Mysore', nine 'Pome', five 'Pome synthetic hybrid', five 'Red', eight 'Sucrier'', five 'Sukali Ndizi', and six 'Synthetic hybrid' cases for sub-genomic group based analysis. The data was pre-processed by scaling using unit standard deviations (each value divided by the variable/element content standard deviation).

2.4.1 Principal Component Analysis and Outlier Identification

Principal component analysis was performed first as a means of extracting the important features/structure (fingerprints) of the data. PCA was done separately (independently) for genomic group, and sub-genomic groups (two separate PCAs). The separate genomic group PCA did not have the sub-genomic group as a variable, and the sub-genomic group PCA did not have genomic group as a variable. The PCAs were done using the NIPALS algorithm, with seven v-fold cross-validation. An initial PCA was done for outliers' identification using both Hoteling T^2 and the squared predictions error (SPE) with a cut-of percentages of 1 % (inclusion of 99 % of the data). Subsequent PCAs were then done. The number of retained components in the subsequent PCAs was based on Eigen value cut-offs of ≥1 (Rea & Rea, 2016). The importance of each variable/element to the resultant PCA models was assessed using modelling power (P), given by the formula: $\text{Power} = 1 - \frac{(SVj)}{(SVi)}$, where: *SVj* is the residual standard variation of the j^{th} variable and *SVi* is its initial standard deviation. Variable loadings were also assessed to ascertain the contribution of each element to the various principal components.

2.4.2 Classification Analysis: Linear Discriminant Analysis, Support Vector Machine, and Artificial Neural Networks analysis

For classification analysis, the PCA scores obtained for each of the separate (independent) sub-genomic, and genomic group based PCAs were subjected to linear discriminant analysis (LDA), support vector machine (SVM), and artificial neural network (ANN) analysis.

Linear discriminant analysis (LDA): Forward stepwise model building was utilized, with probability to enter and remove both set at 0.5. The F-to-enter and remove was also set at 1 with the maximum number of steps set at 100. The sweep and inverse deltas used were 1×10^{7} and 1×10^{12} respectively. The genomic group or sub-genomic group variable was set as the dependent

variable while the components (consisting of scores) from the separate PCAs for each were set as the predictor variables.

Support Vector Machine (SVM): Type 1 SVM was used with training constants capacity of ten. The kernel type used was radial basis function (RBF) with a degree of 3 and gamma of 0.333. The maximum number of iterations was set at 1000 with 0.001 as the stoppage error.

Automated Artificial Neural Network: Artificial neural networks search was done using multiple layer perceptron (MLP) and Radial Basis Functions (RBF). The error functions used were set to both sum of squares and cross entropy. The same functions were set for both the hidden neurons and the output neurons, which were identity, logistic, Tanh, and exponential.

For all the classification analyses, the accuracy of prediction was determined from the resultant classification as the percentage of total correctly predicted samples to the total number of samples. The resultant training/calibration models was saved as Predictive Model Markup Language (PMML). To validate the results, the PMML was applied to the test sample component scores to obtain test predictive accuracy.

3. Results and discussion

3.1 Genomic group and sub-genomic group elemental composition in banana flour

The element content in the unripe banana flour, for the sub-genome groups and genome groups, are presented in Table 2 and 3, respectively. The elemental content of the unripe banana flour varied significantly ($P < 0.05$) with banana genomic and sub-genomic groups. The average element concentration, from the 33 varieties of banana, ranked in the decreasing order as: K > N > Mg > P > Ca > Fe > Mn > B > Zn > Cu. The ranking of the element content in this study was similar to the results obtained by do Prado Ferreira et al. (2020) study (Mg > Ca > Fe ≥ Mn > Zn > Cu), Alkarkhi et al. (2009) (K > Mg > Ca), and in accordance with rankings in Hardisson, Rubio, Baez, Martin, Alvarez, and Diaz (2001) to varying extents. The ranking in the present research was partially different from those recently reported by Devarajan et al. (2021) for ripe banana flour

probably due to the differences in agronomic conditions, varieties utilized, and the ripening of the bananas in their study.

The high K content in the present study is in accordance with the fact that bananas have been reported to be rich in K (Alkarkhi et al., 2009; Devarajan et al., 2021; Hardisson et al., 2001). The K content ranged from 1150 – 1400 mg/100 g for genome groups, and from 772 – 1550 mg/100 g for sub-genome groups (Table 2 and 3, respectively). The genome group AAB presented the lowest K content, and the genome group AAAB presented the highest content. While the sub-genome group 'Sukali Ndizi' presented the lowest K concentration, and the sub-genome group 'Synthetic hybrid' presented the highest content. The genome group presented the most accurate ranges when compared to the ranges obtained for the specific varieties. The range of K contents was comparable to that reported in Brazil (372 – 534 mg/100 g) (Hardisson et al., 2001), Malaysia (543 – 803 mg/100 g) (Alkarkhi et al., 2009), India (209 – 1290 mg/100 g) (Devarajan et al., 2021), and South Africa (905 – 970 mg/100 g) (Anyasi, Jideani, & Mchau, 2018). The K content in unripe banana flour used in this study was between 22 - 44 % of the RDA (WHO, 2012a, 2012b), based on 100 g of flour.

Nitrogen is a fundamental component of amino acids (Tessari, 2006) and had the second highest average element content available in the unripe banana flour. The N content ranged from 206 – 519 mg/100 g for genome groups, and from 108 – 641 mg/100 g for sub-genome groups. The genome group AA presented the lowest N content, and the genome group AAAB presented the highest content. While the sub-genome group 'Sukali Ndizi' presented the lowest N concentration, and the sub-genome group 'Pome synthetic hybrid' presented the highest concentration. The N content in the unripe banana flours was equivalent to protein content of 0.6 - 4 %, weight per weight (w/w). Banana fruit protein contents within the range of 3.2 - 5.5 % have been reported by other researchers (Campuzano, Rosell, & Cornejo, 2018; Rachman, Brennan, Morton, & Brennan, 2021; Rodríguez-Ambriz, Islas-Hernández, Agama-Acevedo, Tovar, & Bello-Pérez, 2008; Suntharalingam & Ravindran, 1993). Hoffmann Sardá et al. (2016) reported

that protein contents between 3 and 6 % are characteristic of unripe banana flour, indicating a typical composition for the flours in the present research.

