PAB Pesquisa Agropecuária Brasileira 60 years

Genetics/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Characterization of the variability of the lipid profile of cashew kernels from a genebank by MicroNIR spectroscopy

Abstract – The objective of this work was to characterize the physicochemical variability of cashew (*Anacardium occidentale*) nut kernels across 50 accessions of the Brazilian cashew germplasm bank, evaluate the potential of portable near-infrared (NIR) spectroscopy to predict quality parameters and identify accessions with superior lipid content. Proximate analyses of moisture, ash, protein, and lipid contents, in addition to fatty acid profiling, revealed a wide genetic variability among the accessions. Lipid content ranged from 38.17 to 47.90%, and protein content from 13.66 to 22.41%, indicating a significant negative correlation between these constituents. Accessions with a high lipid content presented a high proportion of unsaturated fatty acids (~78%), especially oleic acid (52–56%), evidencing their potential for nutritional and industrial applications. The multivariate regression models based on interval partial least squares applied to NIR spectra of whole kernels showed a satisfactory performance in predicting lipid content. Therefore, portable spectroscopy has potential for rapid phenotyping and high-throughput screening, and accessions with desirable lipid profiles can be used as genetic resources for breeding programs and industrial exploitation.

Index terms: *Anacardium occidentale*, fatty acids, genetic variability, germplasm conservation, near-infrared spectroscopy.

Caracterização da variabilidade do perfil lipídico de amêndoas de castanha-de-caju de banco de germoplasma por espectroscopia MicroNIR

Resumo – O objetivo deste trabalho foi caracterizar a variabilidade físico-química de amêndoas de castanha-de-caju (*Anacardium occidentale*) provenientes de 50 acessos do Banco Ativo de Germoplasma de Cajueiro, avaliar o potencial da espectroscopia de infravermelho próximo (NIR) portátil para predizer parâmetros de qualidade e identificar acessos com teor lipídico superior. As análises físico-químicas de umidade, cinzas, proteínas e lipídios, além do perfil de ácidos graxos revelaram ampla variabilidade genética entre os acessos. O teor de lipídios variou de 38,17 a 47,90%, e o teor de proteínas de 13,66 a 22,41%, o que indica correlação negativa significativa entre esses constituintes. Os acessos com maior teor lipídico apresentaram alta proporção de ácidos graxos insaturados (~78%), especialmente ácido oleico (52–56%), o que evidencia seu potencial para aplicações nutricionais e industriais. Os modelos de regressão multivariada baseados em mínimos quadrados parciais por intervalos, aplicados aos espectros NIR de amêndoas inteiras apresentaram desempenho satisfatório para a predição do teor de lipídios. Portanto, a espectroscopia portátil tem potencial para fenotipagem rápida e triagem em

Febriana Correia Ximenes de Carvalho
Universidade Federal do Ceará, Departamento de Biologia, Programa de Pós-Graduação em Sistemática, Uso e Conservação da Biodiversidade, Campus do Pici, Fortaleza, CE, Brazil. E-mail: febriana9293@gmail.com

Elenilson Godoy Alves Filho
Universidade Federal do Ceará, Departamento de Engenharia de Alimentos, Campus do Pici, Fortaleza, CE, Brazil.
E-mail: elenilson.godoy@ufc.br

Ebenezer de Oliveira Silva
Embrapa Agroindústria Tropical, Fortaleza, CE, Brazil. E-mail: ebenezer.silva@embrapa.br

Fernando Antonio Souza de Aragão
Embrapa Agroindústria Tropical, Fortaleza, CE, Brazil. E-mail: fernando.aragao@embrapa.br

Men de Sa Moreira de Souza Filho
Embrapa Agroindústria Tropical, Fortaleza, CE, Brazil. E-mail: men.souza@embrapa.br

Ana Cecília Ribeiro de Castro(✉)
Embrapa Agroindústria Tropical, Fortaleza, CE, Brazil. E-mail: cecilia.castro@embrapa.br

✉ Corresponding author

Received
February 23, 2026

Accepted
July 30, 2026

How to cite
CARVALHO, F.C.X. de; ALVES FILHO, E.G.; SILVA, E. de O.; ARAGÃO, F.A.S. de; SOUZA FILHO, M. de S.M. de; CASTRO, A.C.R. de. Characterization of the variability of the lipid profile of cashew kernels from a genebank by MicroNIR spectroscopy. **Pesquisa Agropecuária Brasileira**, v.61, e04394, 2026. DOI: https://doi.org/10.1590/S1678-3921.pab2026.v61.04394.

This is an open-access article distributed under the Creative Commons Attribution 4.0 International License

larga escala, e os acessos com perfis lipídicos desejáveis podem ser utilizados como fonte de recursos genéticos para programas de melhoramento e exploração industrial.

Termos para indexação: *Anacardium occidentale*, ácidos graxos, variabilidade genética, conservação de germoplasma, espectroscopia no infravermelho próximo.

Introduction

The cashew nut tree *(Anacardium occidentale* L.) is a fruit-bearing species of significant importance for food security and climate change adaptation in tropical regions because of its resilience to water stress and a wide cultivation range (Grüter et al., 2022). The cashew nut kernel (CNK) is the primary commercial product, representing a global market valued at approximately US$12 billion (Hirpara et al., 2025).

Recognized for its exceptional nutritional quality, the CNK is a rich source of healthy lipids, proteins of high biological value, and essential minerals. Among its lipid constituents, the predominance of oleic acid, a stable monounsaturated fatty acid, contributes to both the nutritional and technological value of the kernels, which are associated with cardiovascular health benefits and increased oxidative stability in processed products (Chen et al., 2023). Because these nutritional and compositional attributes may vary substantially among genotypes, understanding such variability is essential for identifying accessions with superior industrial and commercial potential.

The Brazilian cashew germplasm bank represents a strategic genetic resource for the conservation and utilization of cashew diversity. This collection contains a broad range of genetic variability from which superior materials can be identified for breeding programs and industrial applications. However, the characterization of large germplasm collections remains challenging because conventional physicochemical analyses are labor-intensive, costly, and destructive, limiting the number of samples that can be evaluated and consequently slowing the identification of elite genotypes.

Near-infrared (NIR) spectroscopy combined with chemometric tools has emerged as a rapid, reliable, and non-destructive alternative for characterizing food and agricultural products. Portable NIR devices further expand these advantages by enabling real-time analysis with minimal sample preparation and reduced dependence on laboratory infrastructure, thereby facilitating high-throughput phenotyping approaches (Caetano et al., 2025). Although NIR spectroscopy has been widely employed in the quality assessment of agricultural commodities, its application as a strategic tool for screening genetically diverse germplasm collections remains limited.

In this context, the integration of portable NIR spectroscopy and physicochemical characterization offers an opportunity to accelerate the identification of valuable genetic resources within genebanks. By linking non-destructive spectral phenotyping with kernel composition evaluation, this approach can support the rapid detection of accessions with desirable lipid profiles and contribute to the industrial valorization of conserved genetic resources.

The objective of this work was to characterize the physicochemical variability of cashew nut kernels across 50 accessions of the Brazilian cashew germplasm bank, evaluate the potential of portable NIR spectroscopy to predict quality parameters and identify accessions with superior lipid content.

Materials and Methods

Cashew nuts were harvested from 50 cashew accessions belonging to the Brazilian cashew germplasm bank, located at the experimental field of Embrapa Agroindústria Tropical, in Pacajus, Ceará, Brazil (4°11'07"S, 38°30'07"W; at 60 m of altitude). The harvest was conducted during the 2024 season; subsequently, the samples were transported to the Plant Breeding and Genetic Resources Laboratory of Embrapa Agroindústria Tropical and stored at 18°C for six months prior to processing.

