Plant Gene 45 (2026) 100564

Contents lists available at ScienceDirect

Plant Gene

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

Genetic and phenotypic diversity in cannabis genotypes: insights from seed dimensions, mineral profiles, and Short Tandem Repeats (STR) markers

Sabeliwe Langa [a,b,*], Lembe Samukelo Magwaza [a,c,d], Asanda Mditshwa [a], Samson Zeray Tesfay [a]

[a] Horticultural Science, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg 3201, South Africa
[b] Department of Agronomy, Faculty of Science and Agriculture, University of Fort Hare, Alice 5700, South Africa
[c] Plant Science Laboratory, Cranfield University, Cranfield MK43 0AL, United Kingdom
[d] Department of Food Systems and Development, Faculty of Natural and Agricultural Sciences, University of Free State, PO Box 339, Bloemfontein 9300, South Africa

ARTICLE INFO

Keywords:
Cannabis
Mineral element
Seed variety
Sphericity
Short tandem repeats (STR)

ABSTRACT

This study characterized cannabis varieties based on seed physical properties, mineral composition, and Short Tandem Repeats (STR) markers. Eleven cannabis varieties were analyzed, including landraces Ladysmith Ugwayi wesiZulu ('L1') and Iswazi ('L2'), Durban Poison ('H1'), Bergville Ugwayi wesiZulu ('B1'), Natal ('B2') and Iswazi ('B3'), and Msinga Ugwayi wesiZulu ('M1') and Iswazi ('M2'), and the commercially available 'Hemp', 'Cherry bubble gum' ('High-CBD') and 'White Rhino' ('High-THC') varieties. Physical traits measured included seed length, width, thickness, geometric mean diameter, surface area, aspect ratio, and sphericity. Genetic analysis was conducted using sixty-eight STR markers, while mineral composition was assessed using Scanning Electron Microscopy (SEM) with an energy-dispersive X-ray (EDX) detector. Significant differences ($p < 0.001$) were found in seed dimensions and surface area, while aspect ratio and sphericity showed no significant variation ($p > 0.05$). Elemental composition varied significantly ($p < 0.05$), particularly for carbon, oxygen, potassium, phosphorus, sulfur, iron, and silicon. Carbon was highest in 'High-THC' (75.10 %) and lowest in 'H1' (55.58 %), whereas oxygen showed an inverse trend. Variations in seed dimensions and mineral profiles highlight the diverse genetic and phenotypic landscape of the samples. Positive correlations ($p < 0.001$) among varieties suggested similarities in physical traits, mineral composition, and STR markers. However, hierarchical clustering revealed distinct groupings, indicating complex diversity. Landraces could not be reliably classified as 'High-CBD', 'High-THC', or 'Hemp', reflecting STR marker neutrality. Integrating STRs with functional gene markers or metabolic profiling may improve chemotype discrimination and support development of cannabis-specific identification tools. These findings provide insights for breeding programs and optimization of desirable traits.

1. Introduction

Cannabis displays a vast range of physical and chemical diversity due to extensive selective breeding by humans, leading to the development of over 700 varieties worldwide (Vergara et al., 2016; Woodbridge, 2018; Monti et al., 2022). Historically, research on cannabis faced substantial obstacles due to legal restrictions associated with the psychoactive compound delta-9-tetrahydrocannabinol (Δ^9-THC), limiting both scientific inquiry and commercial utilization (Adhikary et al., 2021). These regulatory barriers have also contributed to ongoing debates regarding the plant's taxonomic classification. However, shifts in public perception and legislative frameworks across multiple countries have propelled cannabis into the spotlight for its medicinal and agricultural value, emphasizing the necessity of variety characterization to enhance clinical applications and optimize cultivation practices (Sawler et al., 2015).

Cannabis varieties are generally classified based on both chemical and morphological attributes. The chemical classification system, known as chemotaxonomy, categorizes plants into five primary chemotypes (I to V) depending on the relative concentrations of cannabinoids such as THC, cannabidiol (CBD), and cannabigerol (CBG) (Broseus et al., 2010; Cerrato et al., 2021). For example, Chemotype I is rich in THC, whereas Chemotype III is predominantly CBD-based. Chemotype II features a balanced ratio of THC and CBD, Chemotype IV is

* Corresponding author at: Horticultural Science, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg 3201, South Africa.
E-mail address: sabeliwelanga@gmail.com (S. Langa).

https://doi.org/10.1016/j.plgene.2025.100564
Received 9 February 2025; Received in revised form 28 October 2025; Accepted 3 December 2025
Available online 4 December 2025
2352-4073/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

characterized by high CBG levels, and Chemotype V lacks significant cannabinoid presence (Cerrato et al., 2021; Schluttenhofer and Yuan, 2017). Non-chemical classification methods rely on observable traits, traditionally categorizing cannabis into drug, fiber, and oilseed varieties, although these traits are often influenced by environmental adaptation and domestication (de Meijer, 1994; Salentijn et al., 2015).

Cannabinoid concentrations in cannabis plants fluctuate due to genetic factors, cultivation techniques, and plant physiology, making precise variety identification essential for maintaining consistency in production and quality (Tipparat et al., 2014; Danziger and Bernstein, 2021). The prevalence of hybrid varieties combining features of both *C. sativa* and *C. indica* further complicates classification based solely on morphological characteristics (Jin et al., 2015; Lapierre et al., 2023). Seed-based variety identification provides a more reliable alternative, as different varieties exhibit distinct external morphology and internal chemical profiles (Feng et al., 2018; Rashid et al., 2018; Zhang et al., 2020). For instance, hemp seeds are generally larger than those of marijuana, reflecting differences in resource allocation between fiber and drug-type cultivars (Naraine et al., 2020).

Examining the physical properties of cannabis seeds is crucial for optimizing agricultural practices, including sowing, processing, and storage (Gharibzahedi et al., 2009; Sastry et al., 2019). Advanced analytical techniques such as high-performance liquid chromatography (HPLC), near-infrared hyperspectral imaging (NIR-HSI), and molecular marker technologies like DNA fingerprinting offer precise, environmentally independent methods for distinguishing between cannabis cultivars (Dalusong et al., 2019; Jin et al., 2022). For instance, microsatellites or short tandem repeats (STR) have proven to be highly effective in profiling cannabis genotypes due to their ability to differentiate genetic variations with high specificity (Dalusong et al., 2019). Additionally, sequencing genes such as THCAS and CBDAS enables early differentiation between fiber and drug-type cannabis (Cascini et al., 2019).

Hemp seeds, in particular, are widely recognized for their exceptional nutritional composition, containing abundant oils, proteins, and essential micronutrients, making them valuable for both food production and biofortification initiatives (Mihoc et al., 2012; Kaliniewicz et al., 2021). However, their mineral composition varies depending on genetic and environmental influences, necessitating the identification of superior genotypes for nutritional enhancement (Rehman et al., 2019; Farinon et al., 2020). Despite increasing global interest in cannabis research, studies focusing on genotype characterization within the South African context remain limited.

As regulatory frameworks evolve to accommodate hemp and medicinal cannabis cultivation, variety identification plays a vital role in ensuring consistent yields, quality assurance, and the development of optimized agricultural strategies (Gouws, 2017; Allaf et al., 2023). This study aims to analyze cannabis varieties by assessing seed morphology, STR markers, and mineral composition, providing valuable insights for improving crop performance and expanding industrial applications of landraces.

2. Materials and methods

2.1. Seed collection and experimental design

Eleven cannabis varieties were selected for characterization, including eight landraces collected from local growers in four regions of KwaZulu-Natal, South Africa: Bergville, Hammersdale, Ladysmith, and Msinga. The remaining three commercially available varieties, namely, 'Hemp', 'Cherry bubble gum' ('High-CBD'), and 'White Rhino' ('High-THC'), were sourced from Johannesburg-based supplier, Durbanseedco under Grohydro (Pty) Ltd. and Nirvana Shop South Africa, respectively. Although analyzing seeds grown under the same conditions was preferred, planting was not feasible due to funding limitations. Consequently, the study examined the collected seeds in their existing state.

The seed samples were kept at room temperature in the Horticultural Sciences Laboratory at the University of KwaZulu-Natal in Pietermaritzburg (latitude: −29.636311 and longitude: 30.409060). Seed samples were stored in paper envelopes at room temperature for approximately one week prior to analysis. The study design involved the characterization and comparison of eleven distinct cannabis varieties. Within each variety, seeds were randomly selected for all physical, mineral, and genetic analyses to ensure the results were representative. The landraces assigned pseudonyms to ensure consistent naming throughout the study. Table 1 lists these pseudonyms along with the geographic origin of each landrace. Commercial varieties were excluded from the table because they were store-bought, and the shop location would not accurately reflect the varieties' growing conditions.

2.2. Physical parameters of seeds

The evaluation of the physical traits of the seeds of different cannabis varieties was carried out at the Seed Technology Laboratory, University of KwaZulu-Natal, South Africa. The study followed the procedures by Perea-Flores et al. (2011) and Gojiya et al. (2020). For each variety, the physical trait analysis was conducted with three replicates. For each replicate, twenty seeds were randomly selected and their length (L), width (W), and thickness (T) were measured in millimeters (mm) using a 150 mm (6″) digital calliper (China). The average values of L, W, and T were recorded and used to calculate additional physical properties, including geometric mean diameter, surface area, aspect ratio, and sphericity, using the equations provided by Perea-Flores et al. (2011) and Gojiya et al. (2020).

The geometric mean diameter (D_g) was calculated by using Eq. 1 as follows:

$$D_g = \sqrt[3]{LWT}\ (\text{mm}) \tag{1}$$

where L is the length, W is the width, and T is the thickness, all in mm.

The results obtained from the geometric mean diameter were used to calculate the surface area (S) by using Eq. 2 as follows:

$$S = \pi {D_g}^2\ (\text{mm}^2) \tag{2}$$

The aspect ratio (Ra) was calculated as the ratio of the width to its respective length (expressed as a percentage) using Eq. 3 as follows:

$$Ra = \frac{W}{L} X\ 100 \qquad (\%) \tag{3}$$

where W is the width and L is the length (mm).

The sphericity (ϕ) was calculated as the ratio of the geometric mean diameter to the length of the seed by using Eq. 4 as follows:

$$\phi = \frac{D_g}{L} \tag{4}$$

where D_g = geometric mean diameter and L is the length (mm).

