Oil Crop Science 9 (2024) 132–141

Contents lists available at ScienceDirect

Oil Crop Science

journal homepage: www.keaipublishing.com/en/journals/oil-crop-science

Application of near-infrared spectroscopy for fast germplasm analysis and classification in multi-environment using intact-seed peanut (*Arachis hypogaea* L.)

Fentanesh Chekole Kassie [a,b], Gilles Chaix [c,d,e], Hermine Bille Ngalle [a], Maguette Seye [f,g], Coura Fall [f,g], Hodo-Abalo Tossim [f,g], Aissatou Sambou [f,g], Olivier Gibert [c,d], Fabrice Davrieux [h], Joseph Martin Bell [a], Jean-François Rami [c,d,g], Daniel Fonceka [c,d,f,g], Joël Romaric Nguepjop [c,d,f,g,*]

[a] Department of Plant Biology and Physiology, Faculty of Sciences, University of Yaounde I, Yaounde, P.O. Box 337, Cameroon
[b] Department of Plant Science, College of Agriculture, Wolaita Sodo University, Sodo, P.O. Box 138, Ethiopia
[c] CIRAD, UMR AGAP Institut, Montpellier, France
[d] AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France
[e] ChemHouse Research Group, Montpellier, France
[f] Centre D'Etudes Régional Pour L'Amélioration de L'Adaptation à La Sécheresse (CERAAS/ISRA), Route de Khombole, Thiès, BP 3320, Senegal
[g] Dispositif de Recherche et de Formation en Partenariat, Innovation et Amélioration Variétale en Afrique de L'Ouest (IAVAO), CERAAS, Route de Khombole, Thiès, BP 3320, Senegal
[h] CIRAD, UMR QUALISUD, Saint Pierre, France

ARTICLE INFO

Keywords:
Groundnut
Oilseed
Near infrared spectroscopy
Germplasm analysis
Environment
Nutritional
Breeding

ABSTRACT

Peanut is a worldwide oilseed crop and the need to assess germplasm in a non-destructive manner is important for seed nutritional breeding. In this study, Near Infrared Spectroscopy (NIRS) was applied to rapidly assess germplasm variability from whole seed of 699 samples, field-collected and assembled in four genetic and environment-based sets: one set of 300 varieties of a core-collection and three sets of 133 genotypes of an interspecific population, evaluated in three environments in a large spatial scale of two countries, Mbalmayo and Bafia in Cameroon and Nioro in Senegal, under rainfed conditions. NIR elemental spectra were gathered on six subsets of seeds of each sample, after three rotation scans, with a spectral resolution of 16 cm^{-1} over the spectral range of 867 nm to 2530 nm. Spectra were then processed by principal component analysis (PCA) coupled with Partial least squares-discriminant analysis (PLS-DA). As results, a huge variability was found between varieties and genotypes for all NIR wavelength within and between environments. The magnitude of genetic variation was particularly observed at 11 relevant wavelengths such as 1723 nm, usually related to oil content and fatty acid composition. PCA yielded the most chemical attributes in three significant PCs (i.e., eigenvalues >10), which together captured 93% of the total variation, revealing genetic and environment structure of varieties and genotypes into four clusters, corresponding to the four samples sets. The pattern of genetic variability of the interspecific population covers, remarkably half of spectrum of the core-collection, turning out to be the largest. Interestingly, a PLS-DA model was developed and a strong accuracy of 99.6% was achieved for the four sets, aiming to classify each seed sample according to environment origin. The confusion matrix achieved for the two sets of Bafia and Nioro showed 100% of instances classified correctly with 100% at both sensitivity and specificity, confirming that their seed quality was different from each other and all other samples. Overall, NIRS chemometrics is useful to assess and distinguish seeds from different environments and highlights the value of the interspecific population and core-collection, as a source of nutritional diversity, to support the breeding efforts.

* Corresponding author. CIRAD, UMR AGAP Institut, Montpellier, France.
E-mail address: joel.romaric.nguepjop@cirad.fr (J.R. Nguepjop).

https://doi.org/10.1016/j.ocsci.2024.03.003
Received 18 January 2024; Received in revised form 7 March 2024; Accepted 12 March 2024
Available online 15 May 2024
2096-2428/© 2024 the Oil Crop Research Institute, Chinese Academy of Agriculture Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Peanut is an annual oilseed crop cultivated globally on 36.18 million hectares of area in the world, yielding 71.68 million metric tons of pods in 2020 (FAOSTAT, 2020). As a functional food, peanut seed contains 34% – 56% oil, 22% – 30% protein, 10% – 25% carbohydrates, and 0.05% – 1% of various secondary metabolites, beneficial to human health, such as vitamin E, K and B complex, folic acid, niacin, and minerals (Ca, P, Mg, Zn, and Fe) (Desmae et al., 2019; Harch et al., 1995; Janila et al., 2013; Parilli-Moser et al., 2022). The main production constraints of the crop include drought, pests, diseases, and environmental changes. The oil content of seeds, shelf life, aroma, flavor and cooking quality are all affected by these constraints. Consequently, seed quality traits are targets in genetic breeding (Nawade et al., 2018; Parmar et al., 2022; Tang et al., 2022).

Peanut (*Arachis hypogaea* L.) is an autogamous species, allotetraploid (AABB genome; 2n = 4x = 40) with narrow genetic base (Burow et al., 2009; Simpson et al., 2001). The narrow genetic diversity coupled with low utilization of genetic resources are the major factors limiting peanut breeding. Thus, interspecific hybridization is currently used as a realistic strategy for introgressing prospective diversity from wild species into the cultivated gene pool (Favero et al., 2006; Fonceka et al., 2012a, 2012b; Mallikarjuna et al., 2011b; Tossim et al., 2020). Likewise, genetic diversity assessment and the detection of promising quality-related genotypes are fundamental to germplasm utilization and management in breeding strategies to support food security. To facilitate the investigation of large germplasm, it is reasonable to begin by examining subsets of germplasm that embody appropriate diversity and of manageable size, such as core collections or interspecific populations derived from wild × elite crosses, using appropriate characterization procedures.