The Mg content ranged from 120 – 140 mg/100 g based on genome groups, and from 103 – 145 mg/100 g based on sub-genome groups. The genome group AAB presented the lowest K concentration, and the genome group AAAB presented the highest concentration. While the sub-genome group 'Sukali Ndizi' presented the lowest K concentration, and the sub-genome group 'Pome synthetic hybrid' presented the highest concentration. The Mg concentration determined in the present study were within the ranges reported by several other researchers (Anyasi et al., 2018; do Prado Ferreira et al., 2020; Hardisson et al., 2001; Netshiheni, Omolola, Anyasi, & Jideani, 2019; Wall, 2006). The Mg content in unripe banana flour in this current study was between 47 - 66 % of the RDA (FAO/WHO, 2001), based on 100 g of flour. This indicated that the genome group AAAB and the sub-genome groups 'Goldfinger', 'Red' and 'Pome synthetic hybrid' (not significantly different on $p > 0.05$) could be potential sufficient sources of Mg.

The P content ranged from 67 – 92 mg/100 g for genome groups, and from 47 – 153 mg/100 g for sub-genome groups. The genome group AAA presented the lowest P concentration, and the genome group AAB presented the highest concentration. While the sub-genome group 'Sukali Ndizi' presented the lowest P concentration, and the sub-genome group 'Mysore' presented the highest concentration. Phosphorus functions as a constituent of bones, teeth, adenosine triphosphate (ATP), phosphorylated metabolic intermediates and nucleic acids (Soetan, Olaiya, & Oyewole, 2010). The P content in the unripe banana flour used in this study was comparable to those reported for different varieties grown in other parts of the world, such as 64 mg/100 g for *M. paradisiaca* L (Hardisson et al., 2001); 13 – 87 mg/100 g for Indian accessions (Devarajan et al., 2021); and 66 - 107 mg/100 g for three indigenous varieties and one commercial variety (Williams) grown in South Africa (Anyasi et al., 2018). The P content in unripe banana flour used in this study was between 7 - 22 % of the RDA (FAO/WHO, 2019b), based on 100 g of flour.

Calcium is essential in the formation of strong bones and teeth, body development, cell metabolism, heart function, and for blood clotting (Dotto, Matemu, & Ndakidemi, 2019). The Ca concentration ranged from 18 – 26 mg/100 g for genome groups, and from 13 – 34 mg/100 g for sub-genome groups. The genome group AA presented the lowest Ca concentration, and the genome group AAAB presented the highest concentration. While the sub-genome group 'Inarnibal' presented the lowest Ca concentration, and the sub-genome group 'Pome synthetic hybrid' presented the highest concentration. The Ca content in the unripe banana flour for the present research was comparable to that reported in recent studies, such as 28 – 398 mg/100 g for Indian accessions (Devarajan et al., 2021), a maximum of 74.7 mg/100 g for *M. paradisiaca L* (do Prado Ferreira et al., 2020) and 6 - 95 mg/100 g for South African grown indigenous varieties (Anyasi et al., 2018). The Ca content in unripe banana flour used in this study was between 1 - 5 % of the RDA (FAO/WHO, 2001), based on 100 g of flour.

The Fe concentration ranged from 2.26 – 2.88 mg/100 g for genome groups, and from 2.09 – 3.53 mg/100 g for sub-genome groups. The genome group AA presented the lowest Fe concentration, and the genome group AAB presented the highest concentration. While the sub-genome group 'Red' presented the lowest Fe concentration, and the sub-genome group 'Mysore' presented the highest concentration. The sub-genome groups 'Mysore' and 'Synthetic hybrid' with the highest Fe concentration showed no significant differences ($p > 0.05$). The Fe content determined in this current study were comparable to those published by Devarajan et al. (2021) (0.70 - 45 mg/kg), do Prado Ferreira et al. (2020) (2.57 - 8.98 mg/100 g for *M. paradisiaca* L), and Pillay and Fungo (2016) 0.06 – 42 mg/100 g for East African accessions. The Fe content in unripe banana flour used in this study was between 4 - 7 % of the RDA (FAO/WHO, 2002), based on 100 g of flour.

The composition of the trace elements Mn, Zn, B, and Cu in unripe banana flour were also comparable in varying degrees to those reported in other studies done in Brazil (do Prado Ferreira et al., 2020), India (Devarajan et al., 2021), and East Africa (Pillay et al., 2016). The genome group

AAB presented the highest concentration for the elements Zn and B, while the genome groups AAAB and AAA presented the highest concentration for the elements Mn and Cu, respectively. The sub-genome groups 'Goldfinger', 'Pome synthetic hybrid', 'Inarnibal', and 'Red' presented the highest concentration for the elements Mn, Zn, B, and Cu, respectively. The genome group AA presented the lowest concentration for the elements Mn and Zn, while the genome groups AAA and AAB presented the lowest concentration for the elements B and Cu, respectively. The sub-genome group 'Sukali Ndizi' presented the lowest concentration for the elements Zn and Cu, while the sub-genome groups 'Sucrier' and 'Lakatan" presented the lowest concentration for the elements Mn and B, respectively. The elements Zn, Cu and Mn unripe banana flour ranged between 6 – 13 %, 5 – 22 % and 31 – 78 % of the RDA, respectively. The genomic group AAAB with a Mn content at 78 % of the RDA could be considered sufficient source of Mn relative to other genomic groups. The element B is essential for the development of healthy bones, however there is no RDA since it has not been assigned an essential biological role. Therefore, people consume varying amounts depending on their diet, whereby boron-rich diets contain roughly 3.25 mg/2000 kilocalories of B per day (Pizzorno, 2015).

A study by Hardisson et al. (2001) focusing on 60 randomly selected samples of *M. acuminata* bananas, from two different locations in Brazil, showed that location influenced the banana fruit element content. Forster et al. (2002)'s study on 95 varieties of 'Gran enana' and 'Pequena enana' from Brazil and Equador further showed that the geographic region in which the varieties are grown affected the element content of the banana fruits. The elemental content of bananas maybe also be affected by the stage of fruit development or degree of maturity according to Wadud and Absar (1996) and Goswami and Borthakur (1996). These factors could possibly explain the variation of the elemental content in the present study. However, given the unripe banana flour in the present study was prepared from 33 banana varieties grown on the farm orchard, under the same agronomic conditions and at the same level of maturity, it could be concluded that the observed variation in elemental contents with each banana variety was probably

due to intrinsic genetic differences. The differences would lead to variation in the sequestering and utilization of the different elements based on the sub-genomic or genomic group. This would imply that classification of the banana sub-genomic and genomic groups could be possible based on their multi-elemental fingerprints.

