Journal of Food Composition and Analysis 131 (2024) 106205

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

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

Genomic and subgenomic group discrimination between 100 Indian banana (Musa) accessions using ripe banana pulp multi-elemental fingerprints and chemometrics

Ramajayam Devarajan [a], Siphosethu R. Dibakoane [b], Obiro Cuthbert Wokadala [b,*], Belinda Meiring [c], Victor Mlambo [b], Funso Raphael Kutu [b], July Johannes Sibanyoni [d], Jeyabaskaran Kandallu Jayaraman [e]

[a] ICAR-Indian Institute of Soil and Water Conservation, Research Centre, Sunabeda, Odisha, India
[b] School of Agricultural and Natural Sciences, University of Mpumalanga, Corner R40 and D725 Road, Nelspruit 1200, South Africa
[c] Tshwane University of Technology, Department of Biotechnology and Food Technology, Private Bag X680, Pretoria 0001, South Africa
[d] School of Hospitality and Tourism Management, University of Mpumalanga, Corner R40 and D725 Road, Nelspruit 1200, South Africa
[e] ICAR-National Research Centre for Banana, Tiruchirapalli, Tamil Nadu, India

ARTICLE INFO

Keywords:
Bananas (*Musa*)
Chemometrics, Multi-elemental Fingerprints
Sub-genome groups
Banana Genome
India banana accessions

ABSTRACT

Worldwide, there are over 1000 banana types which are classified in various subgenomic and genomic groups. Distinguishing between the banana types, their genomic and subgenomic groups has been a challenge due to different identities and nomenclature used in different regions of the world. The present study assessed the efficacy of multi-elemental fingerprinting combined with chemometrics to distinguish between genomic and subgenomic groups within 100 Indian banana (Musa) accessions based on ripe banana pulp elemental concentrations. The concentrations of B, Ca, Fe, Mg, Mn K, Zn, Na, and P were analyzed using Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES). Multi-elemental fingerprints plus chemometrics were done using principal component analysis (PCA) then combined with linear discriminant analysis (PCA-LDA), support vector machine (PCA-SVM), and artificial neural network (PCA-ANN) for classification analysis with an 80:20 split between the calibration and verification sets (with total of 300 specimens). The PCA-SVM model was the most effective in classification when applied to the verification set subgenomic and genomic groups data, with accuracies of 83.7% and 100.0% respectively. These results demonstrated that ripe banana pulp multi-elemental fingerprints combined with chemometrics can discriminate between genomic and sub-genomic groups for Indian banana (Musa) accessions.

1. Introduction

Bananas (*Musa spp.*) rank among the most widely grown and consumed subtropical crops worldwide (FAO, 2020, 2021). They are a critical component of the human diet, thanks to their high nutrient density and widespread availability (Arvanitoyannis & Mavromatis, 2009; FAO, 2020, 2021). However, the prevalence of a diverse nomenclature has made it difficult for taxonomist and horticulturists to develop a standard classification system of bananas. There is a need to develop methods that can effectively classify bananas for the benefit of taxonomists and horticulturists. Today there are over 1000 banana cultivars, spanning more than 50 species and sub-genomic groups (Brown et al., 2017; Srivastava & Hu, 2019). The genomic groups consist of six that occur naturally (AA, AAA, AB, AAB, ABB, and ABBB, with various hybrids genomic groups (El-Khishin et al., 2009; Nyombi, 2020).

The utilization of banana chemical composition to distinguish between bananas of different varieties, subgenomic and genomic groups is complicated by various factors. The location in which bananas are grown, was shown to influence the chemical composition more than the cultivar group by Forster et al. (2002b) and Cano et al. (1997) while it was demonstrated that the subgroups (dessert vs plantain) affected the chemical composition of banana (Gibert et al., 2009). The production system (Ambuko et al., 2006; Nyanjage et al., 2001) and altitude (Bugaud et al., 2006) also significantly also affect the chemical

* Corresponding author.
E-mail address: Obiro.wokadala@ump.ac.za (O.C. Wokadala).

https://doi.org/10.1016/j.jfca.2024.106205
Received 21 October 2023; Received in revised form 25 February 2024; Accepted 22 March 2024
Available online 4 April 2024

composition.

The mineral content of ripe bananas was used to distinguish between Malaysia ripe bananas grown in different areas but from the same genomic group (AAA) using multivariate analysis (Alkarkhi et al., 2009). Contrastingly another study of AAA genomic group ripe bananas from different areas (Brazil-Tenerife and Ecuador) was not able to distinguish between the varieties but was able to distinguish between the regions of origin (Forster et al., 2002a). Ripening was suggested to affect the mineral content of bananas (Ayo-Omogie et al., 2021; Izonfuo and Omuaru, 1988) and could have played a role in the different results observed. With elimination of the possible effects of ripening and growing region, using unripe banana flour from bananas grown on the same orchard in South Africa, it was shown that banana genomic and subgenomic groups could be distinguished (Maseko et al., 2022). However, the process of drying unripe banana flour can hinder the adaptability of the method. Hence it is necessary that ripe bananas are used to assess the efficacy of multi-elemental fingerprints in distinguishing between genomic and subgenomic groups. The present study, therefore, is the first study to assess the efficacy multi-elemental fingerprints and chemometrics to discriminate between genomic and sub-genomic based on ripe banana pulp using bananas grown on the same orchard/location.

2. Materials and methods

2.1. Study site, sample collection and elemental analysis

The study site, agronomic conditions, sample collection and elemental analysis were as previously described by Devarajan et al. (2021). One hundred (100) Indian banana accessions were used (Table 1). The accessions belonged to six genomic groups namely, AAB (47), AA (2 genotypes), AAA (7), BB (2), AB (6), and ABB (36). The bananas were collected from the Research Farm of ICAR-National Research Centre for Banana (11.50∘E latitude and 74.50∘E longitude, 90 m in altitude) and utilized for this study based on their bunch availability. The accessions consisted of 18 sub-genomic groups. The names and details of the cultivars that were collected from the germplasm were recorded on the ProMusa website (Crichton et al., 2016).