2.3. Quantification of mineral elements using scanning electron microscopy

The mineral element concentration and composition of embryo tissues from the different cannabis varieties were analyzed at the Microscopy and Microanalysis Unit, University of KwaZulu-Natal. The elements analyzed included aluminum (Al), carbon (C), calcium (Ca), chlorine (Cl), copper (Cu), iron (Fe), potassium (K), magnesium (Mg), manganese (Mn), nitrogen (N), sodium (Na), oxygen (O), phosphorus (P), sulfur (S), and silicon (Si). The mineral element profiling was carried out following the procedure by Perea-Flores et al. (2011), with some modifications. Briefly, five seeds of each variety were cryo-fractured in liquid nitrogen and divided into longitudinal and transversal sections. The halves of each seed were mounted on stubs and secured using

Table 1
The pseudonyms of the cannabis landraces and their geographic location information.

Location	Geographic information					Common local name	Name used for the experiment
	Latitude:	Longitude:	Elevation	Average rainfall	Average temperature		
Ladysmith	−28.57533	29.85948	1009.53 m	108.7 mm	23.3 °C	'Ugwayi wesiZulu'	'L1'
						'Iswazi'	'L2'
Msinga	−28.700960	30.303320	559 m	340.0 mm	23.0 °C	'Ugwayi wesiZulu'	'M1'
						'Iswazi'	'M2'
Bergville	−28.677544	29.125553	1137.86 m	109.4 mm	23.5 °C	'Ugwayi wesiZulu'	'B1'
						'Iswazi'	'B3'
	−28.667679	29.034111				'Natal'	'B2'
Hammersdale	−29.870903	30.633519	597 m	110.3 mm	23.7 °C	'Durban Poison'	'H1'

carbon double-sided insulating tape. The seeds were gold-coated using a Quorum Q150R ES Gold Sputter Coater. The microanalysis of the minerals on the embryo tissue of the seeds was then conducted using a Zeiss Ultra Scanning Electron Microscope (SEM) coupled with an energy-dispersive X-ray (EDX) detector (Zeiss, Oberkochen, Germany). The EDX detector was set at an active area ranging from 8 to 10 mm^2. It was also used to perform the X-ray mapping technique, which constituted using pseudo-colours to represent the bi-dimensional spatial distribution of the energy emission of the chemical elements of the sample. Five fields of view, 100.0 μm^2 each, were analyzed for 60 s at a 20 kV accelerating voltage and at a count rate of 1000–3000 cps. The results obtained from the EDX analysis were processed using Oxford INCA & Aztec EDX Analysis Software. The elemental distribution from the X-ray mapping technique is presented in Fig. 1, with each element represented by a specific colour. This mapping provides visual insight into the spatial distribution of elements such as C, Fe, K, O, P, S, Si, Al, Ca, Cl and Mg, which is essential for understanding the mineral element composition in cannabis seeds. Mineral elements such as Cu, Mn, N and Na were detected in only one or two varieties, with each element found in one field of view per genotype out of a total of 11 genotypes and 5 fields of view. Due to their irregular occurrence and insufficient data, these elements were excluded from further analysis. The proportion of each element was then quantified in percentage of the total mineral concentration in each sample (wt%).

2.4. STR markers analysis

2.4.1. DNA extraction, polymerase chain reaction (PCR) primers and multiplex conditions

DNA was extracted from 150 mg of seeds that were carefully cut into small fragments using sterile blades to expose internal tissues and minimize cross-contamination. Extraction was performed using the Quick-DNA™ Plant/Seed Miniprep Kit (Zymo Research Corp., Irvine, CA, USA) according to the manufacturer's protocol. As recommended by the manufacturer, the samples were soaked overnight in water and then subjected to bead-beating prior to extraction to ensure thorough homogenization. DNA was quantified and assessed for purity using a NanoDrop spectrophotometry (NanoDrop Technologies, Wilmington, Delaware, USA). Spectrophotometric analysis yielded concentrations of 25–60 ng/μL. The purified DNA was immediately analyzed. Based on information from the literature (Gilmore et al., 2003; Di Nunzio et al., 2021; Xia et al., 2022), a panel of sixty-eight STR markers, representative of cannabis, was assembled for genetic analysis of varieties. These markers were selected and combined to create a custom multiplex assay. PCR amplification was conducted in an Eppendorf® Mastercycler® Nexus gradient thermocycler (Eppendorf® 6,331,000,025, Eppendorf AG, Hamburg, Germany). For the PCR protocol, the Q5 High-Fidelity 2× Master Mix (NEB) was used together with 10 μM of each primer pair in a primer mix and nuclease-free water (AMRESCO LLC, OH 44139, USA). Some reactions included modifications for multiplex fluorescent labeling. Forward primers were fluorescently labeled by incorporating universal primers conjugated to distinct fluorophores (M13-FAM, PGEX3-ATTO532, PGEX5-ATTO550, and T7-ATTO565), following the cost-effective tailed-primer method described by Schuelke (2000). The multiplex was carried out in two rounds. The first round amplification was a large, multi-panel multiplex PCR containing all 68 primer pairs. DNA was normalized to a working concentration range of 25–60 ng/μL. A volume of 1–2 μL of this working solution was used per reaction, yielding a final DNA template mass of 25–120 ng per PCR. Owing to the complexity of the 68-plex reaction, a subset of markers exhibited suboptimal amplification. To overcome this, a secondary PCR was conducted in which the underperforming markers were re-amplified using individual primer pairs or small primer sets. Diluted first-round product served as the template. This targeted approach minimized primer competition and markedly improved the amplification efficiency of weak amplicons. Despite the second round re-amplification attempt, three markers (CS1, D02-CANN1, and ANUCS204) continued to perform poorly, producing inconsistent and uninterpretable data that could not be accurately scored. Therefore, these markers were excluded from the final analysis.

For fragment analysis, fragment preparation was carried out using two microliters of the fluorescently labeled PCR amplicons and 15 μL LIZ500 sizing standard and Hi-Di™ Formamide (Thermo Fisher Scientific, Carlsbad, USA). The fragments were run on the ABI PRISM™ 3730xl Genetic Analyser (Applied Biosystems, Thermo Fisher Scientific, Carlsbad, USA), 50 cm capillary array, POP-7TM. The data was analyzed and interpreted using GeneMarker V2.9.5 (SoftGenetics, State College, Pennsylvania, USA).

2.5. Data analysis

Data analysis was carried out using GenStat® 20.1 (VSN International, Hamel Hampstead, UK, 2020) at a 5 % significance level. Duncan's multiple-range test was used to separate the means of significantly different variables. Correlation analysis, Principal Component Analysis (PCA), and Agglomerative Hierarchical Clustering (AHC) were carried out using R software version 4.4.0 within the RStudio platform (Posit Software, PBC, 250 Northern Ave Suite 420, Boston, MA 02210).

3. Results

3.1. Physical parameters

There were statistically significant differences ($p < 0.001$) among the cannabis varieties concerning seed width, thickness, length, geometric mean diameter, and surface area (Table 2). In contrast, no statistically significant differences ($p > 0.05$) were observed in seed sphericity and aspect ratio among the cannabis varieties. The seeds of 'B1' and 'M2' had relatively high mean values of width (3.74 and 3.72 mm), thickness (3.01 and 3.07 mm), length (4.80 and 4.92 mm), geometric mean diameter (3.78 and 3.82 mm) and surface area (44.92 and 46.01 mm^2) compared to the other varieties. On the other hand, the seeds of 'M1' and 'B2' had relatively low mean values of seed width (3.00 and 3.05 mm), thickness (2.48 and 2.46 mm), geometric mean diameter (3.15 and 3.09 mm) and surface area (31.24 and 29.92 mm^2) compared to the other varieties. Furthermore, 'B2' seeds had the lowest mean length of 3.91