The nuts were initially weighed and heat-treated in a non-pressurized autoclave at 80°C for 10 min, followed by sun drying for 24 hours. The shells were then removed using a manual cutting machine to extract the CNKs. To facilitate manual skin removal, the kernels were further dried in an oven at 80°C for 8 hours.

For spectral analysis, 10 whole kernels were randomly selected from each of the 50 accessions. Three replicates were performed per accession, totaling 150 samples. Spectral data were collected in reflectance mode using a portable MicroNIR 1700 spectrometer (VIAVI Solutions, Chandler, AZ, USA) operating in the 908–1676 nm range. Background

noise was accounted for by acquiring reference spectra prior to sample collection.

Spectra were recorded in logarithmic mode (log 1/R). For each sample, five measurements were taken across different kernel regions and averaged for subsequent analysis. All measurements were performed under controlled laboratory conditions in triplicate. Accessory shakers were used to homogenize the samples, with manual agitation performed between measurements.

For chemical analysis, CNK samples from 50 accessions were ground and homogenized at the Food Analysis Laboratory of Embrapa Agroindústria Tropical. The moisture content was determined using the oven-drying method at 105°C until a constant mass was achieved, following standard procedures (Zenebon et al., 2008). Ash content was determined by incineration in a muffle furnace at 550°C until constant mass was reached. Crude protein content was quantified via total nitrogen using the Dumas combustion method (Latimer Jr., 2016) on an NDA 701 automatic nitrogen analyzer (Velp Scientifica, Milan, Italy), applying a conversion factor of 6.25 to calculate total protein concentration. Crude fat was estimated using the semi-automatic XT15 extractor (ANKOM, Macedon, NY, USA) with hexane as the solvent under high temperature and pressure, following method Am 5-04 (AOCS, 2005).

Based on the initial proximate composition, the four accessions with the highest lipid content, along with the commercial cultivar CCP 76 as a control, were selected for oil extraction and fatty acid characterization. Oil yield was determined by Soxhlet extraction using ethyl ether for 24 hours. Subsequently, fatty acids were esterified into fatty acid methyl esters (Zenebon et al., 2008) and analyzed by gas chromatography-mass spectrometry.

Univariate statistical analyses were performed using R software (R Core Team, 2013). Physicochemical analyses and oil yield data were subjected to analysis of variance (ANOVA). When significant differences were detected, means were compared using Scott-Knott's test, at a 5% significance level.

NIR spectra of CNK oil samples were acquired for all five accessions at room temperature (28°C). Measurements were performed in triplicate using suitable cuvettes and cuvette holders, with the average spectrum calculated from three individual readings taken at different regions of each sample. Spectral responses were recorded in absorbance mode using the MicroNIR 1700 spectrometer. Spectral data acquisition parameters were set with an integration time of 50 ms and an average of 100 scans. A Spectralon standard (Labsphere, North Sutton, NH, USA) was used to obtain the reference spectrum for absorbance calculations, while the dark signal was obtained by measuring the instrument's external environment.

The oils were homogenized, and the fatty acids present in the CNK oil were esterified according to the method described by Zenebon et al. (2008). Briefly, 100 mg of oil from each sample were weighed in three replicate, totaling 15 samples. Each sample was solubilized with 3 mL of hexane, followed by the addition of 4.0 mL of 0.50 mol L^{-1} NaOH in methanol, and heated at 60–70°C for 5 min. Subsequently, 5.0 mL of the esterification reagent, prepared from a mixture of 10.0g of ammonium chloride, 300.0 mL of methanol, and 15.0 mL of concentrated sulfuric acid, was added, and the mixture was heated to 60–70°C for another 5 min before adding 4 mL of a saturated NaCl solution. Next, 3 mL of hexane were added to collect the organic phase, while the aqueous phase was discarded. The solvent from the organic phase was evaporated at room temperature using a Speedvac (Thermo Fisher Scientific, Waltham, MA, USA). The methyl esters were solubilized in chromatographic-grade hexane prior to gas chromatography–mass spectrometry analysis.

Multivariate regression analyses by the Partial Least Squares (PLS) regression method were developed to obtain the correlation between NIR spectra as predictor variables, and the variability of four physicochemical parameters of cashew nuts from different accessions as response variables: lipids, moisture, proteins, and ash. In total, 150 NIR spectra from three replicates of 50 accessions were converted into American Standard Code for Information Interchange (ASCII) format, and the spectral region between 900 and 1700 nm was selected for the correlation analysis. Numerical matrices for predictor and response variables were imported using the PLS-Toolbox software version 8.6.2 (Eigenvector Research Inc., Manson, WA, USA) for statistical evaluation. For spectral pre-processing, the Multiplicative Scatter Correction (MSC) algorithm was applied to the predictor variables, which were subsequently autoscaled. The response variables were likewise autoscaled and the SIMPLS algorithm

was applied to decompose the matrices into scores, loadings, and figures of merit.

Cross-validation was performed using the Venetian Blinds method at a 95% confidence level. The optimal number of latent variables was selected based on key statistical parameters and figures of merit, including: the correlation coefficients obtained in calibration (R_{Cal}) and cross-validation (R_{CV}); the coefficients of determination in calibration (R^2_{Cal}) and cross-validation (R^2_{CV}); the root mean square error of calibration (RMSEC); the root mean square error of cross-validation (RMSECV); the ratio of performance to deviation (RPD), calculated as the ratio of the standard deviation from the reference dataset divided by the RMSECV; and the cross-validated bias (CV bias).

Results and Discussion

Significant variability was observed among the 50 accessions for all evaluated physicochemical traits (Table 1), indicating substantial genetic diversity within the germplasm bank. The first two principal components accounted for approximately 70% of the total variance (Figure 1). Analysis of the variable vectors showed a positive association between crude protein and ash content, indicating that accessions with higher protein levels also tended to possess higher mineral concentrations (Figure 1). Conversely, a negative correlation was observed between lipid content and the other major constituents, specifically crude protein and ash (Figures 1 and 2). This suggests that accessions with enhanced lipid accumulation typically exhibit lower protein and mineral contents. Other correlations were not statistically significant, reflecting the relative independence of moisture levels from the remaining nut constituents (Figure 2).

This inverse association between lipid and protein contents has been previously documented in other oilseed species (Guo et al., 2025). During seed development, metabolic processes depend heavily on the availability of photoassimilates, and consequently, plants may prioritize the fatty acid biosynthetic pathway over amino acid deposition (Cao et al., 2024).

Lipid content ranged from 38.17% in BGC 573 to 47.90% in BGC 332 (Table 1). Accessions BGC 332, BGC 305, BGC 543, and BGC 653 exhibited lipid contents exceeding 46%, highlighting their potential for industrial applications. These values are above the 30–35% range generally considered low for cashew kernels, confirming the oil-producing potential of the evaluated germplasm (Liu et al., 2023).

Ash content across the evaluated accessions ranged from 2.31 to 3.34%, with the minimum value recorded in accession BGC 427 and maximum value recorded in accessions BGC 385 and BGC 476 (Table 1). CNK is recognized as an important source of inorganic minerals (Semporé et al., 2021), and the observed variability highlights significant genetic diversity within the germplasm (Caetano et al., 2025). Higher ash content reflects an increased mineral concentration, directly enhancing the kernel's nutritional value (Ibrahim et al., 2024).

Protein content ranged from 13.66% in BGC 427 to 22.41% in BGC 385, aligning with the 13.7–22.1% interval previously reported for different genotypes (Caetano et al., 2025). This consistency across distinct studies suggests that protein accumulation is a stable trait across varying genetic backgrounds. These findings validate cashew kernels as a high biological value protein source, offering significant potential for plant-based protein formulations with complete amino acid profiles (Carmo et al., 2025).