3.2 Chemometrics: Elemental fingerprint multivariate pattern recognition and fingerprint-based classification

3.2.1 Principal component analysis

In the present study, the genomic group based PCA yielded six principal components, with a cumulative R^2 value of 0.789, thereby explaining 78.9 % of the variation in the data, with the first three components contributing 53.2 % (Fig. 1a). The genomic groups were not clearly separated on the PCA scores plot with some apparent clustering mostly around the center (Fig. 1a). PCA is a critical tool for performing exploratory analysis through non-supervised pattern recognition, particularly in the initial stages of a multivariate study, to gain an overview of data and identify patterns in the data (Benabdelkamel et al., 2012; Cheajesadagul et al., 2013). The clustering observed indicated that further application of an appropriate classification method such as LDA, SVM or ANN, was necessary to ensure potential distinguishing of the genomic groups.

A variable is considered a significant contributor to a given principal component if its loading absolute value is ≥ 0.5 (Alkarkhi et al., 2009). The elements with loadings > 0.5 on PC1 were N (0.599), Cu (0.584), Zn (0.564), Mg (0.525), and Ca (0.500) (Appendix Table A2), for the genomic group based PCA. For PC2, Cu (-0.584) and P (0.520) had recognizable loadings (> 0.5), while only Fe (0.749) had loading > 0.5 on PC3 (Fig. 2a). The next three components had only one specific element contributing (≥ 0.5) each. (Appendix Table A2).

When a given variable's modeling power is closer to one, it is considered more relevant to the model, while a variable with lower power is considered of reduced relevance. For the genomic

group based PCA, the variable power values of the elements decreased in the order: Ca (0.853) > P (0.845) > N (0.801) > Cu (0.797) > Fe (0.767) > B (0.747) > K (0.737) > Zn (0.704) > Mg (0.698) > Mn (0.438). This implied that the elements Ca, P, and N contributed most to the unsupervised pattern recognition for distinguishing within the genome groups.

The separate sub-genomic group based PCA gave nine principal components that explained 78.2 % of the variation in the data. The higher number of components (9) compared to the genomic group PCA (6), indicated a more complex model space with reduced influence of each specific element. This suggested that the multi-elemental fingerprint was more effective at defining the genomic group PCA model space than the sub-genomic group PCA model space. The first three PCs of the sub-genomic group based PCA explained 38.9 % of the variation in the data (Fig. 1b). Some sub-genomic groups were however well separated on the PCA loadings plot (Fig. 1b). To different extents, clouds corresponding to the different genomic groups of Cavendish (C), Mysore (M), Inarnibal (I) and Synthetic hybrid (SH) sub-genomic groups were separated from other sub-genomic groups (Fig. 1b). This possibly implied that the element nutrient requirements of these sub-genomic groups are different from those of other sub-genomic groups to varying extents although not distinctly. The sub-genomic group based PCA showed that the elements P (0.793), K (0.580), and Zn (0.523) contributed most to PC1 (Fig. 2b) with loadings > 0.5. The second component had important loadings from the elements N (0.664), Ca (0.654), and Mg (0.536) (Fig. 2b). The general trend was that the latter components (7, 8, 9) had no relevant (> 0.5) loadings due to the elements, while the foremost components (1 - 5) had several relevant elements with loadings (Appendix Table A2). The elements modeling power decreased in the order: Ca (0.889) > P (0.880) > N (0.835) > Zn (0.812) > K (0.781) > Cu (0.774) > Fe (0.720) > B (0.625) > Mg (0.517). It was notable that the elements Ca, P, and N had the highest power for the sub-genomic group based PCA like the genomic group based PCA, hence emphasizing the importance of the three elements at both sub-genomic and genomic level.

These results, based on the separate genomic and sub-genomic group based PCA, implied that the elements Ca, P and N offered the main points of variation between the genomic and sub-genomic groups, even though K was the most abundant element in the unripe banana flours. Therefore, the soil mineral/element requirements of Ca, P, and N for banana genomic and sub-genomic groups need to be considered in agronomic settings and breeding programs. The application of fertilizers such as "calcium ammonium nitrate" (CAN) and "nitrogen, phosphorus, potassium" (NPK) in cases of limited soil contents of Ca, P, and N may have to be adapted to specific banana genomic and sub-genomic groups. Future research should be considered in the differences in response to these specific Ca, P, and N containing fertilizers for bananas from different genomic and sub-genomic groups. The present results indicated relatively limited variable importance for the elements Zn and Fe at their genomic group and sub-genomic group level. This was in accordance with Davey, Van den Bergh, Markham, Swennen, and Keulemans (2009) who showed limited intra-genomic group variation in Zn and Fe levels on different (171) screened genotypes. The present results were also in accordance with Gibert et al. (2009) who reported that Ca significantly varied in 23 unripe cultivated varieties of Colombian Musaceae.

3.2.2 Multivariate multi-elemental fingerprint-based classification

The PCA-LDA, PCA-ANN and PCA-SVM classification results of the training/calibration and test/validation set samples are presented in Table 4. The results showed that PCA-SVM was a more effective classification method compared to PCA-ANN, and PCA-LDA, given it led to higher accuracy values obtained for the genomic and sub-genomic level based classifications. The present research is the first to attempt to utilize these methods to assess the efficacy of multi-elemental fingerprints in the classification of banana genomic groups, and sub-genomic groups. In the present approach, the underlying features (components; fingerprints) of the multi-elemental

data, were first extracted using PCA, and then applied to the pure classification methods (SVM, ANN, and LDA).

A training/calibration-set-only based classification model may have high accuracy due to the calibration set samples' specific characteristics and hence over specify the model such that when applied to new data, the model fails. Studies have shown that the classification (PCA-LDA) of other crops using multi-elemental fingerprint is possible for pistachio nuts (Esteki et al., 2021), cowpeas (Pérez-Rodríguez et al., 2019) and tea (Chen et al., 2009). However, in the present study, multivariate analysis classification of varieties was not possible due to limited replication at variety level.