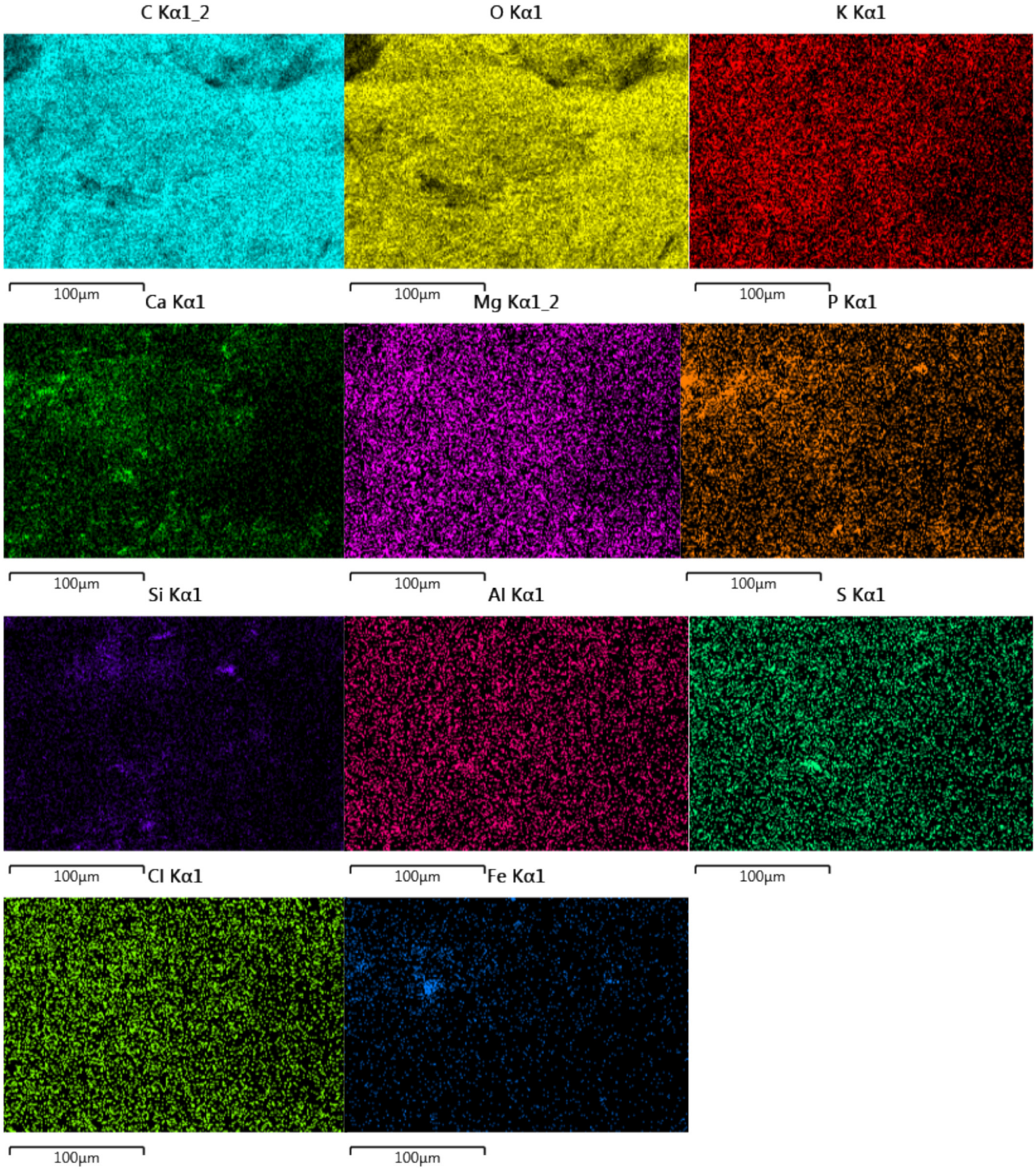

Fig. 1. Scanning electron microscopy micrographs of the embryo tissue cell elemental distribution by X-ray mapping technique from EDX analysis spectrum. **C** (carbon), **O** (oxygen), **Cl** (chlorine), S (sulfur), K (potassium), P (phosphorus), **Ca** (calcium), **Al** (aluminum), **Si** (silicon), **Fe** (iron), **Mg** (magnesium).

mm among the varieties.

3.2. Mineral element composition

The EDX spectrum revealed that embryo tissue cells contained various mineral elements, including C, Fe, K, O, P, S, Si, Al, Ca, Cl, Mn, Cu, N, Na, and Mg (Fig. 2). Among the varieties analyzed, 'M1' was identified as the only variety containing Na, however, in small amounts. 'B3' was also the only variety containing N. 'High-THC' and 'High-CBD' were the only varieties found to have Mn. Furthermore, 'High-CBD' exhibited trace amounts of Cu. To maintain analytical consistency, elements that were irregularly detected across samples were excluded

Table 2
Physical properties of eleven cannabis varieties. The means with different letters are significantly different according to Duncan's multiple range test ($p < 0.05$; number of replicates = 3). LSD = least significance difference.

Variety	Width (mm)	Thickness (mm)	Length (mm)	Geometric mean diameter (mm)	Surface area (mm^2)	Sphericity	Aspect ratio (%)
B1	3.74a	3.01ab	4.80a	3.78a	44.92a	0.79a	77.90ab
M2	3.72a	3.07a	4.92a	3.82a	46.01a	0.78a	75.70ab
L2	3.64ab	2.85abc	4.53ab	3.61ab	40.91ab	0.80a	80.39ab
Hemp	3.59ab	2.71abc	4.79a	3.60ab	40.76ab	0.75a	75.01ab
H1	3.49abc	2.89abc	4.63ab	3.60ab	40.79ab	0.78a	75.62ab
High-THC	3.49abc	2.88abc	4.12bc	3.46bc	37.61bcd	0.84a	84.83a
High-CBD	3.47abc	2.74abc	4.59ab	3.52ab	39.00abc	0.77a	75.60ab
B3	3.27bcd	2.53c	4.42abc	3.32bcd	34.60bcde	0.75a	74.15ab
L1	3.06 cd	2.59bc	4.13bc	3.20 cd	32.10cde	0.78a	74.40ab
B2	3.05 cd	2.46c	3.91c	3.09d	29.92e	0.79a	78.10ab
M1	3.00d	2.48c	4.22bc	3.15 cd	31.24de	0.75a	70.96ab
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.051	0.060
LSD	0.24	0.25	0.29	0.17	3.77	0.05	7.22

from comparative statistical analyses. Additionally, A detailed elemental distribution obtained from the X-ray mapping technique is presented in Table S1.

Moreover, major elements were found in abundance compared to the others, with carbon being dominant ranging between 55.58 and 75.10 wt%, followed by oxygen ranging between 23.79 and 42.82 wt%. The dominance of carbon and oxygen peaks is characteristic of organic biological samples. For comparative purposes, emphasis was therefore placed on the mineral elements, which provide more quantitatively reliable information with this technique.

Notably, there were statistically significant differences ($p < 0.05$) among the cannabis varieties in relation to seed C, Fe, K, O, P, S, and Si. In contrast, no statistically significant differences ($p > 0.05$) were observed among the varieties in relation to Al, Ca, Cl, and Mg (Table S1). The results revealed that K was relatively abundant among the mineral elements in the 'High-CBD', 'H1', 'Hemp', and 'B3' varieties, with values of 1.310, 0.536, 0.182, and 0.694 wt%, respectively. Calcium was the most abundant mineral in 'M1' and 'L2' (1.160 and 0.478 wt%), while Fe predominated in 'B1' and 'M2'. Aluminum was relatively higher in the 'High-THC' variety, and Si was dominant in 'B2' and 'L1'. Among the majority of landraces ('High-CBD', 'H1', 'Hemp', 'B1', 'B2', and 'L1'), calcium was the second most abundant mineral element. Potassium was the second most abundant element in 'M2', 'M1', and 'L2', whereas silicon was the second most abundant in 'B3' and 'High-THC'.

Notably, certain varieties exhibited very low levels of specific elements. For instance, 'B2', 'M2', and 'L1' had aluminum, known as an ultra-trace and toxic element, in the lowest quantity compared to other elements at 0.014, 0.024, and 0.029 wt%, respectively. 'High-CBD', 'B3', 'B1', and 'High-THC' had macro element chlorine in the lowest quantity compared to the elements at 0.064, 0.048, 0.054, and 0.010 wt%, respectively. However, chlorine was generally the lowest among the macro elements in all the varieties 0.064, 0.104, 0.036, 0.048, 0.054, 0.028, 0.036, 0.112, 0.036, 0.068 and 0.01 wt% for 'High-CBD', 'H1', 'Hemp', 'B3', 'B1', 'B2', 'M2', 'M1', 'L2', 'L1' and 'High-THC', respectively. Notably, potassium was the highest among the macro elements in 'High-CBD', 'H1', 'Hemp', 'B3', and 'M2' at 1.31, 0.536, 0.182, 0.694, and 0.392 wt%, respectively. Calcium was the highest among the macro elements in 'B1', 'M1', and 'L2' at 1.04, 1.16 and 0.478 wt%, respectively. Silicon was the highest among the macro elements in 'B2', 'L1', and 'High-THC' at 0.3255, 2.7727, and 0.5677 wt%, respectively. 'H1' and 'Hemp' had iron in the smallest quantities compared to the other elements at 0.057 and 0.030 wt%, respectively. 'Hemp' and 'L2' had sulfur in the smallest quantities compared to the other elements at 0.032 and 0.030 wt%, respectively.

3.3. Short Tandem Repeat markers

Thirty-four markers were identified in the seeds of the different cannabis varieties, distributed across four multiplex sets (Plex 1–4) as presented in Table 3. Allele size ranges varied from 70 bp for marker 3735 to 460 bp for marker B01-CANN1. Markers such as ANUCS205 (172–242 bp) and ANUCS203 (169–267 bp) displayed wide allele ranges, while others, including ANUCS307 (105–108 bp) and ANUCS306 (92–95 bp), showed narrower ranges.

3.4. Multivariate analysis

3.4.1. Multivariate analysis for physical characteristics

3.4.1.1. Correlation matrix. A Pearson's correlation was conducted on the physical parameters and among the varieties (Fig. 3A and B). Under physical parameters, there was a very strong positive correlation between seed sphericity and aspect ratio ($r = 0.93$, $p < 0.001$) (Fig. 3A). Thickness had a positive correlation with length ($r = 0.71$, $p < 0.05$) while width had a strong positive correlation with length ($r = 0.83$, $p < 0.01$) and thickness ($r = 0.91$, $p < 0.001$). Geometric mean diameter had a strong positive correlation with the surface area ($r = 1.00$, $p < 0.001$), width ($r = 0.97$, $p < 0.001$), thickness ($r = 0.93$, $p < 0.001$) and length ($r = 0.90$, $p < 0.001$). On the other hand, there was a very strong positive correlation ($p < 0.001$) between the seeds of different varieties (Fig. 3B), indicating that varieties had similarities in their physical characteristics.

3.4.1.2. Principal component analysis for physical characteristics. The principal component analysis (PCA) was conducted on physical features (Table 4). PCA identified two PCs with an eigenvalue of ≥1, which accounted for 80.11 % of the total variation. PC1 (55.51 %) positively correlated with seed width, thickness, surface area, and geometric mean diameter. PC2 (24.60 %) positively correlated to seed sphericity and aspect ratio.

3.4.1.3. Hierarchal cluster analysis. Agglomerative hierarchical clustering for means of physical traits classified the 11 landraces into four groups (Fig. 4). Cluster 1 comprised of 'B2', 'B3', 'M1' and 'L1'. Cluster 2 comprised of 'B1', 'M2', 'Hemp', 'High-CBD', and 'H1'. Cluster 3 and 4 were comprised of 'L2' and 'High-THC', respectively.