Moisture content in the kernels ranged from 1.89% in BGC 097 to 6.29% in BGC 546, with 48 accessions exhibiting levels below 5%, the maximum threshold established by Brazilian legislation (Brasil, 2009). Maintaining low moisture is critical for the shelf life and quality of oilseeds, as lower levels are associated with sustained physiological vigor and performance. Conversely, higher moisture content tends to reduce germination rates and increase susceptibility to microbial spoilage (Shi et al., 2025).

The lack of a significant correlation between moisture and other chemical constituents is evidenced by the near-zero correlation coefficients. This result aligns with the principal components analysis. In the multivariate analysis, moisture was the primary contributor to principal component 2 and displayed an orientation nearly orthogonal to the lipid, protein, and ash vectors, indicating its independence from the centesimal composition (Figures 1 and 2). Moisture variability across accessions is likely driven by external factors, such as drying or storage conditions

Table 1. Comparison of physicochemical characteristics means evaluated in accessions of the cashew nut genebank[(1)].

Accessions	Moisture (%)	Lipids (%)	Ashes (%)	Proteins (%)
BGC 021	3.33e	45.04c	2.43d	18.63f
BGC 062	3.53e	44.61c	2.34d	15.98k
BGC 075	3.62e	44.12d	2.40d	14.93l
BGC 097	1.89g	44.01d	2.37d	18.32g
BGC 101	3.33e	44.37d	2.53c	16.00k
BGC 229	3.14e	41.70g	2.60c	20.81c
BGC 271	3.96d	44.38d	2.58c	18.26g
BGC 289	3.17e	42.44f	2.44d	19.30e
BGC 292	3.97d	40.38h	2.79b	20.77c
BGC 305	3.05f	47.18a	2.81b	19.02f
BGC 307	3.20e	39.00i	2.79b	19.69e
BGC 321	2.59f	41.22g	2.72c	19.39e
BGC 322	2.67f	43.22e	2.64c	18.70f
BGC 332	3.96d	47.90a	2.65c	19.19e
BGC 342	3.61e	44.79c	2.73c	20.12d
BGC 350	3.44e	38.96i	2.59c	19.73e
BGC 354	3.12e	43.90d	2.63c	19.57e
BGC 355	4.77c	44.30d	2.64c	17.76h
BGC 364	2.72f	44.21d	2.45d	18.60f
BGC 385	3.38e	39.97h	3.34a	22.41a
BGC 397	2.86f	40.10h	2.61c	19.97d
BGC 401	3.29e	40.62h	2.88b	17.79h
BGC 407	3.83d	38.76i	2.63c	18.65f
BGC 423	3.38e	40.87h	2.52c	18.81f
BGC 424	4.28d	40.73h	2.59c	15.81k
BGC 427	3.38e	40.98h	2.31d	13.66m
BGC 446	2.89f	42.03g	2.94b	22.10a
BGC 452	2.97f	41.37g	2.98b	19.29e
BGC 455	3.92d	42.56f	2.56c	16.09k
BGC 457	3.31e	40.52h	2.84b	16.70j
BGC 464	2.90f	46.02b	2.51c	17.02i
BGC 467	3.03f	42.74f	2.68c	20.62c
BGC 469	4.21d	45.07c	2.59c	17.56h
BGC 476	3.44e	42.95f	3.34a	20.36d
BGC 485	3.53e	43.21e	2.97b	18.34g
BGC 512	4.05d	43.69e	2.55c	17.09i
BGC 520	4.05d	43.46e	2.53c	18.58f
BGC 543	3.16e	46.67b	2.65c	18.90f
BGC 546	6.29a	42.58f	2.61c	19.39e
BGC 561	3.61e	43.27e	2.65c	15.96k
BGC 566	2.68f	44.77c	2.70c	18.86f
BGC 573	4.58c	38.17j	2.94b	14.06m
BGC 592	2.83f	37.68j	2.96b	20.38d
BGC 595	5.33b	41.28g	2.69c	18.99f
BGC 605	3.69e	43.26e	2.70c	20.16d
BGC 612	2.66f	45.08c	2.32d	17.40h
BGC 630	2.61f	38.99i	2.79b	19.59e
BGC 640	2.76f	40.74h	2.79b	21.31b
BGC 653	3.24e	46.46b	2.39d	18.36g
BGC 654	3.70e	43.88d	2.37d	19.44e
Average	3.46	42.68	2.66	18.57
Standard deviation	0.80	2.48	0.29	1.89
CV (%)	23.26	5.82	10.88	10.15

[(1)]Means followed by equal letters, in the columns, do not differ from each other by Scott-Knott's test ($p \geq 0.05$).

and kernel size, whereas lipids, proteins, and ash more closely reflect intrinsic genetic characteristics.

Significant variability in the oil yield was observed among accessions with the highest lipid content (BGC 332, BGC 305, BGC 543, and BGC 653). All four accessions presented oil yield exceeding 40%, which is considered exceptionally high for industrial applications. The cultivar CCP 76 and accession BGC 653 stood out with the highest yields, reaching 48.0 and 46.7%, respectively, whereas accessions BGC 305 and BGC 543 presented the lowest yields within this elite group at 43.3 and 42.7%, respectively (Figure 3).

Significant variations were observed in the concentrations of palmitoleic, palmitic, stearic, and arachidic acids, as well as in their corresponding saturated to unsaturated ratios (Table 2 and Figure 6). Notably, accession BGC 543 stood out for its significantly higher levels of stearic and arachidic acids compared to the other genotypes (Figures 6 F and H).

Analysis of total lipid balance revealed that BGC 653 possessed the most unsaturated profile at 76.66%, whereas BGC 543 was characterized as the most saturated at 27.64%. From the perspective of high-quality vegetable oil production, in which high unsaturation is a primary objective, accessions BGC 653 and BGC 305 emerged as the most promising genotypes. Furthermore, BGC 653 offers a distinct industrial advantage by combining this superior lipid quality with a higher oil extraction yield.

Chromatographic analysis identified nine fatty acids, with oleic acid content ranging from 52.0 to 55.9% and linoleic acid from 18.8 to 21.3% as the predominant unsaturated compounds, followed by 11 to 13% of stearic and palmitic acids (Table 2). Total unsaturated fatty acid content ranged from 72.4% to 76.7%, confirming the nutritional quality of cashew