2.2. Sample preparation and processing

For each banana accession, five (5 g) of the fresh fruit pulp were homogenized, weighed in a 150 ml conical flask and dried overnight in a thermostat oven at 60 °C. Thereafter, the fresh fruit pulp was digested in a tri-acid (Sulphuric acid: Nitric acid: Perchloric acid in a ratio of 7:5:3) overnight and at 70° C in a fume hood chamber until the solution became colourless. The digested samples were cooled and transferred to a 100 ml volumetric flasks with distilled water and the final volume of the solution was made to the mark. Filter paper (Whatman no. 42) was used to filter the solution and then it was stored in airtight containers till further use.

2.3. Analysis of mineral contents

The mineral content of the ripe banana fresh fruit pulp was analyzed using a Prodigy7® Inductively Coupled Plasma-Optical Emission Spectroscope (ICP-OES) (Teledyne Leeman Labs, USA). The following parameters were utilized. The RF power was 1200 W with an RF frequency of 40.68 MHz. The sample flow rate was 0.50 L/min, with a plasma gas flow rate of 11 L/m. The integration and stabilization times were 5 s and 15 s respectively. The nebulization pressure was 20 psi while the plasma view was set to axial. Argon gas of 99.99% purity was utilized while nitrogen was used as the purging gas. The analytical wavelengths (nm) for the different elements were set as follows: B (249.772), Ca (393.366), Fe (259.940), Mg (279.553), Mn (257.610), Na (589.592), P (213.618), Zn (206.200), K (766.491). The sample introduction done using a variable speed 4 channel peristaltic pump with a cyclone spray chamber that had a concentric glass nebulizer. The manufacturer's standards were analyzed to verify the accuracy of the calibration before, during and after sample analysis (Tandon, 2005). The calibration R^2 values ranged between 0.9994 and 0.9998 with limits of quantification (LOQs) in the range 0.009–0.0005 mg/L while the limits of detection (LODs) were in the range 0.0001–0.002 mg/L (see supplementary data Table S4). After correcting for the dilution factor brought on by the digestive process, the mineral content was calculated on a fresh weight (f.w.) basis and expressed as mg/kg.

2.4. Classification analysis using multivariate data analysis (Chemometrics)

The multi-variate analysis was done according to methods described by Maseko et al. (2022) and using Statistica® software version 8 (StatSoft, Tulsa, OK, USA). The multivariate data analysis was done independently for the genomic and sub-genomic groups and was subsequently referred to as genomic group-based and sub-genomic group-based analysis, respectively. Principal component analysis (PCA) was used as an exploratory analysis tool for non-supervised highlighting of underlying data structure (fingerprint) of the elemental concentrations. The PCA in combination with linear discriminant analysis (PCALDA), artificial neural network (PCA-ANN) and support vector machine (PCA-SVM) was used to identify multi-elemental mineral fingerprints and for classification analysis.

The elemental content data was separated into training/calibration and verification/test sets with an 80:20 split using MS-Excel random numbers sorting. Sub-genomic group 80/20 split led to 244 specimens in the training set (Table 2) while the verification set consisted of fifty-six (56) specimens. The subgenomic group training set consisted of eighteen (18) sub-genomic groups with each subgenomic group consisting of several specimens (Table 2). The verification set consisted of sixteen (16) subgenomic groups with each subgenomic group represented by at least one (1) specimen) with the highest having 14 specimens. Genomic group 80/20 split between the training and verification set led to a training set of 220 specimens (Table 3), while the verification set consisted of eighty (80) specimens. The genomic groups verification set consisted of all the 6 genomic groups with each group having AA (1) AAA (5), AAB (42), AB (4), ABB (26), and BB (2).

A one-way analysis of variance (ANOVA) was conducted on the subgenomic and genomic group-based training set data in order to ascertain the difference in concentration of each element between the groups with a 95% confidence interval ($p \leq 0.05$). Fisher's least significant difference (LSD) test was conducted as a post-hoc test for distinguishing between the subgenomic and genomic group. The PCA was done based on correlation, with data scaled using unit standard deviations. The NIPALS algorithm was used with cross-validation specifications set as V-fold with a value of 7 and a seed of 1708870. The maximum number of iterations was set at 50 with a convergence criterion of 0.0001. Eigen values were used to determine the number of components with >1 as the limit (Rea & Rea, 2016). Variable importance was determined based on the variable power value while loadings were assessed for determining the importance of the element in the finger-print structure. The PCA model was stored as Predictive Model Markup Language (PMML) and then the scores applied to the classification methods of linear discriminant analysis, support vector machine, and artificial neural networks analysis to yield PCA-LDA, PCA-SVM and PCA-ANN models, respectively. The PCA scores obtained from the genomic and sub-genomic group-based analysis PCA were subjected to linear discriminant analysis, support vector machine, and artificial neural network analysis to generate PCA-LDA, PCA-SVM and PCA-ANN calibration models. The models were stored as PMML consisting of the PCA model and the typical classification models (LDA, ANN and SVM) (see supplementary data, appendix 1a-h) and then applied to the test/-verification data set to assess the efficacy of the classification. The

Table 1
Sub-genomic and genomic groupings of 100 Indian banana varieties utilized for assessment of discrimination between sub-genomic and genomic groups using multi-elemental fingerprints and chemometrics.