Although significant efforts have been devoted to characterizing germplasm for simple traits and for the most important agronomic traits (yield and resistance to pests and diseases) (Fan et al., 2020; Kumari et al., 2014; Mallikarjuna et al., 2011a; Upadhyaya, 2005; Upadhyaya et al., 2011), less is known about quality traits across environments (Grosso et al., 2000; Wang et al., 2023). This is mainly due to the fact that these traits are quantitative and multigenic, with low heritability, and strong genotype environment interactions (Grosso et al., 1994; Isleib et al., 2008). Moreover, the phenotyping of these traits, regularly base on chemical survey, is expensive in terms of both direct monetary input and human labor, time-consuming, complex, and irreversibly destructive. Another main factor limiting chemical studies are the difficulties to assess many samples, each requiring many seeds (Davis et al., 2021; Nawade et al., 2018). Efforts to improve the knowledges of seed attributes might be supported by rapid and non-destructive tools. These include modified refractive index, capacitance sensor (Kandala et al., 2008), hyperspectral imaging (Huang et al., 2014; Rabanera et al., 2021) and near infrared (NIR) spectroscopy (Davis et al., 2021; Govindarajan et al., 2009; Tao et al., 2019; Wang et al., 2022). Among these, NIR-based methods are rapid, make it possible to analyze large number of samples. Moreover, some scholars have already applied machine learning as promising statistical methods to assist humans in the modeling and analysis of complex spectral data (Fordellone et al., 2019; Song et al., 2018) in many research fields including seed quality detection, genotyping of cultivars (Panero et al., 2022), varieties identification (Panero et al., 2018, 2022; Wang and Song, 2023; Xu et al., 2023) and classification (Sampaio et al., 2021; Singh et al., 2023; Tian et al., 2023). Some works previously described the feasibility of near infrared spectrometers to achieve some quick prediction of various peanut chemical compounds (proximal components and secondary metabolites) (Bilal et al., 2020; Li et al., 2019; Liu et al., 2022; Yu et al., 2020). In this paper, we focused on the non-destructive approach by NIR spectroscopy to investigate the environment and the genetic contribution of germplasm variability from intact-peanut-seed spectra without chemical references.

In this study, NIR spectroscopy was applied and coupled with chemometrics to assess germplasm variability from peanut intact-seed of a core-collection and of an interspecific population field-evaluated in three different environments. The objectives were specifically to i) perform a rapid NIR measurement on seeds and check the quality of spectra data, ii) assess genetic variation of varieties and genotypes from seed spectra, iii) study the pattern of genetic variability of the interspecific population in comparison to the core-collection, iv) potentially discriminate genetically related interspecific genotypes within and between environments by developing a classification model using PLS-DA.

2. Materials and methods

2.1. Genetic materials

Two distinct genetic materials were used in this study: an interspecific advanced backcross QTL (AB-QTL) population of 133 genotypes and a core collection of 300 cultivars. The AB-QTL population of 133 BC_2F_4 derivatives was developed from an interspecific cross, using Fleur11 as recurrent cultivated parent and the wild synthetic tetraploid 'ISATGR 278-18' (Nguepjop et al., 2016). The cultivated parent, Fleur 11, is an elite Spanish-type variety, widely cultivated in West Africa. The wild parent, ISATGR 278-18 is derived from a cross between *A. batizocoi* ICG 13160 (GKBSPSc 30,082, PI 468328) and *A. duranensis* ICG 8138 (GKP 10038, PI 262133) (Mallikarjuna et al., 2011b). The core collection of 300 cultivars was defined based on the knowledge of breeders and on diversity data from a collection of 1050 accessions (breeding lines and landraces) held by 10 breeding programs in East, Southern and West Africa (Conde et al., 2023). The detailed information of the 300 varieties of the core-collection and the 133 genotypes of the population are presented in the Supplementary Tables 1 and 2, respectively.

2.2. Trials environment and field experimental design

Whole seed used were collected from field experiments. Experiments were conducted in 3 different locations in 2 countries, Mbalmayo and Bafia in Cameroon and Nioro in Senegal, under rainfall conditions in 2021. The 3 locations were chosen to meet environmental differences, based on different criteria, largely based on ecology (climate and vegetation) and tradition of peanut cultivation and crop rotation (Table 1). Bafia is one of the main areas of peanut production in Cameroon. It is located in tropical savanna and has yellow vertisol soil (Temga et al., 2021) and equatorial climate of the Sudano-Guinean type with an average temperature of 25.1 °C and annual rainfall of 1500 mm. Mbalmayo is located in the tropical forest of Cameroon and has ocher vertisol soil (Temga et al., 2021) with a bimodal humid-forest rainfall climate with an average temperature of 26.5 °C and rainfall of 2402 mm. Nioro is located in the South of the Senegalese peanut basin and have Sahelian

Table 1
Characteristics of the field environments.

	Trial Environments		
	Bafia	Mbalmayo	Nioro
Country	Cameroon	Cameroon	Senegal
Location	Bafia	Mbalmayo	Nioro
Peanut cultivation	++++	+++	+++++
Ecology type	Tropical savanna	Tropical forest	Sahelian
Climate type	Sudano-guinean equatorial	Humid-forest bimodal rainfall	Sahelian semi-humid
Temperature (°C)	25.1	26.5	30.0
Rainfall (mm)	1500	2403	758
Soil type	Yellow vertisol	Ocher vertisol	Deck-Dior (leached ferruginous)
Previous crop	Maize	Maize	Millet
Experiment period	April–July	April–July	July–October

semi-humid ecology with a Deck Dior soil, a leached ferruginous tropical soil (Bogie et al., 2018), and annual rainfall of 758 mm and average temperature of 30 °C. The fields at Bafia and Mbalmayo were one-year fallow land after maize cultivation by farmers and were cleared and plowed for the study. The previous crop at Nioro was millet. The experiments in Bafia and Mbalmayo were conducted during one of the two rainy seasons from April–July, while the Nioro experiment was done during the rainfall season between July and October, at the Research Station of National Agricultural Research Center.

The same experimental design with common agricultural practices, from sowing to harvest, were used in each of the 3 environments. Within each environment, an alpha-lattice design was used with 3 replications, with 10 elementary plots within blocks. A plot consisted of rows of 3 m long on which 10 plants of the same genotype were sown with a spacing of 30 cm between plants on the same row, and 50 cm between two adjacent rows. The seeds were treated with Granox (captafol 10%, benomyl 10%, and carbofuran 20%) before planting to protect them against parasitic attacks and one seed per hill was sown manually at 4 cm depth. According to usual cultural practice, one hundred and fifty kg/ha of mineral fertilizer (6-20-10) were added 20 days after sowing. Throughout the vegetative development, weeds were manually harvested. The harvest was done at 95 days after sowing, followed by free-air drying for one month. At the end of the pod-drying stage, pods of each plant were separated from haulms, stored and dehulled.

2.3. Whole seed sample preparation

Whole seeds from pods of the three agronomic replicates of each genotype were bulked into one specific sample, stored in plastic bag, and labelled according to their respective name and environment. Thus, seeds of each sample for NIR analysis came from pods of 25–30 harvested plants of each genotype. From the expected 699 samples, we discarded 21 who had less than 100 seeds, 3 from Bafia, 9 from Mbalmayo and 9 from the core-collection. Finally, a total of 680 samples of seed were assembled in four genetic- and environment-based sets: one set of the 291 samples from cultivars from Nioro and three sets of the interspecific genotypes (130 samples for Bafia, 124 samples for Mbalmayo, 135 samples for Nioro, including the 133 genotypes and the CS16 variety and the cultivated parent Fleur11, both commonly used as check varieties in Nioro). All sample sealed in hermetic plastic bags were conveyed to the laboratory and kept at ambient temperature prior to spectra acquisition.