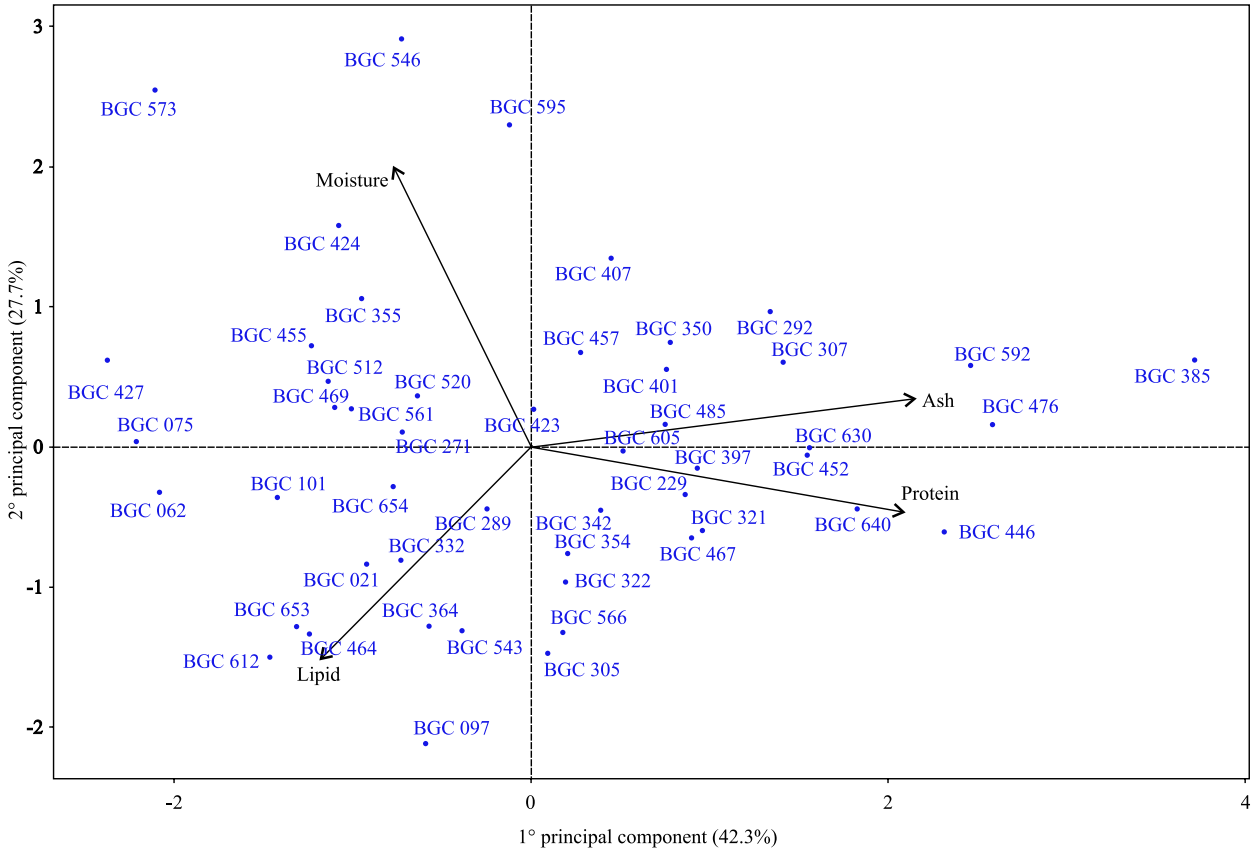

Figure 1. Principal component analysis of proteins, lipids, ash, and moisture in cashew (*Anacardium occidentale*) nuts kernels from different accessions.

kernel oil (Soro et al., 2024). Accessions BGC 653 and BGC 305 showed the most favorable lipid profiles, combining high proportions of unsaturated fatty acids with elevated oil yields. The predominance of unsaturated fatty acids, particularly oleic acid, highlights the nutritional value of the oil and suggests potential benefits associated with cardiovascular protection (Leal et al., 2023; Lu et al., 2024).

The multivariate regression model was developed to correlate NIR spectral data with variations in

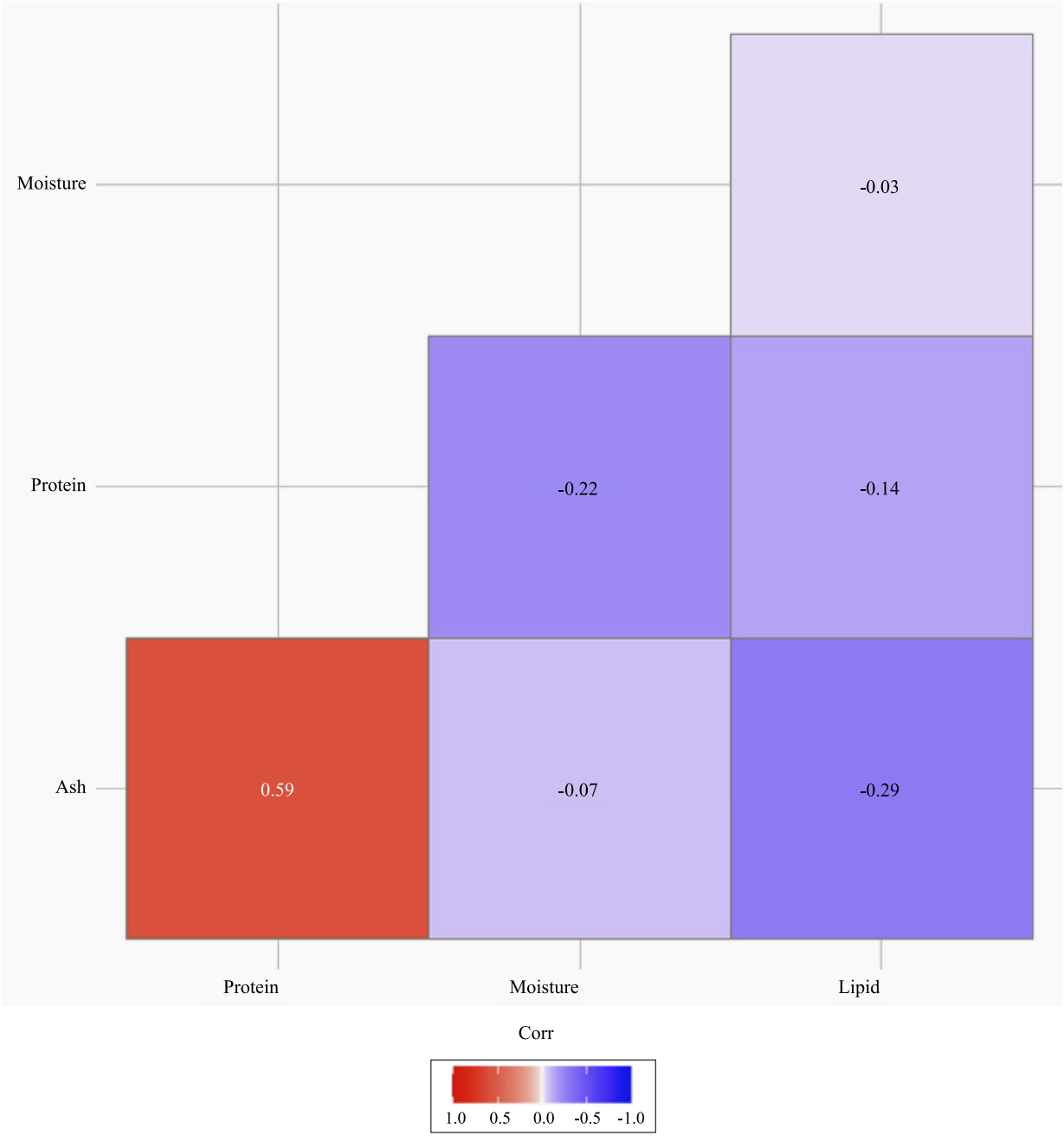

Figure 2. Phenotypic correlation between moisture, lipids, ash, and crude protein variables obtained from the analysis of 50 accessions from the cashew genebank. Pearson correlation coefficients are indicated by colors, with positive values represented in warm tones and negative values in cool tones.

lipid, protein, moisture, and ash content across the 50 cashew accessions. These supervised analyses used PLS regression, which was conducted independently for each physicochemical parameter.

Regression residuals in NIR-based modeling are frequently attributed to intrinsic spectral resolution limits, in which noise, overlapping peaks, and spectral artifacts can partially obscure chemical information (Fodor et al., 2024). Nevertheless, NIR spectroscopy remains a rapid and cost-effective analytical tool for characterizing untreated samples, specifically those containing molecular structures with C–H, N–H, S–H, and O–H bonds (Praiphui et al., 2023).

To mitigate physical light-scattering effects inherent to the technique, the spectra were adjusted to a uniform dispersion level, using the MSC algorithm (Fodor et al., 2024). This correction was calculated based on an average reference spectrum of the dataset, effectively enhancing the signal-to-noise ratio for subsequent modeling.

Preliminary analyses indicated that only lipid and moisture contents exhibited sufficient correlation with NIR spectra. Therefore, interval Partial Least Squares (iPLS) was applied to identify the most informative spectral regions and improve model performance. The selected intervals corresponded to those presenting the lowest RMSECV values, which were subsequently used to construct the final prediction models (Andersen & Bro, 2010; Bertone et al., 2016).