Number	Variety name	Sub-genome group **	Genome group	Number	Variety name	Sub-genome group**	Genome group	Number	Variety name	Sub-genome group**	Genome group
1	Anai Komban	Unique	AA	35	Nanjangud Rasabale*	Silk	AAB	69	Beula	Bluggoe or Monthan	ABB
2	Namarai	Unique	AA	36	Nendran (Quintal Nendran) *	Plantain	AAB	70	Bluggoe*	Monthan	ABB
3	Kunnan*	Kunnan	AB	37	Nendran*	Plantain	AAB	71	BoddidaBukkisa	PisangAwak	ABB
4	Ney Poovan*	Ney Poovan	AB	38	Ladan	Pome	AAB	72	Chinia*	PisangAwak	ABB
5	Njali Poovan*	Ney Poovan	AB	39	Pacha5717	Pome	AAB	73	Chinia (Manohar)*	PisangAwak	ABB
6	Poomkalli	Kunnan	AB	40	Pachanadan*	Pome	AAB	74	Cuba	Bontha	ABB
7	Vadakkan Kadali	Ney Poovan	AB	41	Pacha1674	Pome	AAB	75	DakshinSagar	PisangAwak	ABB
8	Valia Kunnan	Kunnan	AB	42	Pacha Bale	Pome	AAB	76	DeshiKadali	PisangAwak	ABB
9	Borkal Baista	Borkal	BB	43	Padathi	Pome	AAB	77	Gauria (Bluggoe)*	Bluggoe	ABB
10	Musa balbisiana	Balbisiana	BB	44	PeyKadali	Pome	AAB	78	Gouria (PisangAwak)*	PisangAwak	ABB
11	Bhaskara Kali	Red	AAA	45	Poovan*	Mysore	AAB	79	Goukar	Monthan-Ash	ABB
12	Chakkarakeli*	ThellaChakkarakeli	AAA	46	Poovan (Palayankodan)*	Mysore	AAB	80	Jammulapalem Collection	PisangAwak	ABB
13	Dwarf Cavendish*	Cavendish	AAA	47	Poovazhai	Mysore	AAB	81	Kait Khullung	Bontha	ABB
14	Grand Naine*	Cavendish	AAA	48	Rajapuri	Pome	AAB	82	Kait Shjeng	Bontha	ABB
15	Grand Naine BARC Mutant*	Cavendish	AAA	49	Rajapuri India	Plantain	AAB	83	Kapur	Bluggoe	AAB
16	Manoranjitham*	Unique	AAA	50	Rajthali	Pome	AAB	84	Karibale*	Monthan	ABB
17	Singapur*	Cavendish	AAA	51	Malbhog*	Silk	AAB	85	Karpooravalli*	PisangAwak	ABB
18	Alpon*	Mysore	AAB	52	Rasthali (Andhra Rasthali)*	Silk	AAB	86	Kechulepa	Unique	ABB
19	Atrusingan	Pome	AAB	53	Rasthali (Patkapura)*	Silk	AAB	87	Kari Bontha	Bontha	ABB
20	Borchampa	Mysore	AAB	54	Rasthali*	Silk	AAB	88	Kothia*	Bluggoe	ABB
21	Chinali28	Unique	AAB	55	Rasthali (poovan)*	Silk	AAB	89	Madhok Grong	Monthan	ABB
22	Chinali483	Pome	AAB	56	Sabri (Rasthali)*	Silk	AAB	90	Manjavazha	Bontha	ABB
23	Dudh Munger	Pome	AAB	57	Sirumalai*	Pome	AAB	91	Monthan*	Monthan	ABB
24	Eathen*	Plantain	AAB	58	Soneri	Mysore	AAB	92	Nepali Vannan	PisangAwak	ABB
25	Giant*	Pome	AAB	59	Terabun	Mysore	AAB	93	Nute Pong	Bontha	ABB
26	Hybrid Co-I	Pome	AAB	60	Thenkali	Pome	AAB	94	Peyan*	Peyan	ABB
27	Kaali	Pome	AAB	61	Thiruvanandapuram	Unique	AAB	95	Peyan (pome)*	Pome	ABB
28	Hybrid Vannan x PisangLilin	Pome	AAB	62	ThiruvannanThaspulam	Unique	AAB	96	Pidi Monthan	Monthan	ABB
29	Krishnavazhai	Pome	AAB	63	Vannan	Pome	AAB	97	Poombidiyan	PisangAwak	ABB
30	Ladan Pointed	Pome	AAB	64	Aitta Kola	Monthan	ABB	98	Sawai	Monthan	ABB
31	Mala Vazhai*	Pome	AAB	65	Bangrier*	Bluggoe	ABB	99	PeyKunnan	Monthan	ABB
32	Malai Kali243	Pome	AAB	66	Barharia	Monthan	ABB	100	Yenugu Bontha*	Monthan	ABB
33	Malai Kali275	Pome	AAB	67	Batheesa Ash	Monthan	ABB				
34	Mara Bale	Pome	AAB	68	Batheesa Local	Bluggoe or Monthan	ABB				

* Commercial cultivars
** The names and details obtained and verified based on Crichton et al. (2016).

Table 2
Subgenomic group training set specimens and elemental concentrations (mg/kg fresh weight)* of sub-genomic groups consisting of 100 Indian banana varieties utilized for assessment of discrimination between sub-genomic groups using multi-elemental fingerprints and chemometrics.