Accurate classification prediction was well transferred into the test/verification sets for sub-genomic group (Table 4). Accuracies of 78.95 %, 94.87 % and 100 % for the sub-genomic group test sets were achieved by the PCA-LDA, PCA-ANN and PCA-SVM models, respectively. The prediction of genomic groups was relatively more accurate than the sub-genomic group, with all the classification methods yielding 100 % prediction accuracy. This trend was in in contrast to the PCA results (Figure 1a) and indicated the efficacy of the further classification analyses in enabling the distinguishing of the genomics groups. The results further showed that multi-elemental fingerprint was more efficacious for the genomic group than the sub-genomic group least effective for the varieties. These results however differed from those recently reported by Devarajan et al. (2021), which indicated that the sub-genomic and genomic groups may not be classified accurately based on the multi-elemental fingerprint of ripe banana flour. The difference in the present results their results could probably have arisen from the differences in chemometrics data analysis approaches utilized or the difference in the sample types between the two studies. Given the multi-elements in bananas maybe affected by the geographical region and agronomic conditions, sample preparation of methods (Esteki et al., 2021; Maione & Barbosa, 2019), studies similar to the present study, using unripe banana flour from bananas grown under different agronomic and geographical aspects are however necessary to support the results obtained. In

addition, further studies are necessary with appropriate replication, for determining the efficacy of the multi-elemental fingerprint in the classification of banana varieties. The results in the present research suggested that the content of minerals in the soil is more important for bananas of different genomic groups as opposed to sub-genomic groups (lower PC components, and increased classification prediction). The results also suggested that during banana breeding, consideration should be given for the mineral content of the growing soil or medium. A given new hybrid combination of AA and BB may exhibit a specific phenotype/element fingerprint relative to another due to influence by its novel genetic makeup in the suitability to the specific soil mineral content. The results also imply that the elemental content of a banana hybrids can be quite different from that of the initial banana genomic group and sub-genomic group when cultivated under the same agronomic conditions.

4. Conclusion

The unripe banana flour multi-elemental fingerprints were successful facilitating and distinguishing between different banana sub-genomic and genomic groups of bananas grown under the same agronomic conditions through chemometrics based on a combination of PCA with LDA, SVM, and ANN. The chemometrics classification method applied influenced the success rate of the classification, particularly on test/verification set samples. The verification set prediction accuracy rate of the PCA-SVM method was consistently high, exhibiting a better predictive ability than PCA-ANN and PCA-LDA. The application of multi-elemental fingerprint for classification was more effective at genomic group level than sub-genomic group level. Further studies are required based on unripe banana flour from different geographic regions, more varieties and from bananas grown under different agronomic practices.

CRediT authorship contribution statement

Kayise Hypercia Maseko: Experimental analysis, Writing – review & editing. **Thierry Regnier:** Supervision, Writing – review & editing. **Tonna Anyasi:** Writing – review & editing. **Belinda Du Plessis:** Supervision, Writing – review & editing. **Laura Suzzanne Da Silva:** Writing – review & editing. **Funso Kutu:** Writing – review & editing. **Obiro Cuthbert Wokadala:** Conceptualization, Methodology, Data analysis, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is financially supported by the National Research Foundation (NRF) of South Africa [grant number 116308]. Dr Mieke Stephanie Daneel for advice on the varieties cultivated at ARC-TSC gene bank. Dr Elize Jooste for administrative and management support of the project at ARC-TSC. Ms. Aylward Elsabe and Yvette Abercrombie for technical assistance on element analysis.

Appendix A. Supplementary data

References

Alkarkhi, A. F., Ramli, S. B., & Easa, A. M. (2009). Application of multivariate statistical techniques for differentiation of ripe banana flour based on the composition of elements. *International journal of food sciences and nutrition, 60*(sup4), 116-125. https://doi.org/https://doi.org/10.1080/09637480802609368.

Anyasi, T. A., Jideani, A. I., & Mchau, G. R. (2018). Phenolics and essential mineral profile of organic acid pretreated unripe banana flour. *Food Research International, 104*, 100-109. https://doi.org/https://doi.org/10.1016/j.foodres.2017.09.063.

Benabdelkamel, H., Di Donna, L., Mazzotti, F., Naccarato, A., Sindona, G., Tagarelli, A., & Taverna, D. (2012). Authenticity of PGI "Clementine of Calabria" by multielement fingerprint. *Journal of Agricultural and Food Chemistry, 60*(14), 3717-3726. https://doi.org/https://doi.org/10.1021/jf201556e.

Borborah, K., Saikia, D., Rehman, M., Islam, M. A., Mahanta, S., Chutia, J., . . . Tanti, B. (2020). Comparative Analysis of Genetic Diversity in Some Non-commercial Cultivars of Musa L. from Assam, India, Using Morphometric and ISSR Markers. *International Journal of Fruit Science, 20*(sup2), 1814-1828. https://doi.org/https://doi.org/10.1080/15538362.2020.1772178.

Brisibe, E. A., Ubi, G. M., & Ekanem, N. G. (2021). Descriptive and multivariate analyses of morphotaxonomic and yield-related traits in inflorescence dichotomous cultivars of Musa species (AAB genome). *Genetic Resources and Crop Evolution*, 1-16. https://doi.org/https://doi.org/10.1007/s10722-021-01193-6.

Brown, A., Tumuhimbise, R., Amah, D., Uwimana, B., Nyine, M., Mduma, H., . . . Swennen, R. (2017). Bananas and plantains (Musa spp.). In *Genetic improvement of tropical crops* (pp. 219-240): Springer.

Campuzano, A., Rosell, C. M., & Cornejo, F. (2018). Physicochemical and nutritional characteristics of banana flour during ripening. *Food chemistry, 256*, 11-17. https://doi.org/https://doi.org/10.1016/j.foodchem.2018.02.113.

Cheajesadagul, P., Arnaudguilhem, C., Shiowatana, J., Siripinyanond, A., & Szpunar, J. (2013). Discrimination of geographical origin of rice based on multi-element fingerprinting by high resolution inductively coupled plasma mass spectrometry. *Food chemistry, 141*(4), 3504-3509.

Chen, Y., Yu, M., Xu, J., Chen, X., & Shi, J. (2009). Differentiation of eight tea (Camellia sinensis) cultivars in China by elemental fingerprint of their leaves. *Journal of the Science of Food and Agriculture, 89*(14), 2350-2355.

Cruz-Cárdenas, C., Youssef, M., & Escobedo-Graciamedrano, R. (2017). Assessment of genetic relationship in Musa using male flower descriptors and molecular markers. *South African Journal of Botany, 113*, 270-276.

Davey, M. W., Van den Bergh, I., Markham, R., Swennen, R., & Keulemans, J. (2009). Genetic variability in Musa fruit provitamin A carotenoids, lutein and mineral micronutrient contents. *Food chemistry, 115*(3), 806-813.

Devarajan, R., Jayaraman, J. K., Somasundaram, S. M., Ragupathy, S., Raman, P., Sathiamoorthy, K., & Subbaraya, U. (2021). Genetic diversity in fresh fruit pulp mineral profile of 100 Indian Musa accessions. *Food chemistry, 361*, 130080.

do Prado Ferreira, M., & Tarley, C. R. T. (2020). Assessment of in vitro bioacessibility of macrominerals and trace elements in green banana flour. *Journal of Food Composition and Analysis, 92*, 103586.