3.4.2. Multivariate analysis for mineral elements composition

3.4.2.1. Correlation matrix. Generally, there was no significant correlation among the mineral elements. Except for S, which had a strong positive correlation with K ($r = 0.90$, $p < 0.001$) and P ($r = 0.83$, $p < 0.01$) (Fig. 5A). There was also a positive correlation between P and K ($r = 0.71$, $p < 0.05$). On the other hand, C had a significant negative correlation with O ($r = -0.98$, $p < 0.001$) and Cl ($r = -0.61$, $p < 0.05$). Notably, a very strong positive correlation ($p < 0.001$) was evident among the seeds of various varieties (Fig. 5B), suggesting a similarity in

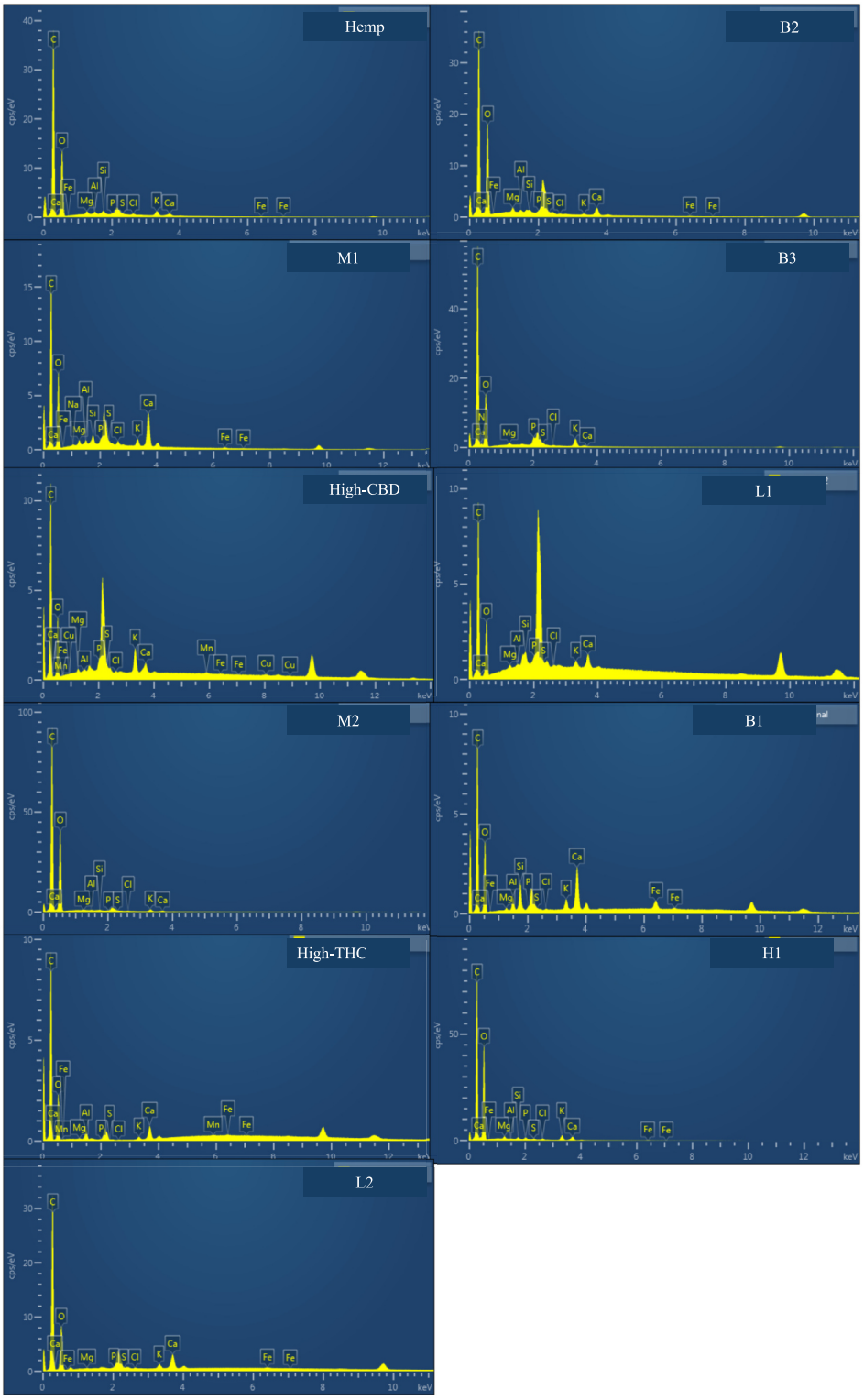

Fig. 2. Scanning electron microscopy micrographs of the EDX spectrum of embryo tissue cells on the mineral elements of different cannabis varieties.

their mineral element composition.

3.4.2.2. Principal component analysis. PCA identified three PCs with an eigenvalue of ≥1, accounting for 73.92 % of the total variation (Table 5). PC1 (35.99 %) negatively correlated with carbon (C) and positively correlated with Cl, K, S, and Mg. PC2 (24.71 %) positively correlated with Al and P and negatively correlated with O. PC3 (13.22 %) positively correlated with Ca.

3.4.2.3. Hierarchal cluster analysis. Agglomerative hierarchical clustering for means of mineral element composition classified the 11 landraces into four groups (Fig. 6). Cluster 1 comprised of 'L2', 'High-THC', 'B2', 'Hemp', and 'M2'. Cluster 2 comprised of 'H1' and cluster 3 comprised of 'B3' and 'High-CBD'. Cluster 4 comprised of 'B1', 'M1' and 'L1'.

Table 3
Detailed list of markers in cannabis varieties and their ranges.

MARKER	RANGE	MARKER	RANGE
Plex 1		**Plex 3**	
B02-CANN2	163–172	D02-CANN1	93–113
H09-CANN2	204–224	ANUCS205	172–242
5159	312–350	nH09	385–430
9269	114–140	ANUCS302	125–173
ANUCS308	177–203	9043	320–355
H11-CANN1	285–297	ANUCS204	128–184
ANUCS303	141–156	ANUCS203	169–267
B05-CANN1	217–244	**Plex 4**	
B01-CANN1	314–460	1528	175–200
ANUCS202	147–185	H06-CANN2	255–273
4910	240–300	DM016_Y	230
Plex 2		ANUCS307	105–108
3735	70–115	ANUCS201	161–223
ANUCS206	159–167	DM029	219
ANUCS301	209–261	E07-CANN1	85–113
C11-CANN1	138–180		
ANUCS305_1	410–450		
ANUCS501_X	72–105		
ANUCS304	167–230		
ANUCS306	92–95		
CS1	134–370		

3.4.3. *Multivariate analysis for short tandem repeats (STR) markers*

For clarity throughout this manuscript, the two alleles for each diploid locus are denoted as the locus name followed by an underscore and the allele number (e.g., 5159_1 and 5159_2). For markers where the locus name itself contains an underscore (e.g., ANUCS305_1), a dot is used to separate the locus name from the allele number (e.g., ANUCS305_1.1 and ANUCS305_1.2) to ensure unambiguous identification.

In addition to CS1, D02-CANN1, and ANUCS204 loci which were eliminated due to displayed inaccuracy of data, B02-CANN2_1, B02-CANN2_2, ANUCS308_1, ANUCS308_2, H11-CANN1_1, H11-CANN1_2, ANUCS202_1, ANUCS202_2, ANUCS206_1, ANUCS501_X.1, ANUCS501_X.2, ANUCS306_1, ANUCS306_2, H06-CANN2_1, DM016_Y.1, DM016_Y.2, ANUCS307_1, ANUCS201_1, ANUCS201_2, DM029_1, DM029_2 and E07-CANN1_1 were excluded from multivariate analysis because they had zero variation. The remaining STR markers were then subjected to further analysis as presented in the subsections 3.4.3.1, 3.4.3.2 and 3.4.3.3 of 3.4.3.

3.4.3.1. Correlation matrix. The Pearson correlation analysis showed variations among the STR markers, revealing significant correlations among the alleles, with some showing strong positive or negative relationships (Fig. 7). Notably, alleles 5159_1 and 5159_2 had a perfect positive correlation ($r = 1.00$; $p < 0.001$). Similarly, 9269_2 had a perfect positive correlation with ANUCS305_1.2 and ANUCS302_1 (r = 1.00; $p < 0.001$). ANUCS303_2 also showed a perfect positive correlation with ANUCS304_1 (r = 1.00; $p < 0.001$). B05-CANN1_1 and B05-CANN1_2 exhibited a perfect positive correlation (r = 1.00; $p < 0.001$). Additionally, allele 4910_1 had a perfect negative correlation with ANUCS304_2, NH09_1, and NH09_2 ($r = -1.00$; $p < 0.001$). ANUCS305_1.1 showed a perfect negative correlation with H06-CANN2_2 (r = −1.00; $p < 0.001$), while ANUCS305_1.2 had a perfect positive correlation with ANUCS302_1 (r = 1.00; $p < 0.001$). ANUCS304_2 displayed a perfect positive correlation with NH09_1 and

Table 4
Summary of factor loadings, eigenvalue, percent, and cumulative variation for eleven cannabis varieties based on their physical traits.

Traits	Dim 1	Dim 2
Width	0.974	−0.068
Thickness	0.969	0.048
Length	−0.497	0.482
Geometric mean diameter	0.236	−0.380
Surface Area	0.963	−0.213
Sphericity	0.410	0.884
Eigenvalue	444.073	196.815
Variability (%)	55.509	24.602
Cumulative (%)	55.509	80.111

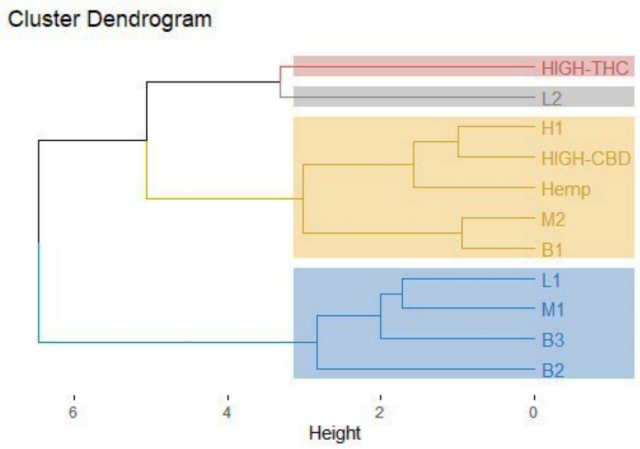

Fig. 4. Agglomerative hierarchical cluster of eleven cannabis varieties based on their physical traits.