Sub Genomic Group	Specimens	Elements**								
		B	Ca	Fe	Mg	Mn	K	Zn	Na	P
Bluggoe	8	4.75 ± 3.509^{bcd}	520 ± 498.3^{abc}	9.13 ± 8.556^{ab}	294 ± 95.5^{abc}	5.16 ± 3.481^{bcde}	629 ± 138.3^{abc}	1.83 ± 0.607^{a}	161 ± 22.2^{abcd}	2.34 ± 0.687^{a}
Bluggoe or Monthan	6	4.52 ± 3.534^{bcd}	506 ± 153.8^{abcd}	7.50 ± 2.279^{ab}	315 ± 60.3^{abc}	3.78 ± 0.941^{abcd}	571 ± 112.0^{abc}	1.79 ± 0.271^{a}	228 ± 61.5^{ce}	2.89 ± 1.399^{abce}
Bontha	13	2.75 ± 2.252^{abcd}	585 ± 377.0^{bcd}	7.73 ± 3.288^{ab}	322 ± 92.4^{abc}	3.65 ± 1.698^{abc}	629 ± 92.3^{abc}	3.11 ± 2.147^{b}	195 ± 57.0^{bcde}	2.66 ± 1.274^{ab}
Cavendish	11	2.52 ± 1.433^{abc}	588 ± 534.7^{bcd}	10.0 ± 4.167^{ab}	297 ± 76.3^{abc}	5.52 ± 2.438^{de}	605 ± 88.1^{abc}	2.02 ± 0.856^{a}	193 ± 34.0^{bcde}	3.73 ± 1.296^{cde}
Kunnan	6	3.53 ± 2.669^{abcd}	262 ± 98.2^{ab}	6.97 ± 1.677^{ab}	288 ± 35.5^{abc}	3.95 ± 2.253^{abcde}	629 ± 62.6^{abc}	2.14 ± 0.612^{ab}	115 ± 70.9^{a}	4.27 ± 0.870^{d}
Monthan	30	4.34 ± 4.173^{cd}	392 ± 262.1^{ab}	7.49 ± 2.27^{a}	288 ± 72.9^{ab}	3.80 ± 1.941^{abc}	632 ± 207.7^{bc}	1.87 ± 0.733^{a}	200 ± 92.7^{ce}	3.97 ± 1.565^{c}
Mysore	17	4.16 ± 3.352^{bcd}	847 ± 738.7^{d}	8.02 ± 5.382^{ab}	325 ± 113.3^{bc}	4.75 ± 1.067^{bcde}	563 ± 101.6^{ab}	2.46 ± 1.256^{ab}	200 ± 63.5^{cde}	2.66 ± 1.512^{ab}
Ney Poovan	5	5.66 ± 4.246^{d}	704 ± 922.8^{bcd}	5.34 ± 1.868^{ab}	323 ± 74.3^{abc}	3.79 ± 1.213^{abcde}	651 ± 119.8^{abc}	2.10 ± 1.261^{ab}	206 ± 79.7^{bcde}	3.31 ± 1.843^{abcde}
Peyan	3	4.14 ± 0.554^{abcd}	531 ± 72.7^{abcd}	9.36 ± 1.281^{ab}	372 ± 50.8^{c}	6.19 ± 0.847^{cde}	657 ± 88.1^{abc}	2.63 ± 0.355^{ab}	183 ± 24.8^{abcde}	4.5 ± 0.522^{de}
PisangAwak	26	2.93 ± 2.096^{abcd}	346 ± 157.5^{ab}	8.12 ± 4.639^{ab}	291 ± 42.6^{abc}	3.62 ± 1.695^{ab}	653 ± 92.3^{bc}	2.01 ± 0.730^{a}	152 ± 43.0^{ab}	2.93 ± 1.035^{abc}
Plantain	12	2.15 ± 1.416^{ab}	226 ± 148.8^{a}	12.2 ± 12.22^{b}	268 ± 32.5^{a}	3.01 ± 1.468^{a}	582 ± 186.4^{abc}	1.82 ± 0.304^{a}	150 ± 26.0^{abd}	3.44 ± 1.130^{bcde}
Pome	64	3.42 ± 3388^{bcd}	416 ± 210.9^{ab}	8.74 ± 8.761^{ab}	307 ± 71.2^{abc}	4.27 ± 1.988^{abcd}	668 ± 200.3^{c}	2.18 ± 1.238^{a}	213 ± 86.4^{e}	2.72 ± 1.080^{ab}
Silk	20	1.47 ± 0.872^{a}	739 ± 515.4^{cd}	7.59 ± 2.142^{ab}	305 ± 89.3^{abc}	4.96 ± 2.167^{cde}	563 ± 158.8^{a}	2.14 ± 0.952^{a}	209 ± 47.1^{ce}	2.82 ± 0.832^{abc}
Unique	15	1.65 ± 1.110^{a}	551 ± 418.1^{bc}	10.1 ± 8.27^{ab}	338 ± 80.0^{c}	5.82 ± 3.373^{e}	556 ± 175.6^{ab}	2.11 ± 0.395^{a}	205 ± 68.8^{ce}	3.37 ± 1.249^{abcde}

*** Numbers with the same superscript in each column, are not significantly different at 95% confidence interval.
* The training set consisted of approximately 80% of the total specimen number and was randomly generated using MS-Excel random number sorting.

Table 3
Genomic group training set specimens and elemental concentrations (mg/kg fresh weight)* of genomic groups consisting of 100 Indian banana varieties utilized for assessment of discrimination between genomic groups using multi-elemental fingerprints and chemometrics.

Genomic Group	Specimens	Element**								
		B	Ca	Fe	Mg	Mn	K	Zn	Na	P
AA	5	1.74 ± 0.392^{a}	616 ± 398.1^{ab}	19.5 ± 9.536^{c}	395 ± 19.3^{c}	6.05 ± 3.494^{c}	565 ± 87.8^{a}	2.33 ± 0.367^{a}	238 ± 66.7^{c}	3.10 ± 0.691^{ac}
AAA	16	3.13 ± 2.671^{a}	669 ± 560.0^{b}	7.48 ± 3.235^{ab}	318 ± 80.0^{a}	4.15 ± 2.646^{abc}	602 ± 92.5^{a}	2.06 ± 0.537^{a}	208 ± 56.3^{abc}	3.18 ± 0.744^{ab}
AAB	99	2.86 ± 1.19^{a}	482 ± 424.9^{ab}	8.80 ± 8.101^{b}	297 ± 79.6^{a}	4.40 ± 2.021^{bc}	621 ± 192.8^{a}	2.07 ± 1.006^{a}	201 ± 74.1^{bc}	2.82 ± 1.182^{a}
AB	14	3.76 ± 3.113^{a}	559 ± 713.1^{ab}	5.98 ± 1.817^{ab}	300 ± 64.6^{a}	3.43 ± 1.562^{ab}	614 ± 99.6^{a}	2.29 ± 1.136^{a}	164 ± 86.7^{ab}	3.21 ± 1.389^{ab}
ABB	82	3.54 ± 2.992^{a}	398 ± 224.2^{a}	6.79 ± 3.570^{a}	282 ± 58.8^{a}	3.56 ± 1.600^{a}	632 ± 145.1^{a}	1.93 ± 0.944^{a}	173 ± 64.5^{a}	3.22 ± 1.464^{b}
BB	4	2.78 ± 0.457^{a}	367 ± 60.4^{ab}	5.19 ± 0.674^{ab}	291 ± 31.7^{a}	3.12 ± 0.510^{ab}	721 ± 77.8^{a}	2.00 ± 0.400^{a}	148 ± 16.7^{abc}	1.86 ± 0.298^{a}

Reference
Devarajan et al. (2021).
* The training set consisted of approximately 80% of the total specimen number and was randomly generated using MS-Excel random number sorting.
** Numbers with the same superscript in each column, are not significantly different at 95% confidence interval.

efficacy was reported as predictive accuracy percentage (number of corrected predicted genomic or subgenomic group/total number of genomic or subgenomic groups×100).

3. Results and discussion

3.1. Genomic group and sub-genomic group elemental composition of fresh fruit pulp

The elemental content of the ripe fresh fruit pulp for both genomic and sub-genomic groups are summarized in Table 2 and Table 3. The concentrations of all the elements in the present study were within the range of previously reported values from bananas in different locations (Table 4). The nutritional implications of the levels of the elements in the ripe bananas in the present study were adequately discussed by Devarajan et al. (2021).