In contrast, the prediction models for proteins and ash showed limited predictive ability with R^2_{CV} <0.20, RMSECV > 1.5 and RPD < 1.2 values. This result is probably attributed to the weak spectral contribution of these constituents in the evaluated NIR range, combined with the use of whole kernels rather than homogenized samples. Caetano et al. (2025) conducted similar analyses using ground cashew kernels, producing a more homogeneous sample matrix than whole kernels. This difference in sample presentation may have contributed to the contrasting spectral responses obtained.

In the iPLS models developed for lipid and moisture contents, the green-shaded regions denote the selected spectral intervals exhibiting the lowest RMSECV values (Figures 4A and 5A). Regression plots (Figures 4B and 5B) demonstrate agreement between predicted and reference values, whereas Hotelling T2 versus Q residual plots (Figures 4C and 5C) were used to evaluate sample influence and model residuals. The calibration

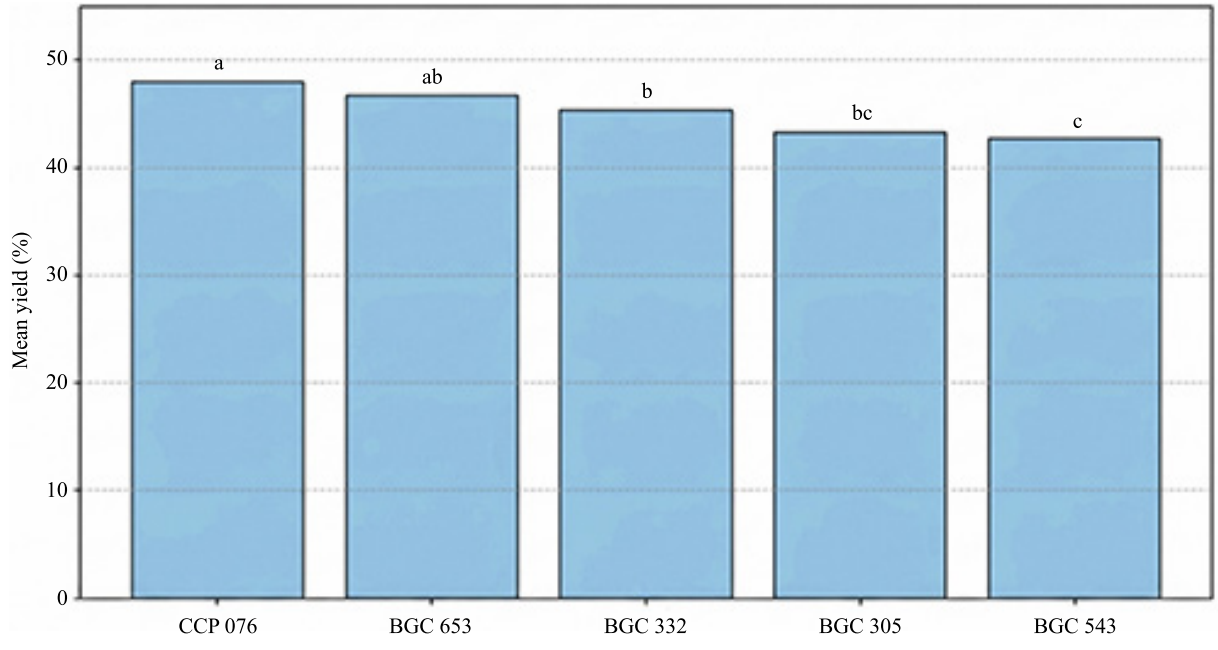

Figure 3. Percentage yield of oil extracted from cashew (*Anacardium occidentale*) nut kernels from the commercial cultivar CCP 76 and accessions BGC 653, BGC 322, BGC 305, BGC 543.

Table 2. Fatty acid methyl esters content in cashew (*Anacardium occidentale*) nut kernel oil[(1)].

Compounds	BGC 305	BGC 332	BGC 543	BGC 653	CCP 076
Palmitoleic acid (%)	0.35a	0.16bc	0.15c	0.16bc	0.26ab
Palmitic acid (%)	13.05a	11.74ab	10.99b	11.35b	12.24ab
Margaric acid (%)	0.08a	0.07a	0.07a	0.05a	0.08a
Linoleic acid (%)	19.81a	18.82a	20.09a	21.28a	20.99a
Oleic acid (%)	55.24a	55.95a	52.01a	55.13a	52.92a
Stearic acid (%)	10.94d	12.76bc	15.83a	11.53cd	12.94b
cis-11-Eicosenoic acid (%)	0.11a	0.08a	0.11a	0.09a	0.10a
Arachidic acid (%)	0.42b	0.42b	0.72a	0.42b	0.47b
Behenic acid(%)	0.00a	0.00a	0.02a	0.00a	0.00a
Saturated/unsaturated ratio	0.32ab	0.33ab	0.38a	0.31b	0.35ab

[(1)]Means followed by equal letters, in the columns, do not differ from each other by Tukey's test ($p \geq 0.05$).