2.4. NIR spectra acquisition

Spectra acquisition was performed to generate a reference database. Prior to recording spectra, a gold reference was used. Spectra were then acquired on six subsets of each 680 samples. The six subsets of each sample were used, as six replicates, to minimize uncertainties due to the hypothetical heterogeneity of seed. Specifically, seeds of each sample, were six-fold randomly sampled to provide biological and analytical replicates, from each other to cover the whole sample. Seeds of each subset were then loaded in the ring cup with an internal diameter of 5 cm and the six subsets of each sample were measured in sequence. Spectra of each of the six subsets were gathered after 3 rotation scans with a spectral resolution of 16 cm^{-1} over the spectral range of 3952 cm^{-1} – 11528 cm^{-1} (867 nm – 2530 nm), using Tango spectrometer from Bruker. At the end, each sample was analyzed in six replicates, and the mean spectra were used for data analyses.

2.5. Statistics and PCA analysis

R software (R Core Team, 2021) with rchemo (Brandolini-Bunlon, et al., 2023) and rnirs packages (Lesnoff, 2021) were used to visualize raw spectra and perform data analysis. PCA over the spectral range selected from 1000 nm to 2500 nm was applied to describe variability according to the varieties and interspecific genotypes within and between environments. PCA is a multivariate unsupervised statistical method able to project multivariate data and describe relevant trends in the analyzed dataset. PCA can also reveal variables with loading that determine some inherent structure of the data, which can be interpreted in chemical terms. The reduction of the number of variables is achieved by making a linear combination of original variables, which yields the so-called principal components (PC) that are decorrelated with each other. PCA was conducted on the pretreated spectra. The full whole spectra have been pre-processed to improve the signal by reducing uncontrolled variations as noise and baseline through Savitsky Golay (SavGol) and derivative.

Mahalanobis distance was computed after PCA to check the 6 replicates distances for each sample. The Mahalanobis distances were determined in units of standard deviations from the center (mean) of the dataset. The 6 replicates were averaged or each sample, and the Mahalanobis distances were computed again.

In this study, the following PCA results were considered (i) the score plot, to visualize the projection of the sample on each PC; and (ii) the loading plot, to evaluate the influence of wavelength, on each PC. Thus, PCA allows emphasizing and interpreting variables and all relevant differences among genotypes within and between environments.

2.6. Classification using PLS-DA modeling on NIR spectra

PLS-DA was used to classifying varieties and interspecific genotypes thorough modeling and prediction of genotype-specific spectra, according to genetic and environmental origin. Data have been split by Duplex method (Snee, 1977) into train set (N = 541, 201, 108, 109, 126 respectively for Core population, AB-QTL Bafia, AB-QTL Mbalmayo, AB-QTL Nioro) and test set (N = 139, 42, 32, 31, 34 respectively for the previous populations) in each group (to keep the same proportionality). The train set was used to train the model, while the test set is used to evaluate its performance. Prior to applying PLS-DA algorithms, the train set spectra were pre-processed by SavGol filter and derivative. The best preprocessing was selected according to the error of classification by cross validation (2 K-fold group repeated 20 times) and the number of latent values was fixed. Then, these parameters were used to build the PLS-DA model and applied on test set spectra.

The resulting confusion matrix of each model were further evaluated to assess the model's performance using the following metrics for each group and for all.

- Recall (the proportion of samples of a specific class that have been predicted by the model as belonging to that class; also known as sensitivity)

$$RECALL = \frac{TP}{FN + TP}$$

- Specificity (The number of samples predicted correctly to be in the negative class out of all the samples in the dataset that actually belong to the negative class.)

$$SPECIFICITY = \frac{TN}{FP + TN}$$

- Precision (the proportion of correct predictions among all predictions for a particular class)

$$PRECISION = \frac{TP}{FP}$$

- Accuracy (the number of samples correctly classified out of all the samples present in the test set)

$$ACCURACY = \frac{TP + TN}{TP + FN + FP + TN}$$

- the proportion of false-negatives (FNR)

$$FNR = \frac{FN}{TP + FN}$$

- the proportion of false-positives (TNR)

$$TNR = \frac{TN}{FP + TN}$$

- the F1-score (the harmonic mean of precision and recall)

$$F_1 score = \frac{2TP}{2TP + FP + FN}$$

True Positive (TP) refers to a sample belonging to the positive class being classified correctly. True Negative (TN) refers to a sample belonging to the negative class being classified correctly. False Positive (FP) refers to a sample belonging to the negative class but being classified wrongly as belonging to the positive class. False Negative (FN) refers to a sample belonging to the positive class but being classified wrongly as belonging to the negative class. Model performances were evaluated by their classification accuracy, which was calculated as the ratio of the number of correctly classified samples to the total number of samples.

3. Results

3.1. Spectra profiles and quality control

From the raw spectra, eleven relevant absorbance peaks were observed around the wavelengths of 929 nm, 1033 nm, 1465 nm, 1763 nm, 2306 nm, 2350 nm and 2510 nm, with four wide spectral peaks appearing close to 1210 nm, 1723 nm, 1932 nm and 2140 nm (Fig. S1). Quality control of spectra was performed to identify atypical spectra and to check variation among the six subsets of each samples. As results, 2 of 4080 spectra (0.04%), were identified as an outlier (Fig. S1) and were discarded for analyses. PCA was performed to check the effect of date on spectra acquisition and no cluster related to date was found (Fig. S2), indicating that there were stable lab conditions during the 6 days of spectra acquisition. With few exceptions, the Mahalanobis distance (MD) among the 6 subsets of each sample was consistent among the 680 samples (Fig. 1). Thus, the spectra graph was presented in Fig. 2 as the average absorption of each sample from the 6 replicated spectra.