The average elemental content, for the 100 banana cultivars ranked in the following decreasing order: K>Mg>Ca>Na>Fe>Mn>P>Zn>B. The elemental ranking found in this study is like those reported by Alkarkhi et al. (2009) (K>Mg>Ca); Devarajan et al. (2021) (K>Ca>Mg>Na>Fe>Mn>B>P>Zn>Cu); do Prado Ferreira and Teixeira Tarley (2020) (Mg > Ca > Fe ≥ Mn > Zn > Cu) and Hardisson et al.

Table 4
Comparison of elemental concentration of the sub-genomic groups of fresh banana fresh pulp with literature Ranges.

Elements	Concentrations for Sub-genomic Groups data set (mg/kg)	Concentrations for Genomic Groups data set (mg/kg)	Literature Reported Ranges (mg/kg)	References
B	1.47 – 5.66	1.74 – 3.76	0.0 – 19.1	(Hardisson et al., 2001, Maseko et al., 2022, Wall, 2006)
Ca	226 – 847	398 – 669	28 – 3986	(Alkarkhi et al., 2009, Anyasi et al., 2018, Ayo-Omogie et al., 2021, Davey et al., 2009, do Prado Ferreira and Teixeira Tarley, 2020, Hardisson et al., 2001, Inyang and Ekop, 2015, Kookal and Thimmaiah, 2018, Maseko et al., 2022, Wall, 2006)
Fe	5.34 – 12.2	5.19 – 19.5	0.4 – 135.1	(Alkarkhi et al., 2009, Ayo-Omogie et al., 2021, Hardisson et al., 2001, Inyang and Ekop, 2015, Kookal and Thimmaiah, 2018, Maseko et al., 2022, Pillay and Fungo, 2016, Wall, 2006)
Mg	268 – 372	282 – 395	9.6 – 4800	(Alkarkhi et al., 2009, Anyasi et al., 2018, Ayo-Omogie et al., 2021, Davey et al., 2009, Hardisson et al., 2001, Inyang and Ekop, 2015, Maseko et al., 2022, Wall, 2006)
Mn	3.01 – 6.19	3.12 – 6.05	0.7 – 80.9	(Alkarkhi et al., 2009, do Prado Ferreira and Teixeira Tarley, 2020, Hardisson et al., 2001, Maseko et al., 2022, Wall, 2006)
K	556 – 668	565–632	209.9 – 18533.3	(Alkarkhi et al., 2009, Anyasi et al., 2018, Ayo-Omogie et al., 2021, Davey et al., 2009, Hardisson et al., 2001, Inyang and Ekop, 2015, Maseko et al., 2022, Wall, 2006)
Zn	1.79 – 3.11	1.93 – 2.33	0.0 – 46.3	(Alkarkhi et al., 2009, Ayo-Omogie et al., 2021, Hardisson et al., 2001, Inyang and Ekop, 2015, Maseko et al., 2022, Pillay and Fungo, 2016, Wall, 2006)
Na	115 – 228	164 – 238	16 – 1292.8	(Alkarkhi et al., 2009, Ayo-Omogie et al., 2021, Hardisson et al., 2001, Inyang and Ekop, 2015, Wall, 2006)
P	2.34 – 4.27	1.86 – 3.21	0.5 – 1530	(Anyasi et al., 2018, Ayo-Omogie et al., 2021, Davey et al., 2009, Hardisson et al., 2001, Inyang and Ekop, 2015, Maseko et al., 2022, Wall, 2006)

(2001) (K>Mg>P>Ca>Na>Fe>B>Zn>Cu>Mn) and Maseko et al. (2021) (K > N > Mg > P > Ca > Fe > Mn > B > Zn > Cu). The findings of the study were slightly different from those reported by Anyasi et al. (2018) (K>Mg>P>Ca>S) probably due to differences in agronomic conditions, varieties utilized and ripening stage of the bananas, and type of methods and treatments used to process the banana samples.

Bananas are bio-accumulators of potassium (Mohapatra et al., 2010; Ranjha et al., 2022), hence in the present study, they have the highest concentrations compared to other elements (Tables 2 and 3). This is in accordance with the findings of Alkarkhi et al. (2009); Davey et al. (2009); Devarajan et al. (2021) and Maseko et al. (2022) who demonstrated that potassium had significantly higher concentrations compared to the other elements in bananas. The K content ranged from 556 to 668 mg/kg for the sub-genomic groups and 565 – 632 mg/kg for the genomic groups ($p<0.05$). The sub-genomic group Pome had the highest K content whilst Unique sub genomic group had the lowest K content. The BB genomic group had the highest content whilst the AA group had the lowest K content.

Calcium and phosphorus are essential for the formation of strong bones and teeth, body development, cell metabolism, heart function, and for blood clotting (Dotto et al., 2019; Soetan et al., 2010). The concentrations of Ca and P reported in the present study falls within

ranges reported in the literature (Table 4). The Silk sub-genomic group had the highest Ca content whilst the Plantain sub-genomic group had the lowest Ca content.

The P content ranged from 2.34 to 4.27 mg/kg for the sub-genomic groups and 1.86–3.21 mg/kg for the genomic groups ($p<0.05$). The concentrations of the P reported in the present study was notably low compared to other studies (Anyasi et al., 2018; Hardisson et al., 2001; Maseko et al., 2022; Wall, 2006). The Monthan sub-genomic group had the highest P content whilst the Bluggoe sub-genomic group had the lowest P content. The ABB genomic group had the highest P content whilst BB genomic group had the lowest P content. The relatively low P concentrations reported in the present study could be attributed to phosphorus deficiencies in the soil where the bananas were grown, and or use of different cultivars, agronomic conditions and practices, maturity stages, diverse sampling and analytical strategies followed as summarized and proposed by Devarajan et al. (2021).

Sodium content ranged from 115 to 338 mg/kg for the sub-genomic groups and 164–238 mg/kg for the genomic groups ($p<0.05$). The Kunnan sub-genomic group had the lowest Na content whilst the Bluggoe or Monthan sub-genomic group had the highest Na content. The AAA genomic group has the highest Na content whilst the BB has the lowest Na content. The Mg content ran from 268 to 372 mg/kg for the sub-genomic groups and 282 – 395 mg/kg for the genomic groups ($p<0.05$). The AAA genomic group had highest Mg content whilst the ABB genomic group had the lowest Mg content. The Peyan sub-genomic group had the highest Mg content whilst Plaintain sub-genomic group had the lowest Mg content.