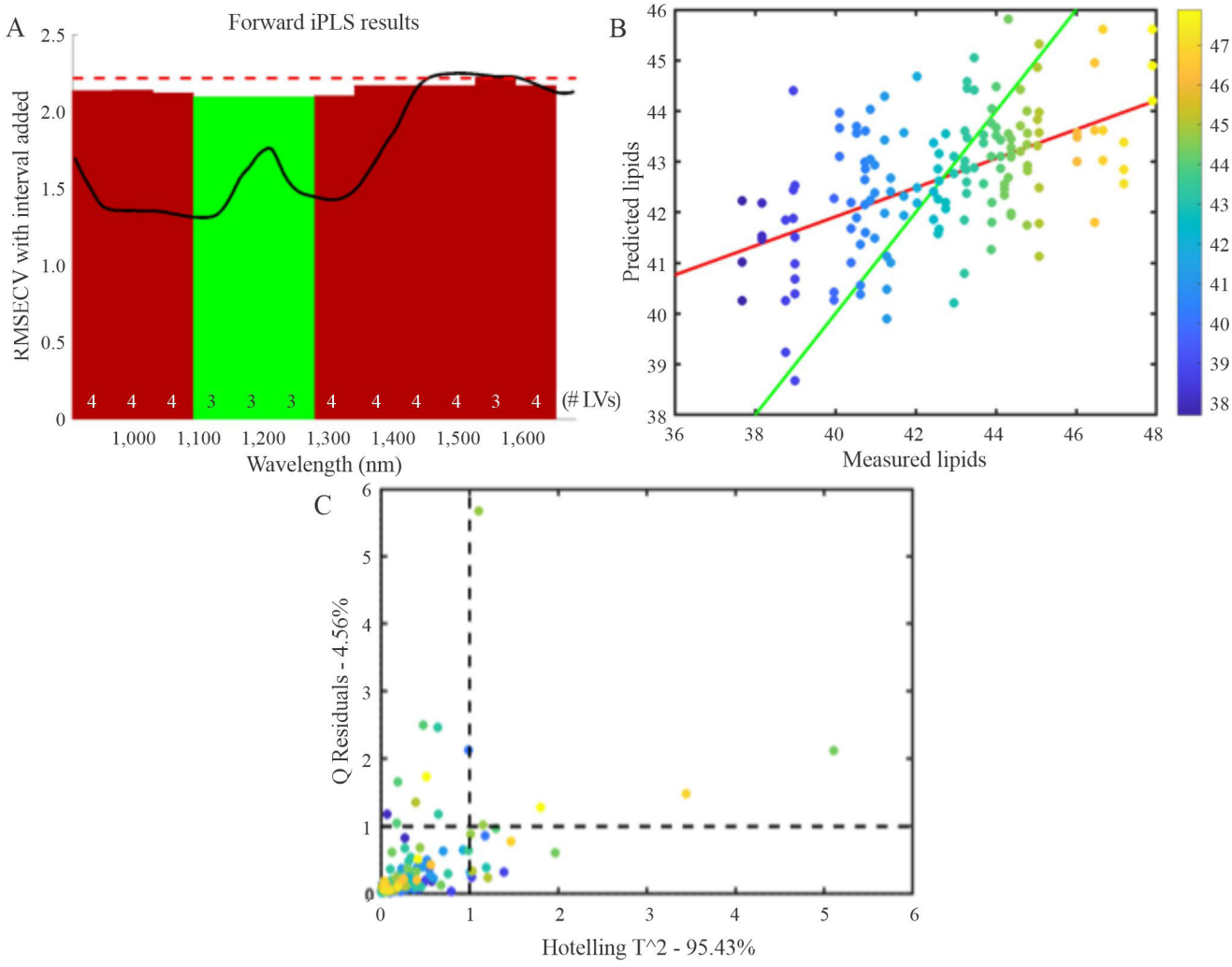

Figure 4. (A) Average of the NIR spectra of cashew (*Anacardium occidentale*) nuts, with the spectral area selected by iPLS in green; (B) results of the multivariate regression for predicting lipid content; and (C) Hotelling T 2 × Q residuals from the regression modeling.

and cross-validation statistics indicate comparable performance for both models, characterized by R2 values near 0.60 and low differences between RMSEC and RMSECV, suggesting stable models with moderate predictive capacity (Table 3).

The iPLS models for lipids and moisture demonstrated potential for preliminary screening, yielding RPD values of 1.18 and 1.57, respectively. Although a RPD < 1.5 indicates that the lipid model is restricted to discriminating between high and low values, the performance of the moisture model (R^2_{CV} = 0.53; RPD = 1.57) is considered satisfactory for initial germplasm characterization.

The minimal discrepancies between RMSEC and RMSECV, reflected in RMSEC/RMSECV ratios of 0.96 and 0.94, indicate stable models with minimal overfitting. Although a few samples were identified as potential outliers in the Hotelling T^2 versus Q-residual plots, they did not substantially affect model performance. Overall, the models explained more than 50% of the observed variance, demonstrating the potential of NIR spectroscopy as a rapid and non-destructive tool for the preliminary screening of lipid and moisture contents in cashew germplasm accessions (Oliveira et al., 2014; Freitas et al., 2018).

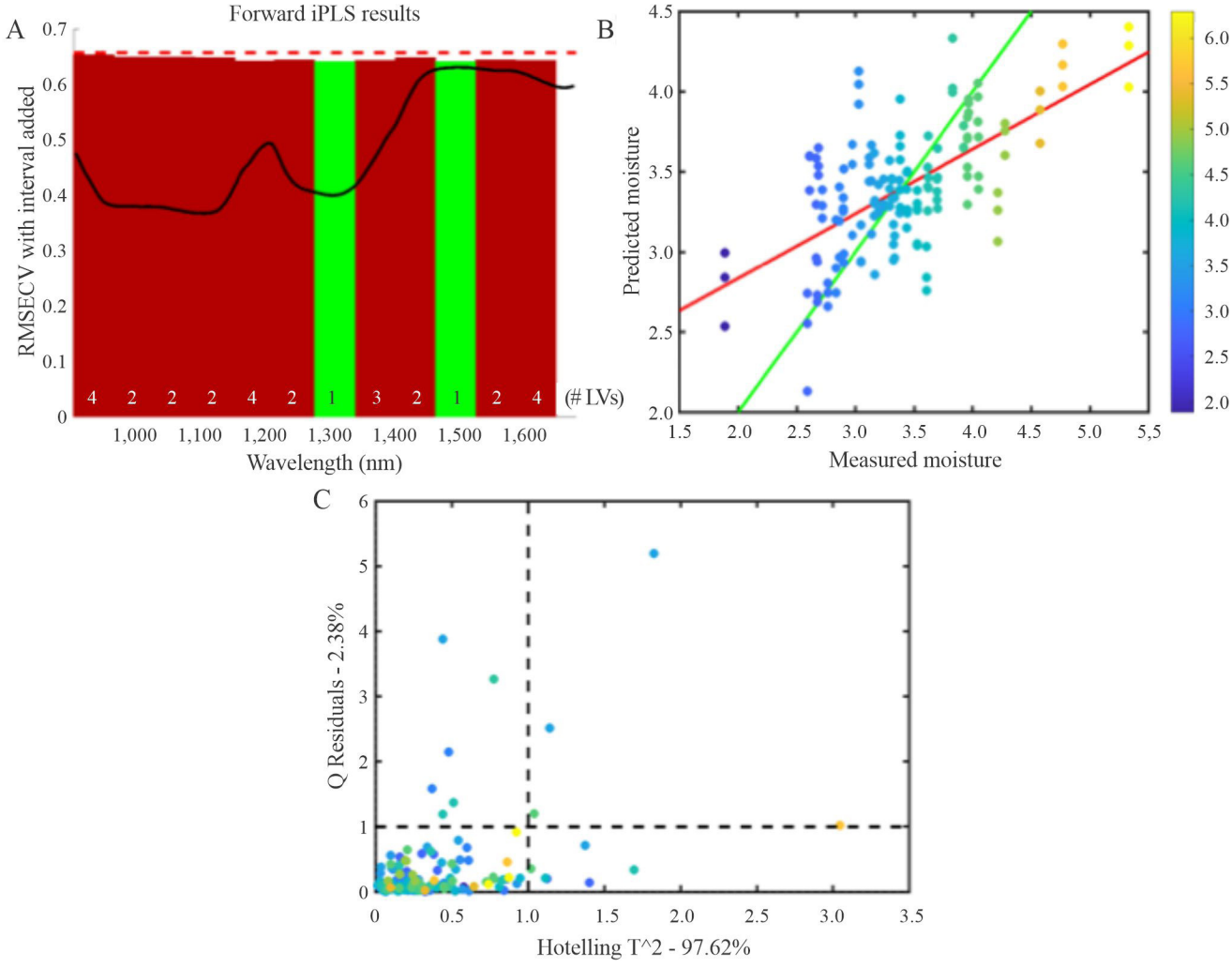

Figure 5. (A) Average of the NIR spectra of cashew (*Anacardium occidentale*) nut kernels, with the spectral area selected by iPLS in green; (B) results of the multivariate regression for predicting moisture content; (C) Hotelling $T^2 \times Q$ residuals from the regression modeling.

Table 3. Statistical parameters of the multivariate regression analysis by iPLS modeling using four latent variables: NIR as predictor variables, and lipids and moisture concentrations as response variables.

Variables[(1)]	Lipids	Moisture
Variance captured (%)	95.44	97.62
R^2_{Cal}	0.59	0.60
RMSEC	2.03	0.48
R^2_{CV}	0.54	0.53
RMSECV	2.11	0.51
RPD	1.18	1.57
RMSEC/RMSECV	0.96	0.94
CV Bias	-0.01	-0.01

[(1)] $R^2_{Cal,}$ coefficient of determination in calibration; RMSEC, root mean square error of calibration; $R^2_{CV,}$ coefficient of determination in cross-validation; RMSECV, root mean square error of cross-validation; RPD (SD/RMSECV), ratio of performance to deviation were calculated using the standard deviation from the reference dataset divided by the RMSECV; RMSEC/RMSECV, similarity criterion (proximity between RMSEC and RMSECV); CV Bias, average difference between estimated and actual values during cross-calibration.