3.2. Genetic variability and environmental impact on intact-seed composition

The mean absorbance spectra of varieties and interspecific genotypes, according to their environment are presented in Fig. 2. A huge variation of absorbance along the spectra was observed among varieties and interspecific genotypes within and between environments. Four spectra group, superimposed on each other, was observed for all wavelength

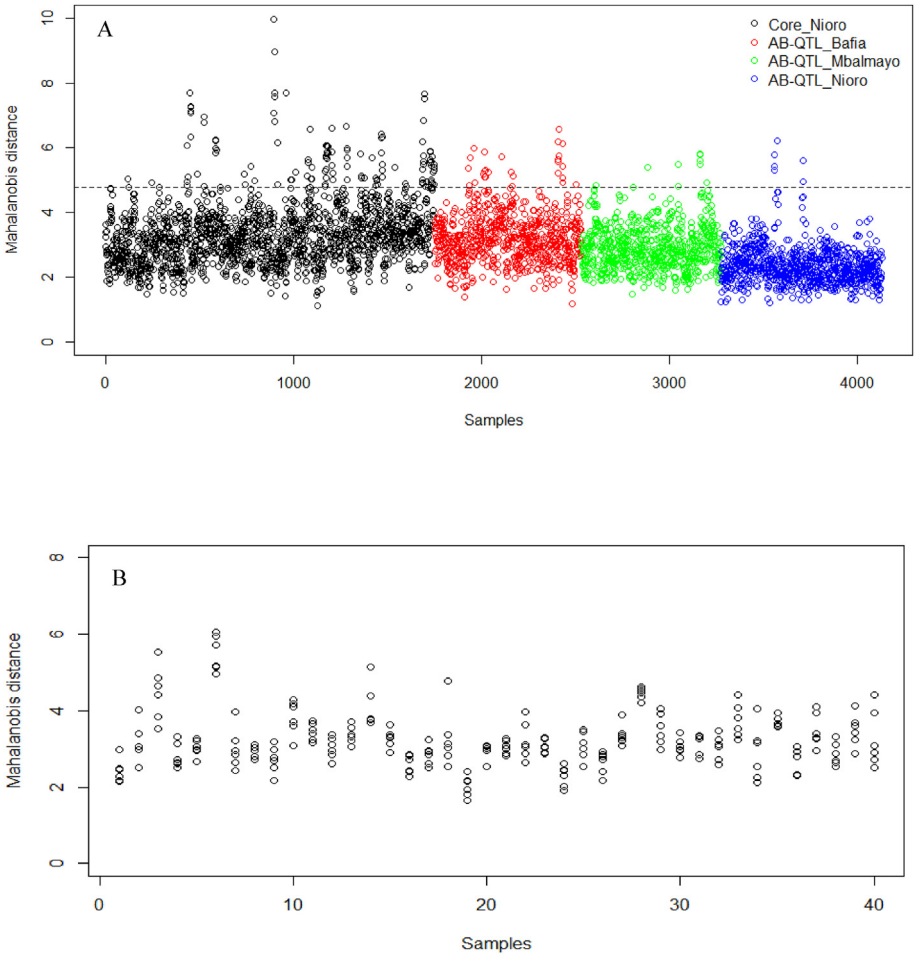

Fig. 1. Plot showing Mahalanobis distance among the six subsets of each sample of four populations. Each dot represents one spectrum. MD Details of 40 samples (B) is figured from the 4080 spectra (A) for a better MD visualization among the 6 spectra of each sample (dotted line: distance cutoff - Chi-squared distribution for Standard Deviation squared (Brandolini-Bunlon, et al., 2023)).

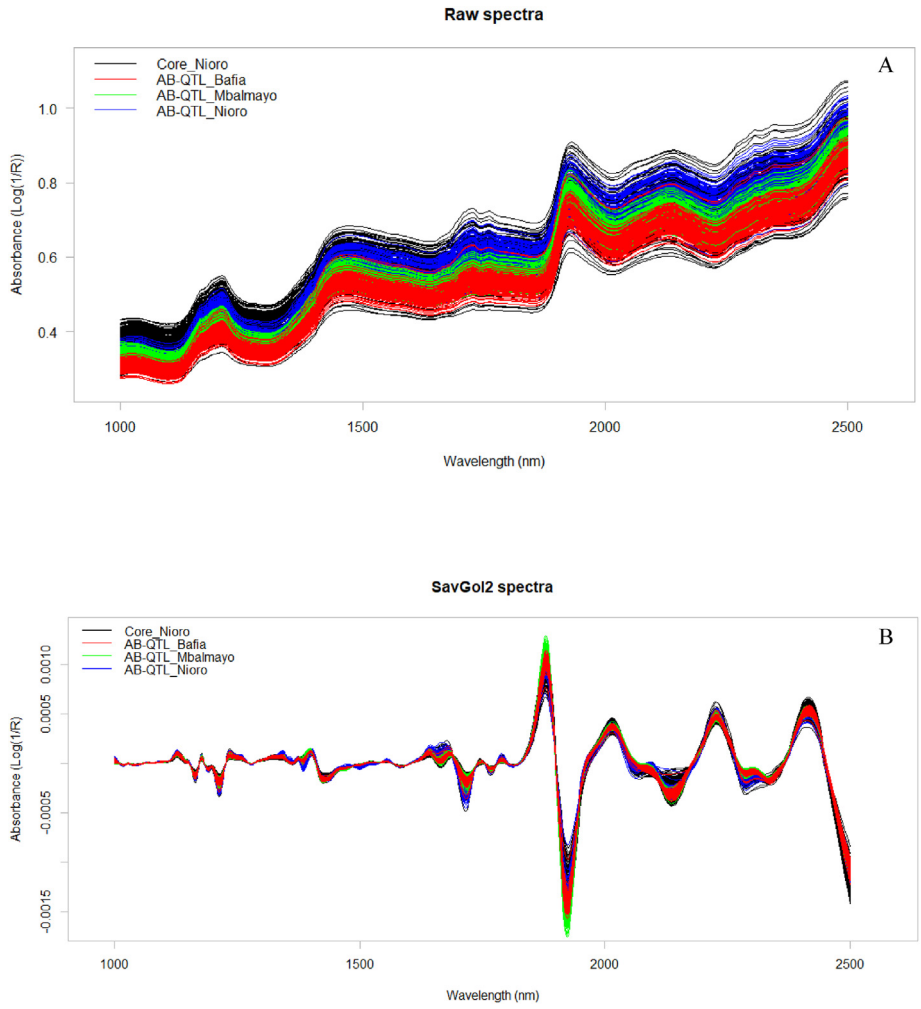

Fig. 2. NIR spectra of intact-seed according to genetic and environment origin of samples without treatment (A) and after Stavisky Golay filter with derivative 2 pre-processing (B). Spectra of varieties of the core collection are labelled in black. Spectra of interspecific AB-QTL genotypes from Bafia, Mbalmayo and Nioro, environments are labelled in red, green and blue, respectively.

from 1000 nm to 2500 nm (Fig. 2 A). Each spectra group corresponds to each of the four studied sets. The widest spectra group corresponded to the set of the core collection while the three other ones were each specific to the three sets of the interspecific population, each from one of the 3 studied environments, Bafia, Mbalmayo and Nioro (Fig. 2 A). The absorbance range of interspecific population was highest in Bafia followed by Mbalmayo and Nioro, pointing out the effect of environmental factors on chemical composition of seeds.