Micronutrients are critical for optimum physiological development of crops and tend to be in low concentrations in most soils and fruits, hence the need for supplementation. The concentrations of Fe, Mn, Zn and B falls within ranges reported in the literature (Table 4). Iron (Fe) ranked from 5.34 – 12.2 mg/kg for the genomic groups and 5.19–195 mg/kg for the genomic groups ($p<0.05$). The plantain sub-genomic group had the highest Fe content whilst the Ney Poovan sub-genomic had the lowest Fe content. The AAB genomic group had the highest Fe content whilst BB genomic group had the lowest Fe content. Manganese (Mn) ranked from 3.01 to 6.19 mg/kg for the sub-genomic groups and 3.12 – 6.05 mg/kg for the genomic groups ($p<0.05$). The Peyan sub-genomic group had the highest Mn, whilst the Plaintain sub-genomic group had the lowest Mn content. The AA had the highest Mn content whilst BB had the lowest Mn content. Zinc (Zn) ranked from 1.79 – 3.11 mg/kg for the sub-genomic groups and 1.93 – 2.33 mg/kg for the genomic groups ($p<0.05$). The Bontha sub—genomic group had the highest Zn content whilst Bluggoe sub-genomic group had the lowest Mn content. The AA had the highest Zn content whilst the ABB genomic group had the lowest Zn content. Boron (B) ranked from 1.47 – 5.66 mg/kg for the sub-genomic groups and 1.76 – 3.76 mg/kg for the genomic groups ($p<0.05$). There was no statistical significance in the concentrations of boron for the genomic groups ($p>0.05$). The Ney Poovan sub-genomic group had the highest B content whilst the silk had the lowest B content. The AB genomic group had the highest B content whilst the AA genomic group had the lowest B content.

The multielement concentration of bananas have been shown to be influenced by location and its association physicochemical and environmental conditions such as pH (Devarajan et al., 2021). Hardisson et al. (2001) found that 60 banana cultivars from two different regions were heavily influenced by location where they were grown. Hence, a study by Forster et al. (2002a) demonstrated that the elemental content of 95 banana varieties from Brazil(Tenerife) and Ecuador varied, affirming the fact that the location has an influence on the element content of bananas. Hence, it could be suggested that the variation in the elemental content of banana fruits in the present study was heavily influenced by the location where they were grown. Maseko et al. (2022) found that the elemental content of banana fruits grown under the same agronomic conditions varied. The authors suggested that the elemental content of banana fruits could also influenced by its intrinsic genetic differences. Hence, it could be concluded that the variation in the elemental content of the banana fruits could also be due to underlying genetic differences, given that the elemental content varied between the genomic and sub-genomic groups. Other factors that could have caused variation in the element content of banana fruits in the present study and previously reported values could be differences in maturity stages of the fruits; sample sizes; analytical methods and experimental protocols followed. These factors may lead to variation in the manner which the genomic and sub-genomic groups sequestrate and utilize the multi-elements. Hence, based on these factors and variations, it is possible to distinguish between genomic and sub-genomic groups based on the multi-elemental concentrations of bananas. The multivariate/chemometrics approach is necessary given that between some variation may occur between various studies due to differences in agronomic conditions and locations, use of different cultivars, and analytic methods/procedures.

3.2. *Elemental fingerprint multivariate pattern recognition and fingerprint-based classification*

The sub-genomic group based PCA extracted 13 components (eigen value $>$ 1.0), which explained 77.85% of variation in the data (see supplementary data 2a). There was limited clustering observed in the sub-genomic groups based PCA plot (Fig. 1a). The limited clustering and substantial number of components extracted indicated an elevated level of variation in the data in sub-genomic group data. The genomic-based PCA yielded seven (7) components which explained 75.51% of the sum of squares (see supplementary data 2b) with more distinct clustering on the PCA plot (Fig. 1b). The lower number of components compared to the sub-genomic group indicated better efficacy for identifying the underlying structure/fingerprint by the data for the genomic group compared to the sub-genomic groups. The overlapping clustering for both the subgenomic and genomics group analysis (Fig. 1a and b, respectively), indicated a need for further analysis using typical classification methods such as LDA, SVM or ANN.

A variable is considered a significant contributor to a given principal component if its loading absolute value is ≥ 0.5. For the sub-genomic PCA, the elements with loadings $>$ 0.5 for PC1 were in the order Mg $>$ Mn $>$ Ca $>$ Fe and Zn $>$ Na (Fig. 2a). There were no elements with loadings $>$ 0.5 for PC2, whilst recognizable loadings for PC3 were K $>$ P $>$ Na (Fig. 2a). The genomic-based PCA element loadings $>$0.5 were in the order Mg$>$Ca$>$Mn$>$Zn$>$Fe$>$ Na for PC1 (Fig. 2b). The notable elements with loadings ($>$0.5) for PC2 were K and P while only the P loading was $>$ 0.5 PC3 (Fig. 1b). The elements variable power for the sub-genomic group based PCA decreased in the order: Mg $>$ Na $>$ P $>$ Ca $>$ Mn $>$ K $>$ Zn $>$ Fe $>$ B (see supplementary data, appendix Table S3). The modeling power for the genomic-based PCA decreased order Mg $>$ Ca $>$ P$>$ Na $>$ Zn $>$ $>$ Fe $>$ B (see supplementary data, Appendix Table S4). Based on the PCA element modelling power results, it was clear that Mg was the most crucial element in the underlying fingerprint of the data for both the subgenomic and genomic group based PCA, while Ca, Na and P also play significant roles. The results were in agreement with those of Alkarkhi et al. (2009) to a reasonable extent. The notable variation from their study could have been due the lack of N in the present analysis, which was found to be important in their study. Hence, amendments of soils with Mg, Na and P for banana sub-genomic group and Mg, K and Ca for banana genomic groups should be taken into consideration in agronomic practices and breeding programs.