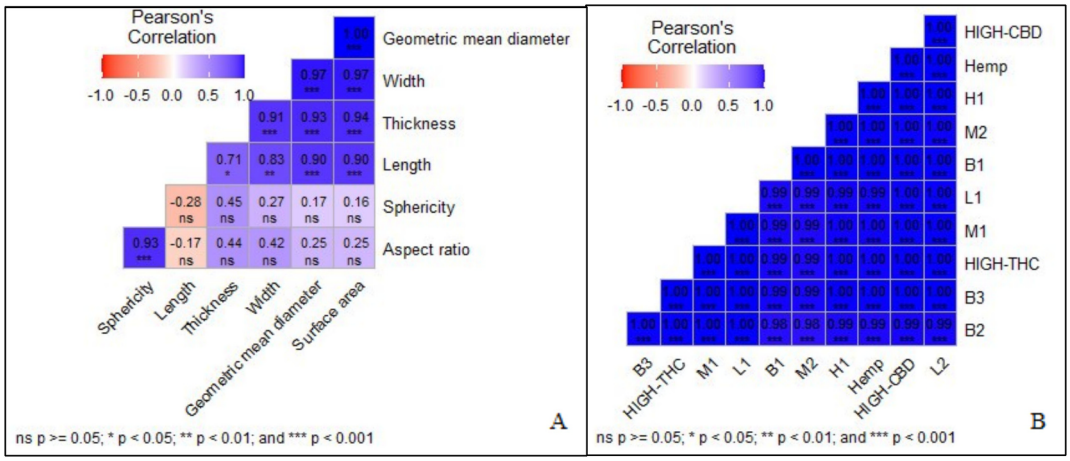

Fig. 3. The Pearson's correlation between (A) physical properties and (B) among varieties based on their physical traits.

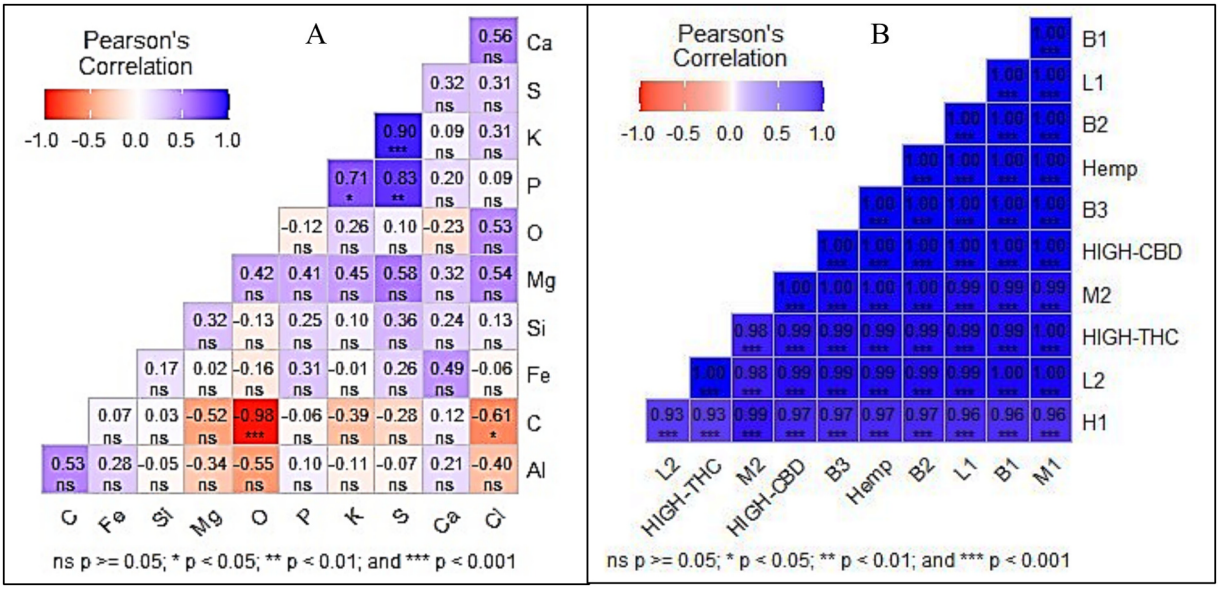

Fig. 5. The Pearson's correlation between (A) mineral element composition and (B) among varieties based on their mineral element composition. **C** (carbon), **Fe** (iron), **K** (potassium), **O** (oxygen), **P** (phosphorus), **S** (sulfur) and **Si** (Silicon), **Al** (aluminum), **Ca** (calcium), **Cl** (chlorine), and **Mg** (magnesium).

Table 5
Summary of factor loadings, eigenvalue, percent, and cumulative variation for eleven cannabis varieties based on their mineral elements. **C** (carbon), **Fe** (iron), **K** (potassium), **O** (oxygen), **P** (phosphorus), **S** (sulfur) and **Si** (Silicon), **Al** (aluminum), **Ca** (calcium), **Cl** (chlorine), and **Mg** (magnesium).

Mineral element	Dim 1	Dim 2	Dim 3
Al	−0.434	0.607	−0.014
C	−0.732	0.582	−0.021
Ca	0.306	0.539	0.720
Cl	0.710	−0.216	0.528
Fe	0.093	0.571	0.345
K	0.759	0.243	−0.488
Mg	0.809	0.013	0.134
O	0.592	−0.723	0.022
P	0.570	0.617	−0.401
S	0.797	0.496	−0.287
Si	0.294	0.375	0.193
Eigenvalue	**3.958**	**2.718**	**1.455**
Variability (%)	**35.986**	**24.711**	**13.225**
Cumulative (%)	**35.986**	**60.697**	**73.922**

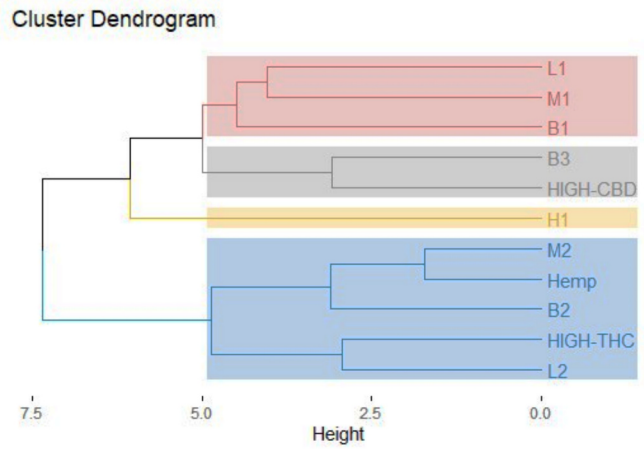

Fig. 6. Agglomerative hierarchical cluster for eleven cannabis varieties based on their mineral element composition.

NH09_2 (r = 1.00; $p < 0.001$), and NH09_1 and NH09_2 also had a perfect positive correlation with each other (r = 1.00; $p < 0.001$).

A very strong positive correlation ($p < 0.001$) was evident among the seeds of various varieties (Fig. 8), suggesting a similarity in their STR markers.

3.4.3.2. Principal component analysis. PCA identified six components with eigenvalues ≥1, explaining 93.53 % of the total variation. The first five components, which together explained 88.03 % of the variation, were retained for interpretation (Table 6). PC1 (35.30 %) was positively correlated with 5159_1, 5159_2, 9269_1, B05-CANN1_1, B05-CANNA1_2, B01-CANNA1_2, ANUCS304_2, NH09_1, NH09_2, ANUCS203_2, and negatively correlated with H09-CANN2_1, H09-CANN2_2, 4910_1, 3735_2, 9043_1, 9043_2, and ANUCS307_2. PC2 (20.30 %) was positively correlated with ANUCS303_1, ANUCS303_2, B01-CANN1_1, C11-CANN1_1, ANUCS304_1, and 9043_1, and negatively correlated with 4910_2, ANUCS301_1, and ANUCS301_2. PC3 (14.67 %) was positively correlated with ANUCS305_1.1, ANUCS205_1, and ANUCS205_2, and negatively correlated with 3735_1 and H06-CANN2_2. PC4 (11.01 %) was positively correlated with 9269_2, ANUCS305_1.2, ANUCS302_1, and ANUCS302_2, and negatively correlated with ANUCS206_2. PC5 (6.75 %) was positively correlated with ANUCS203_1 and the 1528_1 allele calls.

3.4.3.3. Hierarchal cluster analysis. Agglomerative hierarchical clustering for means of STR markers classified the 11 landraces into four groups (Fig. 9). Cluster 1 comprised of 'M1', 'L2', 'B3', 'B1' and 'L1'. Cluster 2 comprised of 'Hemp' and 'High-CBD', and cluster 3 comprised of 'High-THC'. Cluster 4 comprised of 'B2', 'H1' and 'M2'.

4. Discussion

4.1. Physical traits

The findings of the current study showed diverse physical traits of cannabis seeds, with significant variations observed among different varieties. Notably, the seeds of 'B1' and 'M2' exhibited relatively high mean values for width, thickness, length, geometric mean diameter, and surface area compared to other varieties. In contrast, 'M1' and 'B2' seeds displayed relatively lower mean values for width, thickness, geometric

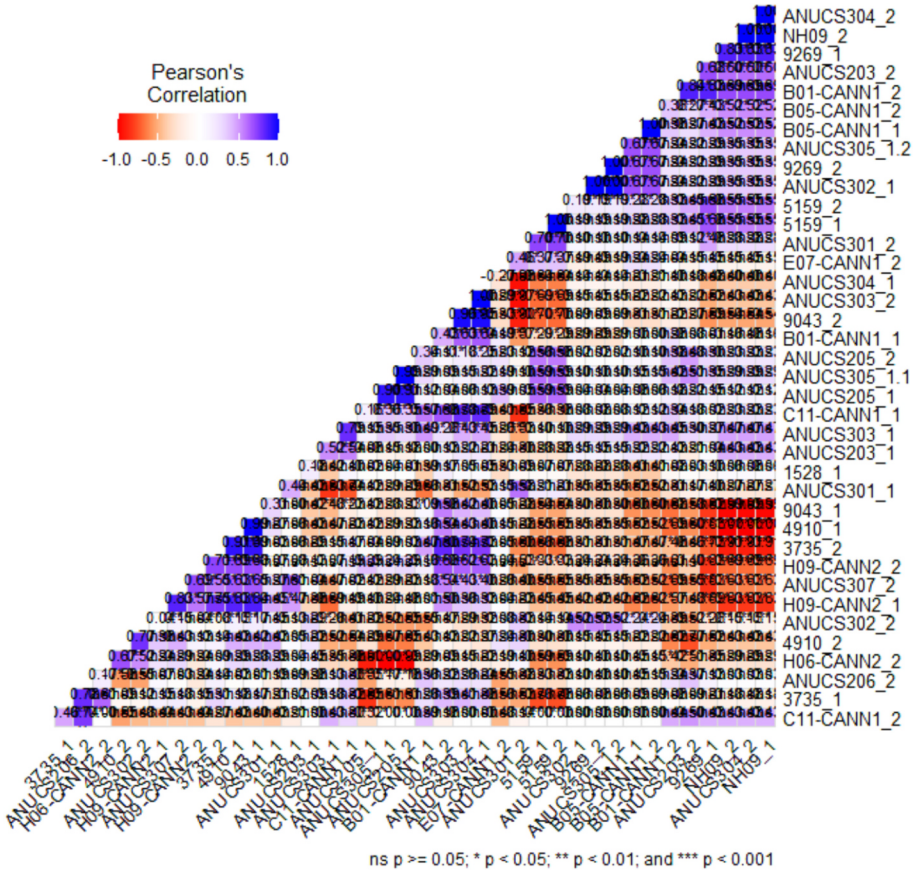

Fig. 7. The Pearson's correlation for Short Tandem Repeats (STR) markers assessed in eleven cannabis varieties.