Dotto, J., Matemu, A. O., & Ndakidemi, P. A. (2019). Nutrient composition and selected physicochemical properties of fifteen Mchare cooking bananas: A study conducted in northern Tanzania. *Scientific African, 6*, e00150.

Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC Trends in Analytical Chemistry, 40*, 38-51. https://doi.org/https://doi.org/10.1016/j.trac.2012.08.003.

El-Khishin, D. A., Belatus, E. L., El-Hamid, A., & Radwan, K. H. (2009). Molecular characterization of banana cultivars (Musa spp.) from Egypt using AFLP. *Research Journal of Agriculture Biological Sciences, 5*(3), 272-279.

Esteki, M., Heydari, E., Simal-Gandara, J., Shahsavari, Z., & Mohammadlou, M. (2021). Discrimination of pistachio cultivars based on multi-elemental fingerprinting by pattern recognition methods. *Food Control, 124*, 107889.

FAO/WHO. (2001). Human Vitamin and Mineral Requirements. Retrieved from: http://www.fao.org/3/y2809e/y2809e.pdf Accessed 17 May 2021.

FAO/WHO. (2002). Human Vitamin and Mineral Requirements. Retrieved from: http://www.fao.org/3/y2809e/y2809e0j.htm#bm19 Accessed 18 May 2021.

FAO/WHO. (2019a). Codex nutrient reference values. Retrieved from: http://www.fao.org/3/ca6969en/CA6969EN.pdf Accessed 18 May 2021.

FAO/WHO. (2019b). Codex nutrient reference values for the Food and Agriculture Organization of the United Nations or World Health Organization. Retrieved from: https://www.fao.org/3/ca6969en/CA6969EN.pdf Accessed.

Fechner, D. C., Hidalgo, M. J., Ruiz Díaz, J. D., Gil, R. A., & Pellerano, R. G. (2020). Geographical origin authentication of honey produced in Argentina. *Food Bioscience, 33*, 100483. https://doi.org/https://doi.org/10.1016/j.fbio.2019.100483.

Forster, M. P., Rodríguez, E. R., Martín, J. D., & Romero, C. D. (2002). Statistical differentiation of bananas according to their mineral composition. *Journal of Agricultural and Food Chemistry, 50*(21), 6130-6135.

Ghag, S. B., & Ganapathi, T. R. (2017). Genetically modified bananas: To mitigate food security concerns. *Scientia horticulturae, 214*, 91-98.

Gibert, O., Dufour, D., Giraldo, A., Sanchez, T., Reynes, M., Pain, J.-P., . . . Diaz, A. (2009). Differentiation between cooking bananas and dessert bananas. 1. Morphological and compositional characterization of cultivated Colombian Musaceae (Musa sp.) in relation to consumer preferences. *Journal of Agricultural and Food Chemistry, 57*(17), 7857-7869.

Goswami, B., & Borthakur, A. (1996). Chemical and biochemical aspects of developing culinary banana (Musa ABB)'Kachkal'. *Food chemistry, 55*(2), 169-172.

Hardisson, A., Rubio, C., Baez, A., Martin, M., Alvarez, R., & Diaz, E. (2001). Mineral composition of the banana (Musa acuminata) from the island of Tenerife. *Food chemistry, 73*(2), 153-161.

Herreros-Chavez, L., Oueghlani, F., Morales-Rubio, A., Cervera, M. L., & de la Guardia, M. (2019). Mineral profiles of legumes and fruits through partial least squares energy dispersive X-ray fluorescence. *Journal of Food Composition and Analysis, 82*, 103240. https://doi.org/https://doi.org/10.1016/j.jfca.2019.103240.

Hoffmann Sardá, F. A., de Lima, F. N. R., Lopes, N. T. T., Santos, A. d. O., Tobaruela, E. d. C., Kato, E. T. M., & Menezes, E. W. (2016). Identification of carbohydrate parameters in commercial unripe banana flour. *Food Research International, 81*, 203-209. https://doi.org/https://doi.org/10.1016/j.foodres.2015.11.016.

Igwe, D. O., Ihearahu, O. C., Osano, A. A., Acquaah, G., & Ude, G. N. (2021). Assessment of genetic diversity of Musa species accessions with variable genomes using ISSR and SCoT markers. *Genetic Resources and Crop Evolution*, 1-22.

Karamura, D. A., Karamura, E., & Tinzaara, W. (2012). Banana cultivar names, synonyms and their usage in Eastern Africa. *Bioversity International, Uganda*, 1-122.

Kment, P., Mihaljevič, M., Ettler, V., Šebek, O., Strnad, L., & Rohlová, L. (2005). Differentiation of Czech wines using multielement composition – A comparison with vineyard soil. *Food chemistry, 91*(1), 157-165. https://doi.org/https://doi.org/10.1016/j.foodchem.2004.06.010.

Kongolo, J. I., Da Silva, L. S., Wokadala, O. C., Du Plessis, B., Husselman, J., Ngcobo, M. E., . . . Daneel, M. (2017). Pasting, thermal, gel texture, resistant starch and colour properties of unripe banana flour from 10 desert banana varieties cultivated in South Africa. *Journal of Food Measurement and Characterization, 11*(3), 1056-1064.

Kruszewski, B., & Obiedziński, M. W. (2018). Multivariate analysis of essential elements in raw cocoa and processed chocolate mass materials from three different manufacturers. *LWT, 98*, 113-123. https://doi.org/https://doi.org/10.1016/j.lwt.2018.08.030.

Maione, C., & Barbosa, R. M. (2019). Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review. *Critical reviews in food science and nutrition, 59*(12), 1868-1879.

Mir-Marqués, A., Martínez-García, M., Garrigues, S., Cervera, M. L., & de la Guardia, M. (2016). Green direct determination of mineral elements in artichokes by infrared spectroscopy and X-ray fluorescence. *Food chemistry, 196*, 1023-1030. https://doi.org/https://doi.org/10.1016/j.foodchem.2015.10.048.

Nelson, S. C., Ploetz, R. C., & Kepler, A. K. (2006). Musa species (banana and plantain). *Species profiles for Pacific Island agroforestry, 15*(2), 251-259.

Netshiheni, R. K., Omolola, A. O., Anyasi, T. A., & Jideani, A. I. (2019). Banana bioactives: absorption, utilisation and health benefits. In *Banana Nutrition-Function and Processing Kinetics*: IntechOpen.

Nyombi, K. (2020). Chapter 44 - Diagnosis and management of nutrient constraints in bananas (Musa spp.). In A. K. Srivastava & C. Hu (Eds.), *Fruit Crops* (pp. 651-659): Elsevier.