3.3. *Chemometrics: Multivariate multi-elemental fingerprint-based classification*

The PCA-LDA, PCA-ANN and PCA-SVM classification results of the training/calibration and test/verification set specimens are presented in Table 5. Successful transfer of a calibration model to a test data set is evidence of effectiveness and robustness of a given model. Although all

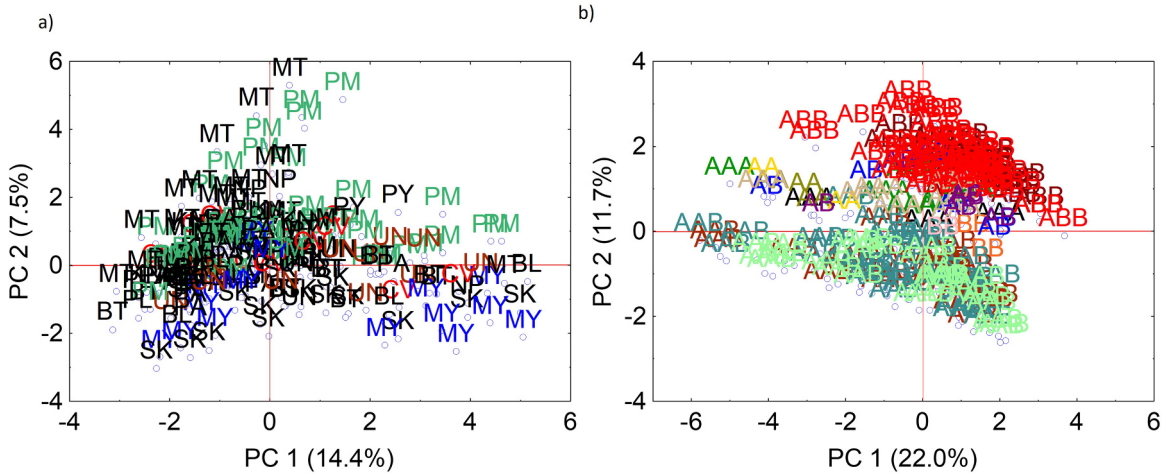

Fig. 1. Principal Component analysis score scatterplots of sub-genomic (a) and genomic groups (b) based on multi-element concentrations. BL=Bluggoe, BLM=Bluggoe or Monthan, BT=Bontha, CV=Cavendish, KN=Kunnan, MT=Monthan, MY=Mysore, NP=Ney Poovan, PA=Pisang Awak, PL=Plantain, PM=Pome, PY=Peyan, SK=Silk, UN=Unique.

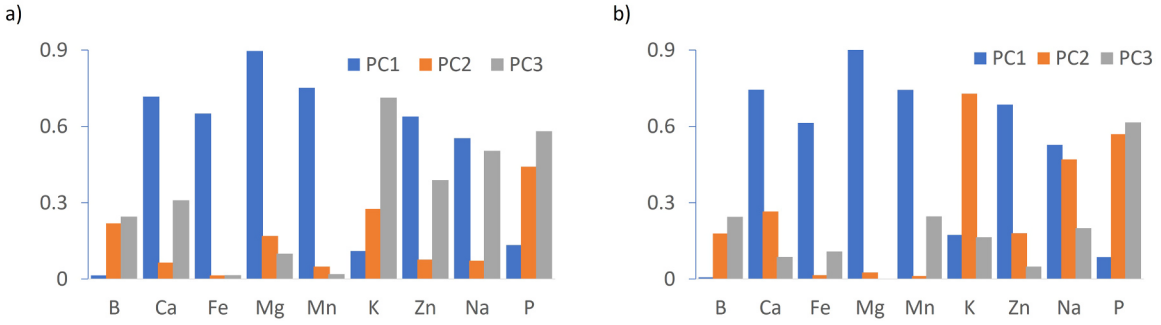

Fig. 2. Principal Components loading of the different elemental contents of for the genomic group based multi-elemental fingerprint. on multi-element concentrations of (a) sub-genome groups and (b) genome group.

Table 5
Classification accuracy of banana sub-genomic and genomic groups based on multi-elemental fingerprints of 100 Indian banana varieties for both training and verification specimen sets.

Specimen Set	Chemometrics Method	Classification Accuracy (%)	
		Subgenomic Group	Genomic Group
Training/ Calibration	PCA-LDA	100.00	0.00
	PCA-ANN	79.13	99.07
	PCA-SVM	100.00	100.00
Test/Verification	PCA-LDA	28.57	0.00
	PCA-ANN	26.53	7.50
	PCA-SVM	83.67	100.00

models gave acceptable accuracy at training/calibration (Table 5), only PCA-SVM model was able to transfer the prediction ability to the verification data set with accuracy values of 83.7% and 100.0% for the subgenomic and genomic group based PCA-SVM, respectively. The higher accuracy of prediction for the genomic group compared to the subgenomic group cold have resulted from the overall higher similarity in elemental fingerprints due to the aggregative effect at a higher classification level. The PCA-SVM model has also been shown by other researchers to be more effective than other models for discrimination for crops such as green tea and coffee beans (Yang et al., 2021). It is suggested that PCA-SVM is preferable to other classification models because it can handle both linear and non-linear datasets, is less sensitive to outliers, has the shortest processing time and high model analysis performance.

4. Conclusion

The present study successfully demonstrated that ripe banana based multi-elemental fingerprints combined with chemometrics can be used to distinguish between banana genomic and sub-genomic groups of 100 Indian Banana (Musa) accessions. The PCA-SVM model recorded the highest classification accuracy on test/verification set prediction accuracy rate compared to the PCA-ANN and PCA-LDA. In addition, the multi-elemental fingerprint classification was more effective at genomic group levels compared to sub-genomic levels. Although effective classification for Indian Musa accessions has been demonstrated in the present study, collaborative studies involving different regions of the world are needed to establish a global model.

CRediT authorship contribution statement

Victor Mlambo: Supervision, Writing – review & editing. **Obiro Cuthbert Wokadala:** Conceptualization, Formal analysis, Investigation, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **Ramajayam Devarajan:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Writing – review & editing. **Siphosethu R. Dibakoane:** Methodology, Writing – original draft. **Belinda Meiring:** Supervision, Writing – review & editing. **Jeyabaskaran Kandallu Jayaraman:** Data curation, Formal analysis,

Methodology. **Funso Raphael Kutu:** Resources, Supervision, Writing – review & editing. **July Johannes Sibanyoni:** Supervision, Writing – review & editing.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that no use of generative AI and AI assisted technologies was used in the writing process of the present manuscript.

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.

Data Availability

Data will be made available on request.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2024.106205.

References

Alkarkhi, A.F.M., Ramli, S.B., Easa, A.M., 2009. Application of multivariate statistical techniques for differentiation of ripe banana flour based on the composition of elements. Int. J. Food Sci. Nutr. *60* (sup4), 116–125. https://doi.org/10.1080/09637480802609368.