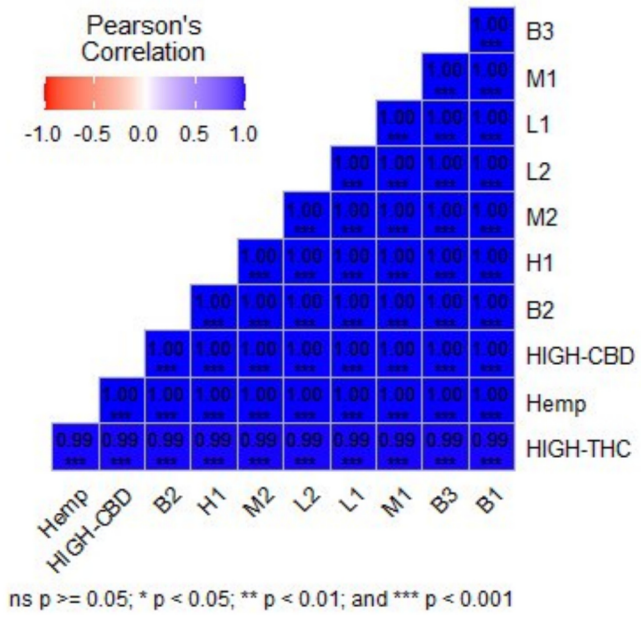

Fig. 8. The Pearson's correlation for eleven cannabis varieties based on their STR markers.

mean diameter, and surface area compared to other varieties. Thus, when designing seed-handling equipment, the dimensions of the seeds' length, width, and thickness must be taken into consideration to determine the appropriate aperture sizes (Taheri-Garavand et al., 2012). On the other hand, the geometric mean diameter is useful for characterizing the seeds with dimensions of irregular shapes (Taheri-Garavand et al., 2012). Additionally, the mean geometric mean diameter is useful for estimating the projected area of a particle moving in a turbulent or near-turbulent air stream (Taheri-Garavand et al., 2012).

Pearson's correlation analysis revealed significant relationships among physical parameters and different cannabis varieties. Notably, seeds with greater thickness correlated with longer lengths, while wider seeds tended to be longer and thicker. Additionally, seeds with larger geometric mean diameters exhibited larger surface areas, widths, thicknesses, and lengths. The principal component analysis (PCA) revealed that PC1 variation was contributed by seed width, thickness, surface area, and geometric mean diameter, which was key in significant differences among the varieties. In contrast, PC2 variation was contributed by seed sphericity and aspect ratio. The Agglomerative hierarchical clustering of physical traits grouped the 11 landraces into four distinct clusters. Cluster 1 included 'B2', 'B3', 'M1', and 'L1', while Cluster 2 comprised 'B1', 'M2', 'Hemp', 'High-CBD', and 'H1'. Cluster 3 consisted of only 'L2', and Cluster 4 was represented by 'High-THC'. Notably, seeds from 'B1' and 'M2' within Cluster 2 displayed relatively higher mean values for width, thickness, length, geometric mean diameter, and surface area compared to other landraces. Contrarily, Cluster 1 consisted of seeds with relatively low surface area, width, and thickness.

4.2. Mineral element composition

The cannabis varieties contained major elements detected by SEM-EDX, including C and O, together with mineral elements such as K, P, Mg, Si, Ca, Cl, and S. In addition, trace and ultra-trace elements, including Fe and Al, were also detected. However, the relative

Table 6
Summary of factor loadings, eigenvalue, percent, and cumulative variation for eleven cannabis varieties based on their Short Tandem Repeats (STR) markers.

Traits	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
H09-CANN2_1	−0.860	−0.146	0.216	−0.092	0.056
H09-CANN2_2	−0.682	0.253	0.455	0.345	0.142
5159_1	0.732	−0.346	0.543	−0.027	−0.025
5159_2	0.732	−0.346	0.543	−0.027	−0.025
9269_1	0.901	0.004	0.009	−0.260	−0.109
9269_2	0.469	0.034	−0.321	0.658	−0.119
ANUCS303_1	0.403	0.767	−0.196	0.213	0.012
ANUCS303_2	−0.552	0.808	0.016	0.160	0.085
B05-CANN1_1	0.636	0.082	−0.240	0.530	−0.215
B05-CANN1_2	0.636	0.082	−0.240	0.530	−0.215
B01-CANN1_1	−0.131	0.730	0.150	−0.120	−0.416
B01-CANN1_2	0.724	0.352	0.007	−0.232	−0.097
4910_1	−0.925	−0.097	0.170	0.132	−0.292
4910_2	−0.474	−0.679	−0.237	0.277	0.196
3735_1	−0.352	0.364	−0.753	−0.318	−0.186
3735_2	−0.872	0.309	0.174	0.122	−0.272
ANUCS206_2	−0.064	0.456	−0.381	−0.617	−0.305
ANUCS301_1	−0.246	−0.774	0.202	−0.350	0.139
ANUCS301_2	0.452	−0.859	0.135	−0.092	−0.152
C11-CANN1_1	0.049	0.938	−0.101	0.081	0.295
C11-CANN1_2	0.357	0.423	−0.343	−0.552	−0.255
ANUCS305_1.1	0.419	0.430	0.777	0.127	0.049
ANUCS305_1.2	0.469	0.034	−0.321	0.658	−0.119
ANUCS304_1	−0.514	0.829	0.073	0.167	0.088
ANUCS304_2	0.925	0.097	−0.170	−0.132	0.292
ANUCS205_1	0.267	0.164	0.866	0.323	0.176
ANUCS205_2	0.353	0.433	0.815	0.088	0.036
NH09_1	0.925	0.097	−0.170	−0.132	0.292
NH09_2	0.925	0.097	−0.170	−0.132	0.292
ANUCS302_1	0.469	0.034	−0.321	0.658	−0.119
ANUCS302_2	−0.107	−0.566	−0.412	0.665	0.093
9043_1	−0.923	−0.103	0.157	0.112	−0.240
9043_2	−0.646	0.671	0.026	0.202	0.128
ANUCS203_1	0.168	0.254	−0.440	0.267	0.658
ANUCS203_2	0.676	0.293	0.152	−0.328	−0.157
1528_1	−0.191	−0.088	0.070	−0.344	0.761
H06-CANN2_2	−0.419	−0.430	−0.777	−0.127	−0.049
ANUCS307_2	−0.807	−0.049	0.030	0.040	0.454
E07-CANN1_2	0.082	−0.364	0.496	0.212	−0.357
Eigenvalue	**13.768**	**7.916**	**5.721**	**4.293**	**2.633**
Variability (%)	**35.302**	**20.297**	**14.669**	**11.007**	**6.752**
Cumulative (%)	**35.302**	**55.600**	**70.269**	**81.276**	**88.029**

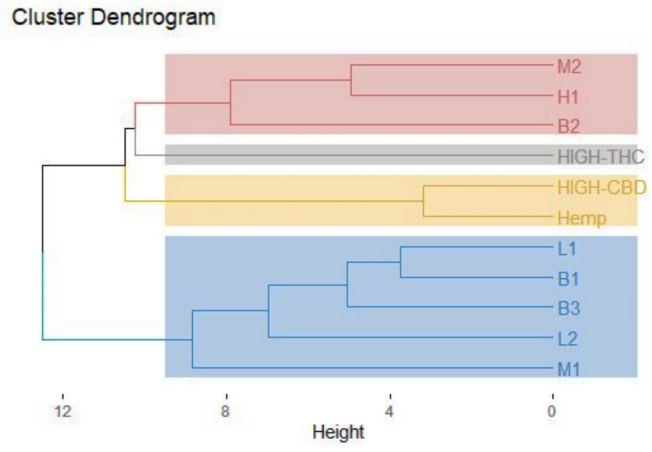

Fig. 9. Agglomerative hierarchical cluster for eleven cannabis varieties based on their STR markers.

abundance of these elements varied among the different cannabis varieties. No single element was consistently dominant; rather, different minerals prevailed in different landraces. For instance, K showed the highest relative abundance in 'High-CBD', 'H1', 'Hemp' and 'B3'. Similarly, Ca was most abundant in 'M1' and 'L2', Si dominated in 'B2' and 'L1', Fe dominated in 'B1'and 'M2', while Al dominated in 'High-THC'. This genotype-specific variation highlights the substantial diversity in mineral composition within the seed embryo tissues.

The major macro elements, such as P, K, Mg, and Ca, as well as trace elements, such as Fe, were previously reported in hempseeds (Farinon et al., 2020). The findings of the current study on the variation in mineral profile of seeds are consistent with Farinon et al. (2020), who reported that it can vary significantly depending on a number of factors, including the type of plant genotype, the type and amount of fertilizer applied, and the mineral composition of the soil. The findings on potassium are also consistent with the reports by Alonso-Esteban et al. (2022), who reported a high relative abundance of potassium in hemp seeds, especially in hulled seeds. Barčauskaitė et al. (2022) also found that hemp seeds were rich in K.

In contrast, the results of the current study are inconsistent with the findings by Farinon et al. (2020) and Alonso-Esteban et al. (2022), who reported that phosphorus was the most abundant mineral found in both whole and hulled hemp seeds. The inconsistency of the findings could be attributed to the differences in the analytical techniques employed and sample preparation. The current study employed SEM-EDX on cryo-fractured seed sections, which identifies the spatial distribution and relative abundance of elements within the embryo tissue. This differs fundamentally from bulk analytical techniques such as Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES or Inductively Coupled Plasma Mass Spectrometry (ICP-MS) used in previous studies (Mattila et al., 2018; Lan et al., 2019), where homogenized whole or dehulled seeds were digested to yield quantitative concentrations (mg/100 g).