Cross-validation yielded coefficients of determination of 0.54 for lipids and 0.53 for moisture contents, indicating that the models explained slightly more than 50% of the variance in the respective datasets. These results suggest that NIR spectroscopy can be used to discriminate accessions with relatively high and low lipid or moisture contents, supporting its application as a rapid screening tool for whole nut kernel characterization. Although a few samples were identified as potential outliers, they did not substantially affect model performance, as evidenced by the validation statistics presented in Table 3.

In order to complement multivariate statistical analysis, a univariate statistical evaluation of accessions BGC 305, BGC 332, BGC 543, BGC 653 and cultivar CCP 076 was performed using one-way ANOVA. The concentration of each organic compound in mg mL^{-1} was considered separately in order to statistically verify significant variation among the selected accessions (Figure 6).

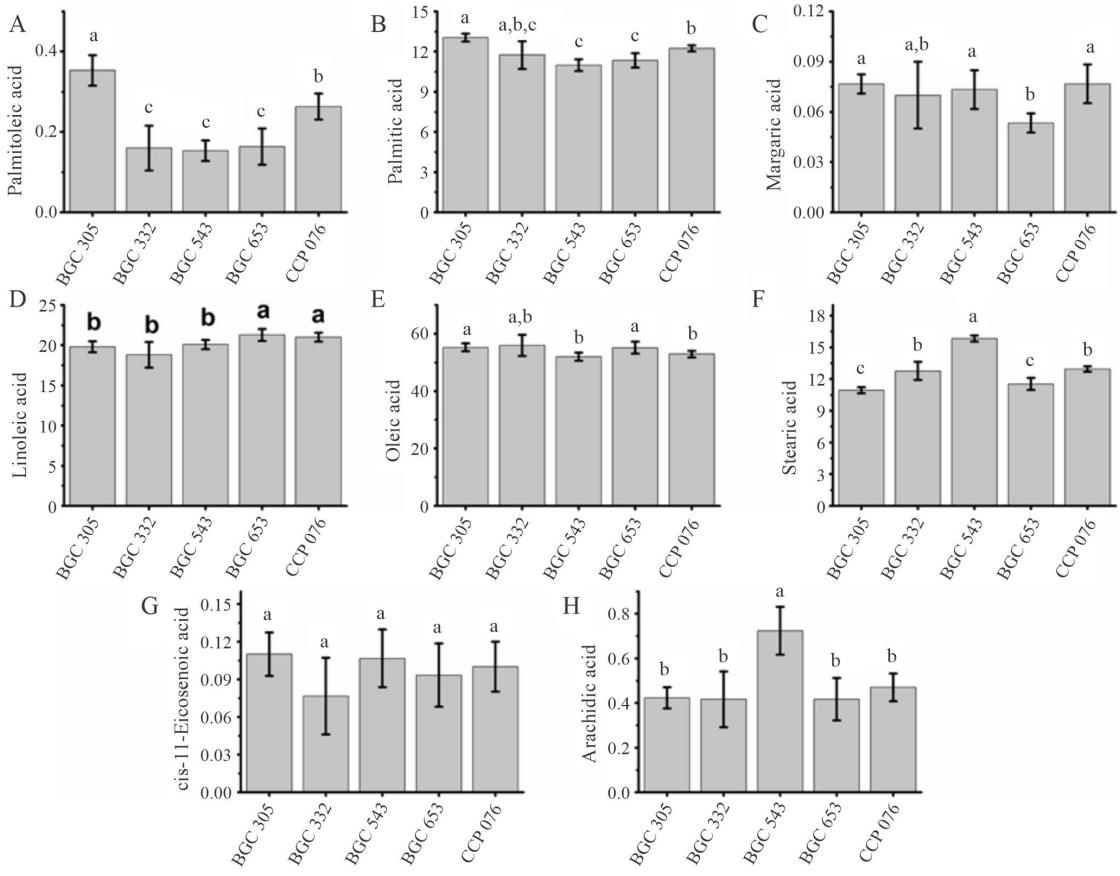

Figure 6. ANOVA of palmitoleic acid, palmitic acid, margaric acid, linoleic acid, oleic acid, stearic acid, cis-11-eicosenoic acid, and arachidic acid concentrations in oils from the BGC 305, BGC 332, BGC 543, BGC 653, and CCP 76 cashew (*Anacardium occidentale*) nut accessions.

Conclusions

1. There is a substantial physicochemical variability among the 50 accessions of the Cashew (*Anacardium occidentale*) genebank, with lipid contents ranging from 38.17 to 47.90%.

2. Accessions BGC 332, BGC 305, BGC 543, and BGC 653 represent superior genotypes characterized by a nutritionally favorable oil composition and high lipid content; among these, accession BGC 653 stands out as it combines an optimal balance between extraction yield and oxidative stability.

3. Cashew nuts prove to be an excellent source of unsaturated fatty acids, predominantly oleic acid, featuring a high biological value protein matrix that meets expected oxidative stability.

4. Portable NIR technology is a viable, fast, and non-destructive tool for predicting lipids and moisture content directly in whole nuts.

5. Although the homogeneity of the whole nut limited the accuracy for protein and ash, the satisfactory performance of the iPLS model for lipids validates the use of the portable NIR sensor as an efficient screening method for genotype characterization, quality monitoring, and early selection of superior genetic materials.

References

ANDERSEN, C.M.; BRO, R. Variable selection in regression - a tutorial. **Journal of Chemometrics**, v.24, p.728-737, 2010. DOI: https://doi.org/10.1002/cem.1360.

AOCS. American Oil Chemists' Society. **Official methods and recommended practices of the American Oil Chemists' Society**. 6th ed. Urbana, 2005. Official Method Am 5-04: rapid determination of oil/fat using high temperature solvent extraction.

BERTONE, E.; VENTURELLO, A.; GIRAUDO, A.; PELLEGRINO, B.; GEOBALDO, F. Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee. **Food Control**, v.59, p.683-689, 2016. DOI: https://doi.org/10.1016/j.foodcont.2015.06.055.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 62, de 15 de dezembro de 2009. [Regulamento Técnico da Amêndoa de Castanha de Caju]. **Diário Oficial da União**, 16 dez. 2009. Seção1, p.29-31.

CAETANO, I.D.E.; ALVES FILHO, E.G.; SILVA, E. de O.; DIONISIO, A.P.; VASCONCELOS SEGUNDO, V.C.; CASTRO, A.C.R. de. Nutritional composition variability of cashew nut kernels from different accessions in a Brazilian genebank. **ACS Food Science & Technology**, v.5, p.3737-3746, 2025. DOI: https://doi.org/10.1021/acsfoodscitech.5c00336.

CAO, D.; MA, Y.; CAO, Z.; HU, S.; LI, Z.; LI, Y.; WANG, K.; WANG, X.; WANG, J.; ZHAO, K.; ZHAO, K.; QIU, D.; LI, Z.; REN, R.; MA, X.; ZHANG, X.; GONG, F.; JUNG, M.Y.; YIN, D. Coordinated lipid mobilization during seed development and germination in peanut (*Arachis hypogaea* L.). **Journal of Agricultural and Food Chemistry**, v.72, p.3218-3230, 2024. DOI: https://doi.org/10.1021/acs.jafc.3c06697.

CARMO, J.S. do; VIANA, J.D. da R.; SOUZA, A.C.R. de; MIRANDA, K.W.E.; MUNIZ, C.R.; SILVA, L.M.A. e; CARVALHO, M.D. de; SATO, A.C.K.; SÁ, D.M.A.T.; OLIVEIRA, W. da S.; DIAS, F.F.G.; DIONÍSIO, A.P. Cashew nut protein concentrate as a potential ingredient for the emerging alternative protein industry. **Food Chemistry**, v.96, art.146855, 2025. DOI: https://doi.org/10.1016/j.foodchem.2025.146855.