Oyeyinka, B. O., & Afolayan, A. J. (2019). Comparative evaluation of the nutritive, mineral, and antinutritive composition of *Musa sinensis* L.(Banana) and *Musa paradisiaca* L.(Plantain) fruit compartments. *Plants, 8*(12), 598. https://doi.org/https://doi.org/10.3390/plants8120598.

Paltridge, N. G., Milham, P. J., Ortiz-Monasterio, J. I., Velu, G., Yasmin, Z., Palmer, L. J., . . . Stangoulis, J. C. (2012). Energy-dispersive X-ray fluorescence spectrometry as a tool for zinc, iron and selenium analysis in whole grain wheat. *Plant and Soil, 361*(1), 261-269.

Pereira, A., & Maraschin, M. (2015). Banana (Musa spp) from peel to pulp: Ethnopharmacology, source of bioactive compounds and its relevance for human health. *Journal of Ethnopharmacology, 160*, 149-163. https://doi.org/https://doi.org/10.1016/j.jep.2014.11.008.

Pérez-Rodríguez, M., Gaiad, J. E., Hidalgo, M. J., Avanza, M. V., & Pellerano, R. G. (2019). Classification of cowpea beans using multielemental fingerprinting combined with supervised learning. *Food Control, 95*, 232-241.

Pillay, M., & Fungo, R. (2016). Diversity of iron and zinc content in bananas from East and Central Africa. *HortScience, 51*(4), 320-324.

Pizzorno, L. (2015). Nothing boring about boron. *Integrative Medicine: A Clinician's Journal, 14*(4), 35.

Rachman, A., Brennan, M. A., Morton, J., & Brennan, C. S. (2021). Starch Pasting Properties, and the Effects of Banana Flour and Cassava Flour Addition to Semolina Flour on Starch and Amino Acid Digestion. *Starch - Stärke, 73*(1-2), 2000137. https://doi.org/https://doi.org/10.1002/star.202000137.

Rea, A., & Rea, W. (2016). How many components should be retained from a multivariate time series PCA? *arXiv preprint arXiv:1610.03588*.

Rodríguez-Ambriz, S. L., Islas-Hernández, J. J., Agama-Acevedo, E., Tovar, J., & Bello-Pérez, L. A. (2008). Characterization of a fibre-rich powder prepared by liquefaction of unripe banana flour. *Food chemistry, 107*(4), 1515-1521. https://doi.org/https://doi.org/10.1016/j.foodchem.2007.10.007.

Ruggiero, L., Fontanella, M. C., Amalfitano, C., Beone, G. M., & Adamo, P. (2021). Provenance discrimination of Sorrento lemon with Protected Geographical indication (PGI) by multi-elemental fingerprinting. *Food chemistry, 362*, 130168. https://doi.org/https://doi.org/10.1016/j.foodchem.2021.130168.

Soetan, K., Olaiya, C., & Oyewole, O. (2010). The importance of mineral elements for humans, domestic animals and plants-A review. *African journal of food science, 4*(5), 200-222.

Sogo-Temi, C., Idowu, O., & Idowu, E. (2014). Effect of biological and chemical ripening agents on the nutritional and metal composition of banana (Musa spp). *Journal of Applied Sciences and Environmental Management, 18*(2), 243-246.

Srivastava, A. K., & Hu, C. (2019). *Fruit Crops: Diagnosis and Management of Nutrient Constraints*: Elsevier.

Suarez-Tapia, A., Kucheryavskiy, S. V., Christensen, B. T., Thomsen, I. K., & Rasmussen, J. (2017). Limitation of multi-elemental fingerprinting of wheat grains: effect of cultivar, sowing date, and nutrient management. *Journal of cereal science, 76*, 76-84.

Sulaiman, S. F., Yusoff, N. A. M., Eldeen, I. M., Seow, E. M., Sajak, A. A. B., & Ooi, K. L. (2011). Correlation between total phenolic and mineral contents with antioxidant activity of eight Malaysian bananas (Musa sp.). *Journal of Food Composition and Analysis, 24*(1), 1-10.

Suntharalingam, S., & Ravindran, G. (1993). Physical and biochemical properties of green banana flour. *Plant Foods for Human Nutrition, 43*(1), 19-27. https://doi.org/10.1007/BF01088092.

Szymczycha-Madeja, A., Welna, M., & Pohl, P. (2015). Determination of essential and non-essential elements in green and black teas by FAAS and ICP OES simplified – multivariate classification of different tea products. *Microchemical Journal, 121*, 122-129. https://doi.org/https://doi.org/10.1016/j.microc.2015.02.009.

Tessari, P. (2006). Nitrogen balance and protein requirements: definition and measurements. In *Cachexia and wasting: a modern approach* (pp. 73-79): Springer.

Tsomele, G. F., Venter, E., Wokadala, O. C., Jooste, E., Dlamini, B. C., Ngobese, N. Z., & Siwela, M. (2021). Structural (gross and micro), physical and nutritional properties of Trichilia emetica and Trichilia dregeana seeds. *CyTA-journal of Food, 19*(1), 483-492.

Valmayor, R., Jamaluddin, S., Silayoi, B., Kusumo, S., Danh, L., Pascua, O., & Espino, R. (2000). *Banana cultivar names and synonyms in Southeast Asia*.

Wadud, S., & Absar, N. (1996). A comparative study on the mineral content of four varieties of banana at different maturity stages. *Indian Journal of Plant Physiology, 1*, 295-297.

Wall, M. M. (2006). Ascorbic acid, vitamin A, and mineral composition of banana (Musa sp.) and papaya (Carica papaya) cultivars grown in Hawaii. *Journal of Food Composition and Analysis, 19*(5), 434-445.

WHO. (2012a). Potassium intake for adults and children. *Geneva: World Health Organization (WHO), 48*.

WHO. (2012b). Potassium intake for adults and children. Retrieved from: https://www.who.int/nutrition/publications/guidelines/potassium_intake_printversion.pdf Accessed 17/05/2021 2021.

Winkler, A., Rauwolf, M., Sterba, J. H., Wobrauschek, P., Streli, C., & Turyanskaya, A. (2020). Total reflection X-ray fluorescence analysis of elemental composition of herbal infusions and teas. *Journal of the Science of Food and Agriculture, 100*(11), 4226-4236. https://doi.org/https://doi.org/10.1002/jsfa.10463.