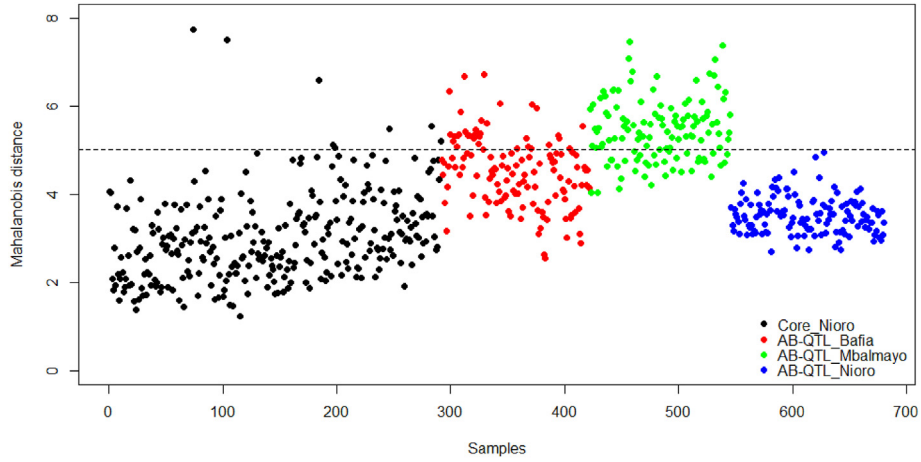

Fig. 3. Plot showing Mahalanobis distance among varieties and interspecific genotypes Each dot on the plot represents a variety or genotype. Varieties of the core collection are labelled in black. Interspecific AB-QTL genotypes from Bafia, Mbalmayo and Nioro, environments are labelled in red, green and blue, respectively **(**dotted line: distance cutoff - Chi-squared distribution for Standard Deviation squared (Brandolini-Bunlon, et al., 2023)**).**

3.3. Pretreatments effects on spectra

Different spectra pretreatments, SNV, Detrend and SavGol were applied in the raw spectra since spectral measurements can be affected by many factors leading to interference (light diffusion, scattering, …) with consequence observed as additive and multiplicative effects on raw spectra data. As example, the absorbance spectra pre-treated by SavGol filter with a window width of 15 points and first derivative was shown in Fig. 2B. As expected, the pretreatments eliminated physical effects due to seed dimension, surface of seed, etc., with consequences on light diffusion. Thus, from pretreated spectra, a huge MD variation, from 1 to 8, was found among varieties and genotypes (Fig. 3). Likewise, a distinct MD was found between 3 environments with a highest value at Mbalmayo followed by Bafia and Nioro.

3.4. Principal component analysis

PCA was performed using pretreated spectra after Savitzky-Golay filter with a window width of 15 points and first derivative. The first 5 PC represent more than 95% of the total variability with the values 60.5, 17.0, 15.5, 3.6 and 1.6, respectively. The PC1/PC2 and PC3/PC4 score plots are shows in Fig. 5. As expected, these figures show greater variability in the core collection and less variability in the other groups. The PC3/PC4 plot allows to distinguish easily the 4 seed lots. These plots showed that samples from different genetic and environmental origins are able to be well clustered and that they have great potential to be correctly identified.

Loading plots showing how each variable correlates with PC are shown in Fig. 4. The first loading indicates that the regions around 1900 nm and 2150 nm have a higher influence on PC1. Likewise, regions around 1210 nm, 1720 nm and 2300 nm seemed more related to PC2. For PC3, the region around 2400 nm seemed to be more important. PC4 was more related to 1400 nm, 1800 nm, 1950 nm and 2150 nm regions. The varieties and interspecific genotypes demonstrating contrasted scores in the top PCs were recorded (Fig. S3) and could be used further in peanut breeding programs.

3.5. Discrimination of genetically related interspecific genotypes among environments

The score plots illustrated that data could be grouped into four clusters, with overlapping the main clusters at the margin, with some interspecific genotypes and varieties superimposed, particularly, at the Nioro environment-set cluster (Fig. 5). The two most separated environments in the plane determined by plot scores were Mbalmayo and Bafia. With few exceptions, all interspecific genotypes from Mbalmayo exhibited high positive values at the PC3 compared to the other environments. This suggests that Mbalmayo environment might positively increases the seed traits associated with PC3. Finally, the African varieties studied in one environment added genetic variability to the environmental variability, resulting in a wide range of differences.

3.6. Classification based on whole seed spectra

A PLS-DA model was developed and the classification results of the model were shown in Table 2. The classification accuracy on the test set was 99.6% with correctly classified instances of the 4 samples sets i.e. African varieties in one environment and the interspecific genotypes from the 3 environments (Table 2). Interestingly, the confusion matrix achieved for the two sets, Bafia and Nioro shows 100% of instances classified correctly with 100% at both sensitivity and specificity. These two sets did not show incorrect instances, even in the model generated when all other sets were considered, thus confirming that their seed composition seemed very different from each other and from those of the other seed samples. These results showed that NIRS combined with discrimination analysis based on PLS regression is a simple and efficient tool for the classification of peanut genotypes, depending on each combination of the genetic and environment origins, which determine plant nutritional availability.

4. Discussion

The efficiency of NIR spectroscopy, as tool for fast and non-destructive large germplasm characterization in multi-environment were later discussed under the umbrella of breeding in intra and interspecific context.

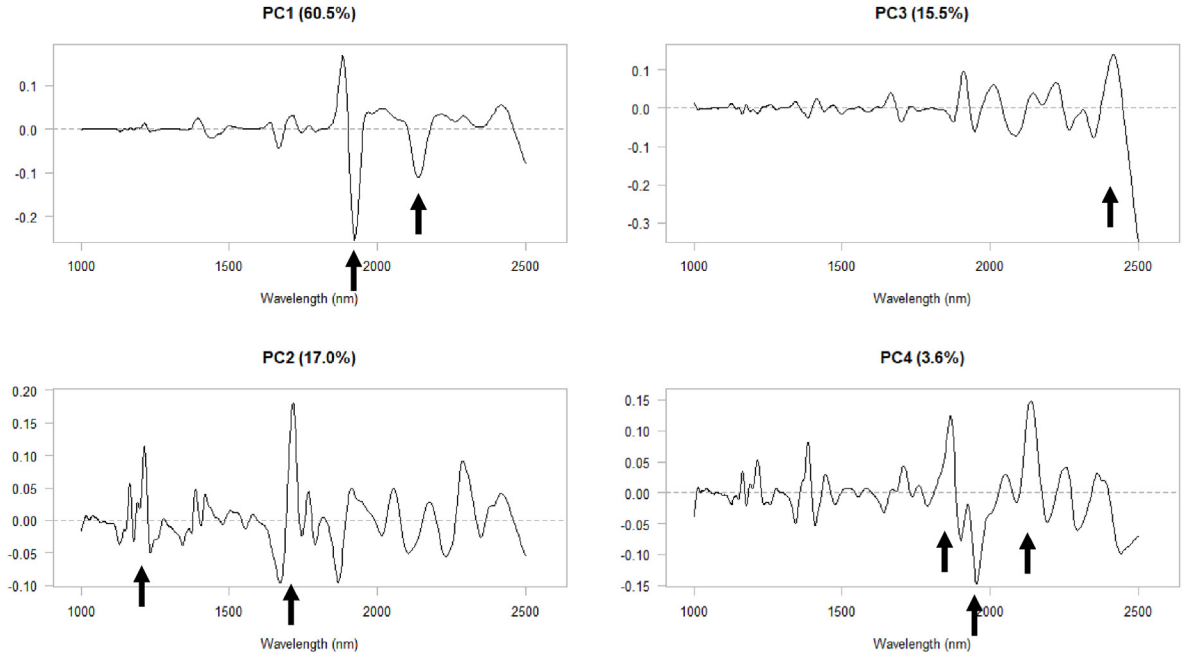

Fig. 4. PCA loading plots for the fourth first PCs showing how each variable correlate to each PC for wavelength.