Furthermore, specific varieties exhibited exceptionally low levels of certain elements. For instance, 'B2', 'M2', and 'L1' having the lowest Al levels, and 'High-CBD' varieties, along with 'B3', 'B1', and 'High-THC' having lowest chlorine levels. Moreover, 'H1' and 'Hemp' demonstrated the lowest iron levels. Iron, however, was the only microelement in the cannabis varieties that was present in significant amounts. Notably, iron was found to be the primary microelement in whole hemp seeds by Alonso-Esteban et al. (2022) and Barčauskaitė et al. (2022). Whereas 'Hemp' and 'L2' showed the lowest sulfur levels. However, it is important to highlight that several anti-nutrient elements in food, such as phytates and oxalates, which are included in hemp seeds, may reduce the bioavailability of some mineral elements (Alonso-Esteban et al., 2022).

In general, there were minimal correlations observed among the mineral elements, except for S, which exhibited a strong positive correlation with K and P. Additionally, a positive correlation was found between P and K. Conversely, C demonstrated a significant negative correlation with O and Cl. The correlation between these elements is important for the development and selection of a variety of the desired minerals, which could improve the level of another mineral (Yeken et al., 2018). However, the interactions between the mineral elements may be influenced by genetic and environmental factors (Yeken et al., 2018; Nadeem, 2021). Nonetheless, a very strong positive correlation was observed among the seeds of various varieties, indicating a similarity in their mineral element composition.

The findings of the current study showed distinct patterns of variation in mineral element composition among the cannabis genotypes. PC1 primarily negatively associated with C and positively correlated with Cl, K, S, and Mg. PC2 was linked to Al, P, and O, and PC3 was related to Ca. On the other hand, the Agglomerative hierarchical clustering approach showed similarities and differences in mineral element composition among the cannabis landraces. These findings provide insights into potential relationships and distinctions in their genetic backgrounds. For instance, Cluster 1 included 'L2', 'High-THC', 'B2', 'Hemp', and 'M2', indicating similarities in their mineral element composition. Cluster 2 consisted of 'H1', while cluster 3 comprised 'B3' and 'High-CBD'. Finally, cluster 4 encompassed 'B1', 'M1', and 'L1'.

4.3. Short Tandem Repeat markers

The findings of Pearson's correlation analysis indicated a strong network of interactions, both positive and negative, between the different alleles, with multiple significant correlations among them. The correlations observed suggest complex interactions that may provide insight into the underlying biological processes and genetic linkages. The perfect correlations observed, such as those between 5159_1 and 5159_2 and ANUCS304_2 with NH09_1 and NH09_2, highlighted a very close genetic linkage. Similarly, the correlation analysis revealed a strong relationship among the different varieties, suggesting genetic linkage. The STR markers used in this study are among those commonly applied and validated for species identification, gender determination, and individual profiling of *C. sativa* (Xia et al., 2022). Therefore, the inconsistency between the current study and findings by Köhnemann et al. (2012) and Marcińska et al. (2022) may be associated with plant material used as they analyzed leaves or inflorescence. Thus, this could imply that there is still a need for the development and validation of a unique identification technique for cannabis seed variety using the STR system. Otherwise, other genes may be critical for the characterization of the varieties used in the current study at a seed level. Tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS) gene sequencing from 21*C. sativa* L. varieties, for example, has been used to distinguish drug-type from fiber-type cannabis seeds, achieving an AUC of 1.00 on that specific dataset in the study by Cascini et al. (2019).

Principal Component Analysis (PCA) further explained the genetic structure, identifying six principal components, with the first five explaining a substantial portion of the total variance. The PCA identified six principal components, explaining 93.53 % of the total variation. PC1, the most influential component, positively correlated with several markers while negatively correlated with others. PC2, PC3, PC4, PC5, and PC6 each contributed to explaining variance in the dataset with distinct marker associations. Agglomerative hierarchical clustering grouped the 11 landraces into four clusters, revealing patterns of similarity and dissimilarity in their STR marker profiles. Cluster 1 included 'M1', 'L2', 'B3', 'B1', and 'L1', while Cluster 2 comprised 'Hemp' and 'High-CBD'. 'High-THC' formed Cluster 3, and Cluster 4 consisted of 'B2', 'H1', and 'M2'.

In this study, STR markers which are commonly used as neutral genetic markers (Di Nunzio et al., 2024) were applied to evaluate whether variation at these loci could assist in distinguishing cannabis chemotypes. Since STRs are not directly linked to cannabinoid biosynthesis genes, no strong correlation with chemotype was expected, and the landraces could not be reliably classified as 'High-CBD', 'High-THC', or 'Hemp'. Nevertheless, their high polymorphism makes them valuable for capturing genetic diversity and structure. In future applications, STRs could be combined with functional gene markers such as THCAS and CBDAS, or integrated with metabolic profiling, to achieve more robust chemotype discrimination and to support the development of cannabis-specific functional markers.

In summary, the correlation revealed similarities among the cannabis varieties. However, further analysis, particularly agglomerative hierarchical clustering, showed differences among the varieties. This study identified four distinct clusters among the landraces by evaluating their physical traits, mineral element composition, and STR markers separately using agglomerative hierarchical clustering. However, the clustering based on mineral composition grouped the landraces differently compared to the clustering based on physical traits. Similarly, clustering based on STR markers resulted in distinct groupings that differed from those based on physical traits and mineral composition. These differences across clustering analyses highlighted the complex nature of the diversity of the varieties and the need for consideration of multiple characteristics for comprehensive landrace characterization and classification.

5. Conclusion

Physical parameters, particularly of 'M1' and 'B2', are fundamental in determining suitable aperture sizes for efficient processing and aiding in characterizing seeds with irregular shapes. These findings emphasize the need for variety-specific considerations in equipment design and offer valuable insights for breeding programs targeting seed trait optimization, ultimately enhancing processing techniques and equipment efficiency. Potassium, calcium, iron, aluminum and silicon were the major macro elements, and their distribution differed among varieties. Understanding these variations is vital for optimizing agricultural practices and processing techniques for facilitating the development of high-quality hemp-based products. Strong correlations between certain alleles suggested close genetic linkage, emphasizing the importance of understanding genetic relationships among cannabis varieties. Generally, the correlation highlighted similarities among cannabis varieties. However, clustering based on mineral composition differed from that based on physical traits, and STR markers also resulted in distinct groupings, indicating a complex relationship among these characteristics. Therefore, the landraces could not be categorized as 'High-CBD', 'High-THC', or 'Hemp' based on the measured parameters. Further studies are recommended to develop and validate a unique identification technique for cannabis seed varieties using the STR system. Given that STRs are neutral markers not directly linked to cannabinoid biosynthesis, complementary approaches are warranted. For example, Cascini et al. (2019) successfully discriminated between drug-type and fiber-type cannabis seeds using Tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS) gene sequencing. Future work should explore integrating STR markers with functional gene markers or with metabolic profiling to achieve more robust chemotype discrimination and support the development of cannabis-specific functional markers.

CRediT authorship contribution statement

Sabeliwe Langa: Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Lembe Samukelo Magwaza:** Writing – review & editing, Supervision, Resources, Conceptualization. **Asanda Mditshwa:** Writing – review & editing, Supervision, Conceptualization. **Samson Zeray Tesfay:** Writing – review & editing, Supervision, Conceptualization.

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

We gratefully acknowledge the cannabis growers in Ladysmith, Msinga, Hammarsdale, and Bergville, particularly Gogo Gwijo, Khekhe Mtshali, and Sizwe, for their invaluable assistance in providing landrace cannabis seeds for this study. We also thank Durbanseedco under Grohydro (Pty) Ltd and Nirvana Shop South Africa for their additional seed donations. Finally, we express our sincere appreciation to the Moses Kotane Institute for its financial support, which made this research possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plgene.2025.100564.

Data availability

Research data is available upon request and can be requested by contacting the corresponding author of the article.

References

Adhikary, D., Kulkarni, M., El-Mezawy, A., Mobini, S., Elhiti, M., Gjuric, R., Ray, A., Polowick, P., Slaski, J.J., Jones, M.P., Bhowmik, P., 2021. Medical cannabis and industrial hemp tissue culture: present status and future potential. Front. Plant Sci. 12, 627240.

Allaf, S., Lim, J.S., Buckley, N.A., Cairns, R., 2023. The impact of cannabis legalization and decriminalization on acute poisoning: a systematic review. Addiction. https://doi.org/10.1111/add.16280.

Alonso-Esteban, J.I., Torija-Isasa, M.E., de Cortes Sánchez-Mata, M., 2022. Mineral elements and related antinutrients, in whole and hulled hemp (*Cannabis sativa* L.) seeds. J. Food Compos. Anal. 109, 104516.

Barčauskaitė, K., Žydelis, R., Ruzgas, R., Bakšinskaitė, A., Tilvikienė, V., 2022. The seeds of industrial hemp (*Cannabis sativa* L.) a source of minerals and biologically active compounds. Journal of Natural Fibers 19 (16), 13025–13039.

Broseus, J., Anglada, F., Esseiva, P., 2010. The differentiation of fibre- and drug type Cannabis seedlings by gas chromatography/mass spectrometry and chemometric tools. Forensic Sci. Int. 200 (1–3), 87–92.

Cascini, F., Farcomeni, A., Migliorini, D., Baldassarri, L., Boschi, I., Martello, S., Amaducci, S., Lucini, L., Bernardi, J., 2019. Highly predictive genetic markers distinguish drug-type from fiber-type *Cannabis sativa* L. Plants 8 (11), 496.

Cerrato, A., Citti, C., Cannazza, G., Capriotti, A.L., Cavaliere, C., Grassi, G., Marini, F., Montone, C.M., Paris, R., Piovesana, S., Laganà, A., 2021. Phytocannabinomics: untargeted metabolomics as a tool for cannabis chemovar differentiation. Talanta 230, 122313.