Table captions:

Table 1

Genomic and sub-genomic groups of the 33 banana varieties of bananas that were utilized for preparation of unripe banana flour for discrimination of Musa banana genomic and sub-genomic groups based on multi-elemental fingerprints and chemometrics

Variety	Genome group	Sub-genome group
Khai Thong Ruang	AA	Sucrier
Lady Finger	AA	Sucrier
Pisang Lemak Manis	AA	Inarnibal
Americanii	AAA	Cavendish
B38	AAA	Cavendish
Chinese Cavendish	AAA	Cavendish
E1	AAA	Cavendish
Eldorado	AAA	Cavendish
FOR-916	AAA	Cavendish
Grand Nain	AAA	Cavendish
Green-Red	AAA	Red
IBP 5-B	AAA	Cavendish
J6	AAA	Cavendish
Lakatan	AAA	Lakatan
N6	AAA	Cavendish
Nandi	AAA	Cavendish
NV1	AAA	Cavendish
Poyo	AAA	Cavendish
RSS 3	AAA	Cavendish
Selangor	AAA	Cavendish
Sordwana	AAA	Cavendish
T2	AAA	Cavendish
Williams	AAA	Cavendish
FHIA-18	AAAB	Pome synthetic hybrid
Goldfinger (FHIA-01)	AAAB	Goldfinger selection
PK6	AAAB	Goldfinger selection
PK2 40	AAAB	Goldfinger selection
PKZ	AAAB	Goldfinger selection
FHIA-25	AAB	Synthetic hybrid
Foconah	AAB	Pome
Kam22	AAB	Sukali Ndizi
Pisang Ceylan	AAB	Mysore
Pome	AAB	Pome

Varieties were grown in a gene bank plantation in Burgershall, Mpumalanga province, South Africa.
The bananas were grown on the same plot under same agronomic conditions and management practices in Burgershall.
Chromosomes inherited from *Musa acuminata* Colla has 2n=2x=22 (AA-genome).
Chromosomes inherited from *Musa balbisiana* Colla has 2n=2x=22 (BB-genome).
Genomic and sub-genomic group classifications were based on Promusa (http://www.promusa.org/Banana + cultivar + checklist).
Abbreviations:
Kam22: Kamaramasenge is a synonym of Sukali Ndizi in East Africa.
FHIA: Fundación Hondureña de Investigación Agrícola.

Table 2

Element composition of unripe banana flour for 33 varieties, that was applied for elemental fingerprint-based classification of genomic groups

Genome groups	Element (mg/100 g)									
	K	**N**	**Mg**	**P**	**Ca**	**Fe**	**Mn**	**Zn**	**B**	**Cu**
AA	1280 ± 240(57)^{a}	206 ± 73(17)^{a}	128 ± 12(3)^{a}	80 ± 25(6)^{a}	18 ± 3(0.79)^{a}	2.26 ± 0.23(0.05)^{b}	1.13 ± 0.36(0.09)^{a}	0.75 ± 0.12(0.03)^{a}	0.76 ± 0.19(0.05)^{a}	0.36 ± 0.11(0.02)^{a}
AAA	1270 ± 100(9)^{a}	463 ± 160(15)^{b}	133 ± 13(1)^{a}	67 ± 12(1)^{b}	18 ± 6(0.54)^{a}	2.55 ± 0.36(0.03)^{a}	1.44 ± 0.68(0.06)^{b}	0.79 ± 0.10(0.01)^{a}	0.61 ± 0.26(0.02)^{b}	0.41 ± 0.09(0.01)^{b}
AAAB	1400 ± 160(29)^{c}	519 ± 170(32)^{b}	140 ± 9(2)^{c}	87 ± 20(4)^{a}	26 ± 4(0.77)^{b}	2.50 ± 0.36(0.07)^{a}	2.18 ± 1(0.23)^{a}	0.78 ± 0.15(0.03)^{a}	0.77 ± 0.17(0.03)^{a}	0.36 ± 0.09(0.02)^{a}
AAB	1150 ± 250(47)^{b}	293 ± 180(33)^{a}	120 ± 11(2)^{b}	92 ± 43(8)^{a}	24 ± 4(0.71)^{b}	2.88 ± 0.61(0.11)^{c}	1.52 ± 0.25(0.05)^{a}	0.86 ± 0.13(0.02)^{b}	0.80 ± 0.15(0.03)^{a}	0.34 ± 0.12(0.02)^{a}

Values are presented as means ± standard deviation (standard error of mean) on dry weight basis (n=6).

Within a given factor, different superscript letters within a column indicate statistically significant differences between genome groups with Fisher's least significant (LSD) test ($p < 0.05$).

Table 3

Element composition of unripe banana flour for 33 varieties, that was applied for elemental fingerprint-based classification of sub-genomic groups