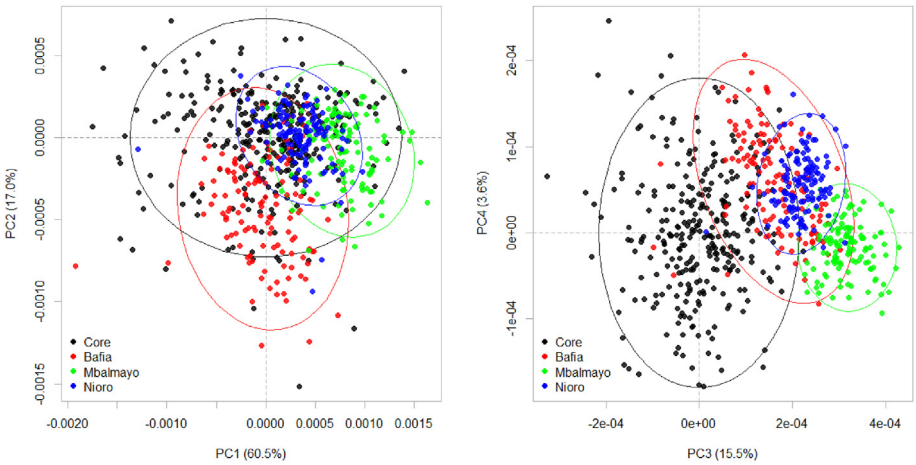

Fig. 5. PCA visualization of core varieties and interspecific genotypes among environments. PCA 2-dimensional score plots of PC2 and PC1 (A) and PC3 and PC4 (B) using NIRS spectra. Each dot on the plot represents a variety or genotype. Varieties of the core collection are labelled in black. Interspecific AB-QTL genotypes from Bafia, Mbalmayo and Nioro, environments are labelled in red, green and blue, respectively.

Table 2
Confusion matrix showing classification performance of PLS-DA model applied to test set sample samples (N = 139, Class 1: Core, Class 2: AB-QTL Bafia, Class 3: AB-QTL Mbalmayo, Class 4: AB-QTL Nioro).

		Predicted								
		1	2	3	4	Actual	Accuracy	Precison	Recall	F1-score
Actual	1	41	0	0	0	42	0.993	0.976	1.000	0.988
	2	0	32	0	0	32	1.000	1.000	1.000	1.000
	3	1	0	31	0	31	0.993	1.000	0.969	0.984
	4	0	0	0	34	34	1.000	1.000	1.000	1.000
	Pred	41	32	31	34	139				
Accuracy										0.996
Specificity										0.998
Recall										0.993
Precision										0.993
Proportion of false-negatives										0.007
Proportion of false-positives										0.002

Peanut is an important oil seed crop and the need to characterize peanut germplasm is essential as the demand for peanut is increasing continuously in various end product applications. According to the rapid and non-destructive attributes of the NIR, a total of 6 days was required to obtain all spectra of the six subsets of seed of 680 samples. The low level (0.04%) of outlier spectra on the global data set was considered as a good basis for analysis. Typical spectra observed in this study were in accordance with those reported in past studies. From raw spectra, eleven major peaks were observed. The region around 1210 nm, 1720 nm, 1763 nm, 2306 nm and 2350 nm could be assigned to fatty acids or oil content, which are generally considered as major components of peanut kernels (Govindarajan et al., 2009; Sundaram et al., 2009; Tao et al., 2019). The spectral peak around 2140 nm would likely result from the absorbance of proteins. The absorbance peak around 1465 nm might be related to the O–H overtone bond. The sharp peak around 1932 nm was due to the strong absorption of water contained in peanut kernels (Govindarajan et al., 2009; Sundaram et al., 2009; Tao et al., 2019). In the future, predictive models will be developed for nutritional content of peanut seeds.

A wide genetic variation was found among varieties and interspecific genotypes within environments. An environmental effect on seed compounds was highlighted by using the same interspecific population, thorough 3 environments. The largest variation was found in Bafia, followed by Mbalmayo and Nioro. Bafia in savanna and Mbalmayo in forest, grown under yellow and ocher vertisol, respectively in Cameroun while Nioro in Sahel in Senegal exhibited leached ferruginous soil. The interaction between all agroecological scenarios (climate, vegetation and soil) and spatial factors create a complex system of environments that affect peanut plant growth and development, leading to a discrimination among genotypes within and between environments. As previously reported by chemical studies, seed composition is influenced by environment but also has a strong genetic component. The variation of oil composition has been related to temperature (Harris and James, 1969), planting date (Andersen and Gorbet, 2002), location and soil moisture (Holaday and Pearson, 1974; Young et al., 1974), photoperiod (Dwivedi et al., 2000), market grade (Mozingo et al., 1988) and genotype (Gimode et al., 2020; Harch et al., 1995; Holaday and Pearson, 1974; Norden et al., 1987; Worthington and Hammons, 1971). However, with multiple environmental factors mentioned above, it is difficult to decipher factors underlining variation in this study. Likewise, identifying suitable peanut genotypes for global ecological zones remains a challenging task due to the significant genotype variability across environments. Finally, the African varieties studied in one environment added genetic variability to the environmental one, resulting in a wide range of variability.

According to the spectra profiles and PCA plot, the genetic pattern of interspecific population covers, remarkably half of the spectrum of the core-collection, that turned out to be largest, as we expected. Interestingly, we found specific genetic variation among interspecific genotypes that was not subtle cover by the core-collection at the common Nioro environment. Interspecific genotypes with positive value on the main PCA axis were recorded as promising genotypes for quality traits. These genotypes could be recommended for further breeding for developing suitable varieties. In this respect, evaluation of the segregating interspecific population could further ease the discovery of QTL and valuable

wild genes that contribute to improved seed quality.