Dalusong, V.G., Santos, L.H., Perez, L.M., 2019. Establishment of core Short Tandem Repeat (STR) markers for rice (*Oryza sativa* L.) varietal identification. Philipp. J. Crop Sci. (Philippines) 44 (3).

Danziger, N., Bernstein, N., 2021. Plant architecture manipulation increases cannabinoid standardization in 'drug-type' medical cannabis. Ind. Crop. Prod. 167, 113528. https://doi.org/10.1016/j.indcrop.2021.113528.

de Meijer, E.P.M., 1994. Diversity in Cannabis. Doctoral thesis,. Wageningen Agricultural University, Wageningen, The Netherlands.

Di Nunzio, M., Agostini, V., Alessandrini, F., Barrot-Feixat, C., Berti, A., Bini, C., Bottinelli, M., Carnevali, E., Corradini, B., Fabbri, M., Fattorini, P., 2021. A Ge. FI–ISFG European collaborative study on DNA identification of *Cannabis sativa* samples using a 13-locus multiplex STR method. Forensic Sci. Int. 329, 111053.

Di Nunzio, M., Pieri, M., Gangitano, D., Di Nunzio, C., Tinto, N., Niola, M., Barrot-Feixat, C., 2024. Leveraging genetics to support forensic toxicology analysis: demonstrating concordance among marijuana samples. J. Appl. Res. Med. Aromat. Plants 43, 100598.

Farinon, B., Molinari, R., Costantini, L., Merendino, N., 2020. The seed of industrial hemp (*Cannabis sativa* L.): nutritional quality and potential functionality for human health and nutrition. Nutrients 12 (7), 1935.

Feng, L., Zhu, S., Zhang, C., Bao, Y., Feng, X., He, Y., 2018. Identification of maize kernel vigor under different accelerated aging times using hyperspectral imaging. Molecules 23 (12), 3078.

Gharibzahedi, S.M.T., Mousavi, S.M., Razavi, S.H., Akavan-Borna, M., 2009, October. Determination of nutritional and physical properties of sesame seed (*Sesamum Indicum* L.). In: Proceedings of Rural Development Conference. Lithuanian University, pp. 304–309.

Gilmore, S., Peakall, R., Robertson, J., 2003. Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. Forensic Sci. Int. 131 (1), 65–74.

Gojiya, D.K., Dobariya, U.D., Pandya, P.A., Gojiya, K.M., 2020. Studies on physical properties of peanut seed. Acta Sci. Agric. 4 (3), 1–5.

Gouws, J., 2017. Cultivation of cannabis and manufacture of cannabis-related pharmaceutical products for medicinal and research purposes. In: Medicines Control Council: Application Guideline, V1, 2017.

Jin, D., Jin, S., Chen, J., 2015, October. Cannabis classification systems and growth trends of the North American medical cannabis industry. In: Second Annual Academic Conference of the Specialty Committee of TCM Pharmacognosy of the World Federation of Chinese Medicine Societies, pp. 23–27.

Jin, B., Zhang, C., Jia, L., Tang, Q., Gao, L., Zhao, G., Qi, H., 2022. Identification of rice seed varieties based on near-infrared hyperspectral imaging technology combined with deep learning. ACS Omega 7 (6), 4735–4749.

Kaliniewicz, Z., Jadwisienczak, K., Żuk, Z., Lipinski, A., 2021. Selected physical and mechanical properties of hemp seeds. BioResources 16 (1), 1411.

Köhnemann, S., Nedele, J., Schwotzer, D., Morzfeld, J., Pfeiffer, H., 2012. The validation of a 15 STR multiplex PCR for Cannabis species. Int. J. Legal Med. 126, 601–606.

Lan, Y., Zha, F., Peckrul, A., Hanson, B., Johnson, B., Rao, J., Chen, B., 2019. Genotype x environmental effects on yielding ability and seed chemical composition of industrial hemp (*Cannabis sativa* L.) varieties grown in North Dakota, USA. J. Am. Oil Chem. Soc. 96 (12), 1417–1425.

Lapierre, É., Monthony, A.S., Torkamaneh, D., 2023. Genomics-based taxonomy to clarify cannabis classification. Genome 66 (08), 202–211.

Marcińska, M., Wróbel, M., Sekuła, A., Kowalczyk, M., Sekuła, K., Czarny, J., Powierska-Czarny, J., Nawotka, R., Bachliński, R., Duszyńska, A., Makowska, Ż., 2022. Genetic identification of Cannabis varieties based on marker STR analysis: preliminary studies. Forens. Sci. Int. Genet. Suppl. Ser. 8, 341–343.

Mattila, P., Mäkinen, S., Eurola, M., Jalava, T., Pihlava, J.M., Hellström, J., Pihlanto, A., 2018. Nutritional value of commercial protein-rich plant products. Plant Foods Hum. Nutr. 73 (2), 108–115.

Mihoc, M., Pop, G., Alexa, E., Radulov, I., 2012. Nutritive quality of romanian hemp varieties (*Cannabis sativa* L.) with special focus on oil and metal contents of seeds. Chem. Cent. J. 6, 1–12.

Monti, M.C., Frei, P., Weber, S., Scheurer, E., Mercer-Chalmers-Bender, K., 2022. Beyond Δ^9-tetrahydrocannabinol and cannabidiol: chemical differentiation of cannabis varieties applying targeted and untargeted analysis. Anal. Bioanal. Chem. 414 (13), 3847–3862.

Nadeem, M.A., 2021. Exploring the phenotypic diversity for seed mineral contents in Turkish faba bean germplasm. Int. J. Agric. Wildl. Sci. (IJAWS) 7 (3), 540–550. https://doi.org/10.24180/ijaws.949496.

Naraine, S.G., Small, E., Laursen, A.E., Campbell, L.G., 2020. A multivariate analysis of morphological divergence of "seeds" (achenes) among ruderal, fibre, oilseed, dioecious/monoecious and marijuana variants of *Cannabis sativa* L. Genet. Resour. Crop. Evol. 67 (3), 703–714.

Perea-Flores, M.J., Chanona-Perez, J.J., Garibay-Febles, V., Calderon-Dominguez, G., Terrés-Rojas, E., Mendoza-Perez, J.A., Herrera-Bucio, R., 2011. Microscopy techniques and image analysis for evaluation of some chemical and physical properties and morphological features for seeds of the castor oil plant (*Ricinus communis*). Ind. Crop. Prod. 34 (1), 1057–1065.

Rashid, N., Zafar, M., Ahmad, M., Malik, K., Haq, I.U., Shah, S.N., Mateen, A., Ahmed, T., 2018. Intraspecific variation in seed morphology of tribe *vicieae* (Papilionoidae) using scanning electron microscopy techniques. Microsc. Res. Tech. 81 (3), 298–307.

Rehman, H.M., Cooper, J.W., Lam, H.M., Yang, S.H., 2019. Legume biofortification is an underexploited strategy for combatting hidden hunger. Plant Cell Environ. 42 (1), 52–70.

Salentijn, E.M.J., Zhang, Q., Amaducci, S., Yang, M., Trindade, L.M., 2015. New developments in fiber hemp (*Cannabis sativa* L.) breeding. Ind. Crop. Prod. 68, 32–41.

Sastry, D.V.S.S.R., Upadhyaya, H.D., Srinivas, T.R., 2019. Variation for seed physical and hydration properties of chickpea (*Cicer arietinum* L.) mini core collection and their relevance to conservation and utilization. Plant Genet. Resour. 17 (4), 311–324.

Sawler, J., Stout, J.M., Gardner, K.M., Hudson, D., Vidmar, J., Butler, L., Page, J.E., Myles, S., 2015. The genetic structure of marijuana and hemp. PLoS One 10 (8), e0133292.

Schluttenhofer, C., Yuan, L., 2017. Challenges towards revitalizing hemp: a multifaceted crop. Trends Plant Sci. 22 (11), 917–929.

Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. 18 (2), 233–234.

Taheri-Garavand, A., Nassiri, A., Gharibzahedi, S.M.T., 2012. Physical and mechanical properties of hemp seed. Int. Agrophys. 26 (2).

Tipparat, P., Kunkaew, W., Julsrigival, S., Pinmanee, S., Natakankitkul, S., 2014. Classification of cannabis plants grown in Northern Thailand using physico-chemical properties. J. Nat. Sci. Res. 10, 46–54.

Vergara, D., Baker, H., Clancy, K., Keepers, K.G., Paul Mendieta, J., Pauli, C.S., Tittes, S. B., White, K.H., Kane, N.C., 2016. Genetic and genomic tools for *Cannabis sativa*. Crit. Rev. Plant Sci. 35, 5–6.

Woodbridge, M., 2018. A Primer to Medicinal cannabis: An Introductory Text to the Therapeutic Use of cannabis. Bedrocan International.

Xia, R., Tao, R., Qu, Y., Zhang, X., Yu, H., Yuan, C., Zhang, S., Li, C., 2022. Development and validation of a novel and fast detection method for cannabis sativa: a 19-plex short tandem repeat typing system. Front. Plant Sci. 13, 837945.

Yeken, M.Z., Akpolat, H., Karaköy, T., Çiftçi, V., 2018. Assessment of mineral content variations for biofortification of the bean seed. Uluslararası Tarım ve Yaban Hayatı Bilim. Derg 4, 261–269. https://doi.org/10.24180/ijaws.455311.

Zhang, C., Zhao, Y., Yan, T., Bai, X., Xiao, Q., Gao, P., Li, M., Huang, W., Bao, Y., He, Y., Liu, F., 2020. Application of near-infrared hyperspectral imaging for variety identification of coated maize kernels with deep learning. Infrared Phys. Technol. 111, 103550.

https://dspace.org/

Genetic and phenotypic diversity in cannabis genotypes: insights from seed dimensions, mineral profiles, and Short Tandem Repeats (STR) markers

Langa, Sabeliwe; Magwaza, Lembe Samukelo; Mditshwa, Asanda; Tesfay, Samson Zeray

2026-03
Elsevier
https://doi.org/10.1016/j.plgene.2025.100564
Langa S, Magwaza LS, Mditshwa A, Tesfay SZ. (2026) Genetic and phenotypic diversity in cannabis genotypes: insights from seed dimensions, mineral profiles, and Short Tandem Repeats (STR) markers. Plant Gene, Volume 45, March 2026, Article number 100564

Downloaded from a DSpace repository. https://dspace.org/