Sub-genome group	Element (mg/100 g)									
	K	N	Mg	P	Ca	Fe	Mn	Zn	B	Cu
Sucrier	1180 ± 230(66)^{b}	156 ± 18(5)^{ef}	130 ± 14(4)^{ac}	65 ± 16(5)^{bc}	20 ± 2(0.54)^{cd}	2.17 ± 0.23(0.07)^{b}	0.92 ± 0.25(0.07)^{c}	0.67 ± 0.04(0.01)^{ab}	0.63 ± 0.08(0.02)^{bc}	0.31 ± 0.11(0.03)^{b}
Inarnibal	1490 ± 81(33)^{de}	304 ± 4(2)^{abf}	126 ± 9(4)^{abc}	109 ± 2(0.77)^{e}	13 ± 0.52(0.21)^{a}	2.43 ± 0.10(0.04)^{ab}	1.54 ± 0.04(0.02)^{abc}	0.90 ± 0.00(0.00)^{cd}	1.01 ± 0.02(0.01)^{d}	0.44 ± 0.01(0.00)^{a}
Cavendish	1260 ± 97(9)^{cf}	461 ± 164(16)^{cd}	132 ± 13(1)^{a}	68 ± 12(1)^{c}	17 ± 5(0.49)^{bc}	2.58 ± 0.36(0.03)^{a}	1.38 ± 0.69(0.07)^{a}	0.80 ± 0.09(0.01)^{e}	0.62 ± 0.27(0.03)^{b}	0.40 ± 0.09(0.01)^{a}
Red	1350 ± 26(11)^{ac}	572 ± 22(9)^{dg}	141 ± 5(2)^{d}	56 ± 0.96(0.39)^{ab}	33 ± 0.88(0.36)^{g}	2.09 ± 0.07(0.03)^{b}	1.99 ± 0.18(0.07)^{bd}	0.78 ± 0.05(0.02)^{e}	0.48 ± 0.07(0.03)^{e}	0.44 ± 0.04(0.01)^{a}
Lakatan	1440 ± 16(7)^{ade}	399 ± 11(4)^{abc}	138 ± 3(1)^{ad}	50 ± 1(0.58)^{a}	15 ± 0.39(0.16)^{ab}	2.45 ± 0.17(0.07)^{ab}	1.80 ± 0.06(0.02)^{abd}	0.66 ± 0.02(0.01)^{ab}	0.47 ± 0.05(0.02)^{e}	0.41 ± 0.01(0.00)^{ac}
Pome synthetic hybrid	1340 ± 6(3)^{ac}	641 ± 23(9)^{g}	145 ± 3(1)^{d}	85 ± 4(2)^{d}	34 ± 0.46(0.19)^{g}	2.45 ± 0.05(0.02)^{ab}	1.49 ± 0.01(0.00)^{abc}	1.06 ± 0.04(0.02)^{f}	0.87 ± 0.06(0.02)^{ad}	0.42 ± 0.02(0.01)^{a}
Goldfinger selection	1420 ± 170(35)^{ad}	489 ± 180(37)^{cd}	138 ± 10(2)^{d}	87 ± 23(5)^{d}	24 ± 1(0.33)^{ef}	2.51 ± 0.40(0.08)^{a}	2.35 ± 1(0.27)^{d}	0.71 ± 0.05(0.01)^{b}	0.75 ± 0.18(0.04)^{ac}	0.34 ± 0.10(0.02)^{bc}
Synthetic hybrid	1550 ± 40(17)^{e}	236 ± 54(22)^{aef}	118 ± 5(2)^{b}	129 ± 4(2)^{f}	18 ± 0.74(0.30)^{abc}	3.47 ± 0.11(0.05)^{c}	1.57 ± 0.05(0.02)^{abc}	0.97 ± 0.01(0.00)^{df}	0.76 ± 0.04(0.02)^{abcd}	0.39 ± 0.01(0.00)^{abc}
Pome	1120 ± 60(17)^{b}	336 ± 220(64)^{ab}	125 ± 8(2)^{bc}	65 ± 9(3)^{bc}	27 ± 3(0.72)^{f}	2.39 ± 0.29(0.08)^{ab}	1.56 ± 0.09(0.03)^{abc}	0.87 ± 0.05(0.01)^{c}	0.85 ± 0.22(0.06)^{ad}	0.38 ± 0.13(0.04)^{abc}
Sukali Ndizi	772 ± 6(3)^{g}	108 ± 12(5)^{e}	103 ± 1(0.48)^{e}	47 ± 3(1)^{a}	23 ± 1(0.51)^{de}	2.59 ± 0.27(0.11)^{a}	1.81 ± 0.05(0.02)^{abd}	0.62 ± 0.02(0.01)^{a}	0.75 ± 0.08(0.03)^{abc}	0.15 ± 0.01(0.00)^{d}
Mysore	1190 ± 14(6)^{bf}	449 ± 7(3)^{bcd}	127 ± 5(2)^{abc}	153 ± 3(1)^{g}	23 ± 0.22(0.09)^{def}	3.53 ± 0.55(0.22)^{c}	1.07 ± 0.02(0.01)^{ac}	0.94 ± 0.05(0.02)^{cd}	0.81 ± 0.00(0.00)^{abcd}	0.37 ± 0.01(0.00)^{abc}

Values are presented as means ± standard deviation (standard error of mean) on dry weight basis (n=6).

Within a given element column, the different superscript letters indicate statistically significant differences between the sub-genome groups with Fisher's least significant test (LSD) ($p < 0.05$).

Table 4

Multivariate multi-elemental fingerprint classification accuracy results for the training/calibration and the test/verification set samples of unripe banana flour from 33 varieties.

Analysis sample test	Chemometrics method	Classification accuracy %	
		Sub-genomic group	Genomic group
Training/calibration set	PCA-LDA	100.00	100.00
	PCA-LNN	100.00	100.00
	PCA-SVM	100.00	100.00
Test/verification set	PCA-LDA	78.95	100.00
	PCA-LNN	94.87	100.00
	PCA-SVM	100.00	100.00

Figure Captions:

Fig. 1. Principal Component Analysis (PCA) score scatterplots based on multi-element concentrations of (a) genome groups (b) sub-genome groups

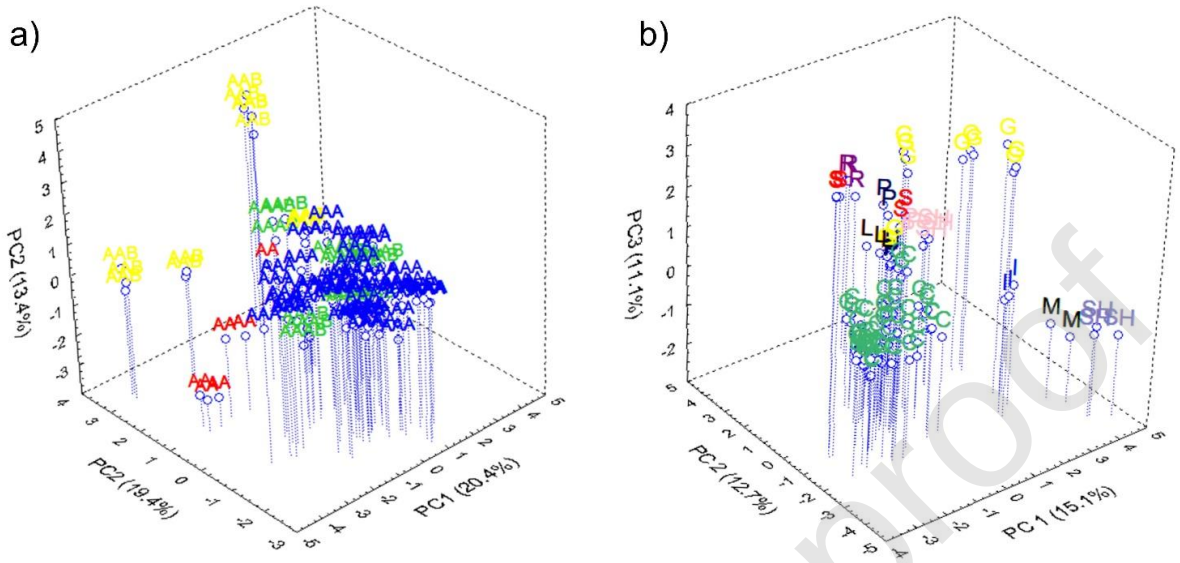


Fig. 2. Principal Component Analysis (PCA) loadings scatterplots based on multi-element concentrations of (a) genome groups and (b) sub-genome groups (c).