A PLS-DA model was successful developed from seed spectra to classify varieties and genotypes according to their genetic and environmental origin. A robust prediction accuracy of 99.6% was achieved. The confusion matrix achieved for the two environments, Nioro and Bafia shows 100% of instances classified correctly with 100% at both sensitivity and specificity. This confirm that their seed chemical composition was very different from each other and from those of the other seed samples. These results suggested that PLS-DA model could be used to classify peanut genotypes depending on the combination of the genetic and environment origins of seeds, which influence plant nutritional properties. In further studies, the current model would be confronted to wide others breeding populations in different environment to predict genetic and environment origin and nutritional content of whole seeds.

Breeding programs need germplasm diversity with extreme values for any nutritional trait. The magnitude of the genetic influence among varieties and genotypes suggested that nutritional related traits were amenable to improvement through intra and interspecific breeding. Nowadays, the availability of NIR data, might accelerate the utilization of germplasm and genetic diversity both in breeding programs. The observed genotypic variations and their variability across environments have deep implications for breeding programs. It seems feasible to achieve a fruitful goal in breeding on the basis of seed composition, because both the environmental effects found in the different locations and the genetic effects of the different varieties and interspecific derivatives influence the seed chemical compounds. Interestingly, even if the core-collection turned out to be the widest, a huge, specific and subtle genetic variation was found among interspecific genotypes, that was not covered by the 300 varieties. This offers the possibility of discovering new sources of diverse nutritional polymorphisms from wild derivatives. As early reported, three introgression lines with elevated Oleic/Linoleic profiles were found using chemical survey of 77 interspecific lines (Gimode et al., 2020). Interspecific hybridization has recently played an important role in accessing useful alleles from the wild (Favero et al., 2006; Mallikarjuna et al., 2011b; Simpson, 2001). We recorded varieties and interspecific lines with favorable spectra profiles. Thus, those potential chemotypes, with favorable chemical profiles could be further evaluated and promoted as a valuable genetic material to develop suitable varieties. Moreover, the comprehension of the genetic and environments determinants of nutritional traits might help in marker-assisted selection, accelerating the breeding of superior varieties tailored for specific environments and end-user demands.

5. Conclusion

The present study was carried out to investigate the potential of NIR coupled with chemometric to rapidly assess peanut germplasm from whole seed of a core-collection and an interspecific population, field-evaluated in 3 environments. This paper describes the NIR inputs to control breeding populations and assess germplasm variability, as we expected before the genetic studies. A wide variability of seed compounds was observed in the given germplasm, within and between environments, as revealed by spectra and multivariate analysis. A PLS-DA model was developed with a strong classification accuracy, aiming to properly predict each whole seed sample according to environment. These results indicate that NIR coupled with chemometric seem useful to accurately assess and distinguish intact-seed within different environments, that would ease further prediction of intact-seed nutritional content and utilization of germplasm in breeding programs.

CRediT authorship contribution statement

Fentanesh Chekole Kassie: Writing – original draft, Investigation. **Gilles Chaix:** Writing – review & editing, Validation, Formal analysis. **Hermine Bille Ngalle:** Supervision, Resources, Project administration, Funding acquisition. **Maguette Seye:** Investigation. **Coura Fall:** Investigation. **Hodo-Abalo Tossim:** Investigation. **Aissatou Sambou:** Investigation. **Olivier Gibert:** Writing – review & editing. **Fabrice Davrieux:** Data curation. **Joseph Martin Bell:** Supervision, Resources, Project administration, Funding acquisition. **Jean-François Rami:** Supervision, Resources, Project administration, Funding acquisition. **Daniel Fonceka:** Supervision, Resources, Project administration, Funding acquisition. **Joël Romaric Nguepjop:** Writing – review & editing, Methodology, Formal analysis, 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

This study was supported by the GENES intra-Africa Academic Mobility scheme of the European Union (EU-GENES: EACEA/2917/2552) and by the DESIRA-ABEE project funded by European Union. The authors thank Paulo Salgado, Rokhaya Diarra, Aissatou Bathily and Laurent Bonnal, for technical advices for spectra acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ocsci.2024.03.003.

References

Andersen, P.C., Gorbet, D.W., 2002. Influence of year and planting date on fatty acid chemistry of high oleic acid and normal peanut genotypes. J. Agric. Food Chem. 50, 1298–1305. https://doi.org/10.1021/jf0113171.

Bilal, M., Xiaobo, Z., Arslan, M., Tahir, H.E., Azam, M., Junjun, Z., Basheer, S., Abdullah, 2020. Rapid determination of the chemical compositions of peanut seed (Arachis hypogaea.) Using portable near-infrared spectroscopy. Vib. Spectrosc. 110, 103138. https://doi.org/10.1016/j.vibspec.2020.103138.

Bogie, N.A., Bayala, R., Diedhiou, I., Dick, R.P., Ghezzehei, T.A., 2018. Alteration of soil physical properties and processes after ten years of intercropping with native shrubs in the Sahel. Soil Tillage Res. 182, 153–163. https://doi.org/10.1016/j.still.2018.05.010.

Brandolini-Bunlon, M., Jaillais, B., Roger, J.M., Lesnoff, M., 2023. rchemo: Dimension Reduction, Regression and Discrimination for Chemometrics. R package version 0.1-1. https://CRAN.R-project.org/package=rchemo.

Burow, M.D., Simpson, C.E., Faries, M.W., Starr, J.L., Paterson, A.H., 2009. Molecular biogeographic study of recently described B- and A-genome Arachis species, also providing new insights into the origins of cultivated peanut. Genome 52, 107–119.

Conde, S., Rami, J.-F., Okello, D.K., Sambou, A., Muitia, A., Oteng-Frimpong, R., Makweti, L., Sako, D., Faye, I., Chintu, J., Coulibaly, A.M., Miningou, A., Asibuo, J.Y., Konate, M., Banla, E.M., Seye, M., Djiboune, Y.R., Tossim, H.-A., Sylla, S.N., Hoisington, D., Clevenger, J., Chu, Y., Tallury, S., Ozias-Akins, P., Fonceka, D., 2023. The Groundnut Improvement Network for Africa (GINA) Germplasm collection: a unique genetic resource for breeding and gene discovery. G3 GenesGenomesGenetics jkad244. https://doi.org/10.1093/g3journal/jkad244.

Davis, B.I., Agraz, C.B., Kline, M., Gottschall, E., Nolt, M., Whitaker, T.B., Osborne, J.A., Tengstrand, E., Ostrowski, K., Teixeira, R., Davis, J.P., 2021. Measurements of high oleic purity in peanut lots using rapid, single kernel near-infrared reflectance spectroscopy. J. Am. Oil Chem. Soc. 98, 621–632. https://doi.org/10.1002/aocs.12487.