Ambuko, J.L., Sekozawa, Y., Sugaya, S., Itoh, F., Nakamura, K., & Gemma, H. (2006). Effect of seasonal variation, cultivar and production system on some postharvest characteristics of the banana. (712 ed., pp. 505-510): International Society for Horticultural Science (ISHS), Leuven, Belgium.

Anyasi, T.A., Jideani, A.I.O., McHau, G.R.A., 2018. Phenolics and essential mineral profile of organic acid pretreated unripe banana flour. Food Res. Int. *104*, 100–109. https://doi.org/10.1016/j.foodres.2017.09.063.

Arvanitoyannis, I.S., Mavromatis, A., 2009. Banana cultivars, cultivation practices, and physicochemical properties. Crit. Rev. Food Sci. Nutr. *49* (2), 113–135. https://doi.org/10.1080/10408390701764344.

Ayo-Omogie, H.N., Jolayemi, O.S., Chinma, C.E., 2021. Fermentation and blanching as adaptable strategies to improve nutritional and functional properties of unripe Cardaba banana flour. J. Agric. Food Res. *6*, 100214 https://doi.org/10.1016/j.jafr.2021.100214.

Brown, A., Tumuhimbise, R., Amah, D., Uwimana, B., Nyine, M., Mduma, H., Swennen, R., 2017. Bananas and plantains (Musa spp). Genetic Improvement of Tropical Crops. Springer, pp. 219–240.

Bugaud, C., Chillet, M., Beauté, M.P., Dubois, C., 2006. Physicochemical analysis of mountain bananas from the French West Indies. Sci. Hortic. *108* (2), 167–172. https://doi.org/10.1016/j.scienta.2006.01.024.

Cano, M.P., de Ancos, B., Matallana, M.C., Cámara, M., Reglero, G., Tabera, J., 1997. Differences among Spanish and Latin-American banana cultivars: morphological, chemical and sensory characteristics. Food Chem. *59* (3), 411–419. https://doi.org/10.1016/S0308-8146(96)00285-3.

Crichton, R., Vezina, A., & Van den Bergh, I. (2016). An online checklist of banana cultivars. (1114 ed., pp. 13-18): International Society for Horticultural Science (ISHS), Leuven, Belgium.

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 Chem. *115* (3), 806–813. https://doi.org/10.1016/j.foodchem.2008.12.088.

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 Chem. *361*, 130080 https://doi.org/10.1016/j.foodchem.2021.130080.

do Prado Ferreira, M., Teixeira Tarley, C.R., 2020. Assessment of in vitro bioacessibility of macrominerals and trace elements in green banana flour. J. Food Compos. Anal. *92*, 103586 https://doi.org/10.1016/j.jfca.2020.103586.

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. Sci. Afr. *6*, e00150.

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. Res. J. Agric. Biol. Sci. *5* (3), 272–279.

FAO. (2020). Medium-term Outlook: Prospects for global production and trade in bananas and tropical fruits 2019 to 2028. Rome.

FAO. (2021). Banana facts and figures.

Forster, M.P., Rodríguez, E.R., Martín, J.D., Romero, C.D., 2002. Statistical differentiation of bananas according to their mineral composition. J. Agric. Food Chem. *50* (21), 6130–6135. https://doi.org/10.1021/jf0255578.

Forster, M.P., Rodríguez Rodríguez, E., 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.

Gibert, O., Dufour, D., Giraldo, A., Sánchez, T., Reynes, M., Pain, J.-P., Díaz, 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. J. Agric. Food Chem. *57* (17), 7857–7869. https://doi.org/10.1021/jf901788x.

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 Chem. *73* (2), 153–161. https://doi.org/10.1016/S0308-8146(00)00252-1.

Inyang, U.E., Ekop, V.O., 2015. Physico-chemical properties and anti-nutrient contents of unripe banana and African yam bean flour blends. Int. J. Food Sci. Nutr. *4* (5), 549–554.

Izonfuo, W.-A.L., Omuaru, V.O.T., 1988. Effect of ripening on the chemical composition of plantain peels and pulps (Musa paradisiaca). J. Sci. Food Agric. *45* (4), 333–336. https://doi.org/10.1002/jsfa.2740450407.

Kookal, S.K., Thimmaiah, A., 2018. Nutritional composition of staple food bananas of three cultivars in India. Am. J. Plant Sci. *9* (12), 2480–2493.

Maseko, K.H., Regnier, T., Anyasi, T.A., Du Plessis, B., Da Silva, L.S., Kutu, F.R., Wokadala, O.C., 2022. Discrimination of Musa banana genomic and sub-genomic groups based on multi-elemental fingerprints and chemometrics. J. Food Compos. Anal. *106*, 104334 https://doi.org/10.1016/j.jfca.2021.104334.

Mohapatra, D., Mishra, S., Sutar, N., 2010. Banana and its by-product utilisation: an overview. J. Sci. Ind. Res. 69 (5), 323–329.

Nyanjage, M.O., Wainwright, H., Bishop, C.F.H., Cullum, F.J., 2001. A comparative study on the ripening and mineral content of organically and conventionally grown cavendish bananas. Biol. Agric. Hortic. *18* (3), 221–234. https://doi.org/10.1080/01448765.2001.9754885.

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

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

Ranjha, M.M.A.N., Irfan, S., Nadeem, M., Mahmood, S., 2022. A comprehensive review on nutritional value, medicinal uses, and processing of banana. Food Rev. Int. *38* (2), 199–225.

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

Soetan, K., Olaiya, C., Oyewole, O., 2010. The importance of mineral elements for humans, domestic animals and plants: a review. Afr. J. Food Sci. *4* (5), 200–222.

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

Tandon, H.L.S., 2005. Methods of Analysis of Soils, Plants, Waters, Fertilisers & Organic Manures. Fertiliser Development and Consultation Organisation.

Wall, M.M., 2006. Ascorbic acid, vitamin A, and mineral composition of banana (Musa sp.) and papaya (Carica papaya) cultivars grown in Hawaii. J. Food Compos. Anal. *19* (5), 434–445. https://doi.org/10.1016/j.jfca.2006.01.002.

Yang, S., Li, C., Mei, Y., Liu, W., Liu, R., Chen, W., Xu, K., 2021. Determination of the geographical origin of coffee beans using terahertz spectroscopy combined with machine learning methods. Front. Nutr. *8*, 680627.