CHEN, Y.Y.; LI, N.Y.; GUO, X.; HUANG, H.J.; GARCIA-OLIVEIRA, P.; SUN, J.Y; ZHANG, J.D.; PRIETO, M.A.; GUO, Z.Q.; LIU, C. The nutritional and bio-active constituents, functional activities, and industrial applications of cashew (*Anacardium occidentale*): a review. **Food Frontiers**, v.4, p.1606-1621, 2023. DOI: https://doi.org/10.1002/fft2.250.

FODOR, M.; MATKOVITS, A.; BENES, E.L.; JÓKAI, Z. The role of near-infrared spectroscopy in food quality assurance: a review of the past two decades. **Foods**, v.13, art.3501, 2024. DOI: https://doi.org/10.3390/foods13213501.

FREITAS, J.V.B.; ALVES FILHO, E.G.; SILVA, L.M.A.; ZOCOLO, G.J.; BRITO, E.S. de; GRAMOSA, N.V. Chemometric analysis of NMR and GC datasets for chemotype characterization of essential oils from different species of *Ocimum*. **Talanta**, v.180, p.329-336, 2018. DOI: https://doi.org/10.1016/j.talanta.2017.12.053.

GRÜTER, R.; TRACHSEL, T.; LAUBE, P.; JAISLI, I. Expected global suitability of coffee, cashew and avocado due to climate change. **PLoS ONE**, v.17, e0261976, 2022. DOI: https://doi.org/10.1371/journal.pone.0261976.

GUO, H.; LI, J.; LIU, Y.; FERNÁNDEZ-PASCUAL, E. Lipid metabolism during seed germination of *Pistacia chinensis* and its response to gibberellic acid. **Plant Physiology and Biochemistry**, v.219, art.109371, 2025. DOI: https://doi.org/10.1016/j.plaphy.2024.109371.

HIRPARA, S.R.; VAHONIYA, D.R.; BAGDA, B.; RAJWADI, A. Cashew nut industry in India: production, processing, and marketing. **International Journal of Agriculture and Food Science**, v.7, p.326-333, 2025. DOI: https://doi.org/10.33545/2664844X.2025.v7.i9d.763.

IBRAHIM, M.; AKHTAR, N.; KHAN, A.; SARA; ANWAR, Y.; WONG, L.S.; SUBRAMANIYAN, V.; KUMARASAMY, V.; RAUF, A.; AHLAWAT, Y.K.; SHARMA, N.; LOPES, L. de A. Nutritional profiling of wild edible plants: quantification of macronutrients and minerals via energy dispersive x-ray (EDX) fluorescence spectroscopy. **Brazilian Journal of Biology**, v.84, e287349, 2024.

LATIMER JR., G.W. (Ed.). **Official Methods of Analysis of AOAC International**. 20th ed. Rockville: AOAC International, 2016. Official Method 992.23: Crude protein in cereal grains and oilseeds.

LEAL, A.R.; DIONÍSIO, A.P.; ABREU, F.A.P. de; OLIVEIRA, G.F. de; ARAÚJO, I.M. da S.; MAGALHÃES, H.C.R.; LEITE, A.B.; SILVA, E.K.M. da; NASCIMENTO, R.F. do; NASCIMENTO, H.O. do; SOUSA, P.H.M. de. Impact of different kernel grades on volatile compounds profile, fatty acids, and oxidative quality of cashew nut oil. **Food Research International**, v.165, art.112526, 2023. DOI: https://doi.org/10.1016/j.foodres.2023.112526.

LIU, Y.; LI, L.; XIA, Q.; LIN, L. Analysis of physicochemical properties, lipid composition, and oxidative stability of cashew nut kernel oil. **Foods**, v.12, art.693, 2023. DOI: https://doi.org/10.3390/foods12040693.

LU, Y.; ZHAO, J.; XIN, Q.; YUAN, R.; MIAO, Y.; YANG, M.; MO, H.; CHEN, K.; CONG, W. Protective effects of oleic acid and polyphenols in extra virgin olive oil on cardiovascular diseases. **Food Science and Human Wellness**, v.13, p.529-540, 2024. DOI: https://doi.org/10.26599/FSHW.2022.9250047.

OLIVEIRA, C.R. de; CARNEIRO, R.L.; FERREIRA, A.G. Tracking the degradation of fresh orange juice and discrimination of orange varieties: an example of NMR in coordination with chemometrics analyses. **Food Chemistry**, v.164, p.446-453, 2014. DOI: https://doi.org/10.1016/j.foodchem.2014.05.026.

PRAIPHUI, A.; LOPIN, K.V.; KIELAR, F. Construction and evaluation of a low cost NIR-spectrometer for the determination of mango quality parameters. **Journal of Food Measurement and Characterization**, v.17, p.4125-4139, 2023. DOI: https://doi.org/10.1007/s11694-023-01948-y.

R CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013.

SEMPORÉ, J.N.; SONGRÉ-OUATTARA, L.T.; TARPAGA, W.V.; BATIONO, F.; DICKO, M.H. Comparison of proximate composition and nutritional qualities of fifty-three cashew accessions from Burkina Faso. **Food and Nutrition Sciences**, v.12, p.1191-1203, 2021. DOI: https://doi.org/10.4236/fns.2021.1212087.

SHI, P.; SUN, H.; WANG, Y.; HAN, N.; REN, L.; LV, J.; GUO, Q.; HE, K.; WANG, H.; YU, G. The impact of storage conditions on peanut seed quality, growth, and yield. **Plants**, v.14, art.2944, 2025. DOI: https://doi.org/10.3390/plants14192944.

SORO, K.F.D.; BOUKACHABINE, K.; EL-ANTARI, A. Comparative study of morphological aspects and oil composition of cashew nuts from nine locaties in Ivory Coast. **Food Nutrition Chemistry**, v.2, art.257, 2024. DOI: https://doi.org/10.18686/fnc257.

ZENEBON, O.; PASCUET, N.S.; TIGLEA, P. (Coord.). **Métodos físico-químicos para análise de alimentos**. 4.ed., 1.ed. digital. São Paulo: Instituto Adolfo Lutz, 2008. 1020p. Versão eletrônica.

Author contributions

Febriana Correia Ximenes de Carvalho: data curation (equal), writing – original draft (equal); **Elenilson Godoy Alves Filho**: data curation (equal), formal analysis (equal); writing – original draft (equal); **Ebenezer de Oliveira Silva**: writing – review & editing (equal); **Fernando Antonio Souza de Aragão**: data curation (equal); **Men de Sa Moreira de Souza Filho**: project administration (equal), methodology (equal), writing – original draft (equal); **Ana Cecília Ribeiro de Castro**: project administration (equal); conceptualization (equal), resources (equal), data curation (equal), writing – original draft (equal).

Chief editor: Edemar Corazza
Edited by: Madalena Rinaldi

Data availability statement

Data available upon request: research data are only available upon reasonable request to the corresponding author.

Declaration of use of AI technologies

During the preparation of this work, the authors used Notebook LM/Gemini in order to achieve linguistic refinement and grammatical correction in English. After this use, the authors reviewed and edited the content as needed and takes full responsibility for it.

Conflict of interest statement

The authors declare no conflicts of interest.

Acknowledgment

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for the financial support to the first author.

Disclaimer/Publisher's note

The statements, opinions, and data contained in all texts published in Pesquisa Agropecuária Brasileira (PAB) are solely those of the individual author(s) and not of the journal's publisher, editor, and editorial team, who disclaim responsibility for any injury to people or property resulting from any referred ideas, methods, instructions, or products.

The mention of specific chemical products, machines, and commercial equipment in the texts published in this journal does not imply their recommendation by the publisher.