Desmae, H., Janila, P., Okori, P., Pandey, M.K., Motagi, B.N., Monyo, E., Mponda, O., Okello, D., Sako, D., Echeckwu, C., Oteng-Frimpong, R., Miningou, A., Ojiewo, C., Varshney, R.K., 2019. Genetics, genomics and breeding of groundnut (Arachis hypogaea L.). Plant Breed. 138, 425–444. https://doi.org/10.1111/pbr.12645.

Dwivedi, S.L., Nigam, S.N., Rao, R.C.N., 2000. Photoperiod effects on seed quality traits in peanut. Crop Sci. 40, 1223–1227. https://doi.org/10.2135/cropsci2000.4051223x.

Fan, P., Song, W., Kang, Y., Wan, L., Lei, Y., Huai, D., Chen, Y., Wang, X., Jiang, H., Yan, L., Liao, B., 2020. Phenotypic identification of peanut germplasm for resistance to southern stem rot. Oil Crop Sci 5, 174–179. https://doi.org/10.1016/j.ocsci.2020.12.001.

FAOSTAT, 2020. Statistical Database. Food and Agriculture Organization of the United Nations, Rome.

Favero, A.P., Simpson, C.E., Valls, J.F.M., Vello, N.A., 2006. Study of the evolution of cultivated peanut through crossability studies among Arachis ipaensis, A. Duranensis, and A. hypogaea. Crop Sci. 46, 1546–1552. https://doi.org/10.2135/cropsci2005.09-0331.

Fonceka, D., Tossim, H.-A., Rivallan, R., Vignes, H., Faye, I., Ndoye, O., Moretzsohn, M.C., Bertioli, D.J., Glaszmann, J.-C., Courtois, B., Rami, J.-F., 2012a. Fostered and left

behind alleles in peanut: interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. BMC Plant Biol. 12, 26. https://doi.org/10.1186/1471-2229-12-26.
Fonceka, D., Tossim, H.-A., Rivallan, R., Vignes, H., Lacut, E., de Bellis, F., Faye, I., Ndoye, O., Leal-Bertioli, S.C.M., Valls, J.F.M., Bertioli, D.J., Glaszmann, J.-C., Courtois, B., Rami, J.-F., 2012b. Construction of chromosome segment substitution lines in peanut (Arachis hypogaea L.) using a wild synthetic and QTL mapping for plant morphology. PLoS One 7, e48642. https://doi.org/10.1371/journal.pone.0048642.
Fordellone, M., Bellincontro, A., Mencarelli, F., 2019. Partial least squares discriminant analysis: a dimensionality reduction method to classify hyperspectral data. Stat. Appl. - Ital. J. Appl. Stat. 181–200. https://doi.org/10.26398/IJAS.0031-010.
Gimode, D., Chu, Y., Dean, L., Holbrook, C., Fonceka, D., Ozias-Akins, P., 2020. Seed composition survey of a peanut CSSL population reveals introgression lines with elevated oleic/linoleic profiles. Peanut Sci. 47, 139–149. https://doi.org/10.3146/PS20-17.1.
Govindarajan, K.N., Kandala, V., Subbiah, J., 2009. NIR reflectance spectroscopy for nondestructive moisture content determination in peanut kernels. Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 52. https://doi.org/10.13031/2013.29117.
Grosso, N.R., Lamarque, A., Maestri, D.M., Zygadlo, J.A., Guzmán, C.A., 1994. Fatty acid variation of runner peanut (Arachis hypogaea L.) among geographic localities from Córdoba (Argentina). J. Am. Oil Chem. Soc. 71, 541–542. https://doi.org/10.1007/BF02540669.
Grosso, N.R., Nepote, V., Guzman, C.A., 2000. Chemical composition of some wild peanut species (Arachis L.) seeds. J. Agric. Food Chem. 48, 806–809. https://doi.org/10.1021/jf9901744.
Harch, B.D., Basford, K.E., DeLacy, I.H., Lawrence, P.K., Cruickshank, A., 1995. Patterns of diversity in fatty acid composition in the Australian groundnut germplasm collection. Genet. Resour. Crop Evol. 42, 243–256. https://doi.org/10.1007/BF02431259.
Harris, P., James, A.T., 1969. The effect of low temperatures on fatty acid biosynthesis in plants. Biochem. J. 112, 325–330.
Holaday, C.E., Pearson, J.L., 1974. Effects of genotype and production area on the fatty acid composition, total oil and total protein in peanuts. J. Food Sci. 39, 1206–1209.
Huang, H., Liu, L., Ngadi, M.O., 2014. Recent developments in hyperspectral imaging for assessment of food quality and safety. Sensors 14, 7248–7276. https://doi.org/10.3390/s140407248.
Isleib, T.G., Tillman, B.L., Pattee, H.E., Sanders, T.H., Hendrix, K.W., Dean, L.O., 2008. Genotype-by-Environment interactions for seed composition traits of breeding lines in the uniform peanut performance test. Peanut Sci. 35 (1), 130–138. https://doi.org/10.3146/PS08-001.
Janila, P., Nigam, S.N., Pandey, M.K., Nagesh, P., Varshney, R.K., 2013. Groundnut improvement: use of genetic and genomic tools. Front. Plant Sci. 4, 23. https://doi.org/10.3389/fpls.2013.00023.
Kandala, C.V.K., Butts, C.L., Lamb, M.C., 2008. Moisture content determination for in-shell peanuts with a low-cost impedance analyzer and capacitor sensor. Moisture Content Determ. -Shell Peanuts Low-Cost Impedance Anal. Capacit. Sens. 51, 1377–1381.
Kumari, V., Gowda, M.V.C., Tasiwal, V., Pandey, M.K., Bhat, R.S., Mallikarjuna, N., Upadhyaya, H.D., Varshney, R.K., 2014. Diversification of primary gene pool through introgression of resistance to foliar diseases from synthetic amphidiploids to cultivated groundnut (Arachis hypogaea L.). Crops J 2, 110–119. https://doi.org/10.1016/j.cj.2014.03.002.
Lesnoff, M., 2021. R package rnirs: dimension reduction, regression and discrimination for chemometrics [WWW Document]. https://github.com/mlesnoff/rnirs.URL. (Accessed 8 January 2023).
Li, Z., Tang, X., Shen, Z., Yang, K., Zhao, L., Li, Y., 2019. Comprehensive comparison of multiple quantitative near-infrared spectroscopy models for Aspergillus flavus contamination detection in peanut. J. Sci. Food Agric. 99, 5671–5679. https://doi.org/10.1002/jsfa.9828.
Liu, L., Jiang, H., Chen, Q., 2022. Rapid determination of acidity index of peanut during storage by a portable near-infrared spectroscopy system. Infrared Phys. Technol. 127, 104472. https://doi.org/10.1016/j.infrared.2022.104472.
Mallikarjuna, N., Jadhav, D.R., Reddy, K., Husain, F., Das, K., 2011a. Screening new Arachis amphidiploids, and autotetraploids for resistance to late leaf spot by detached leaf technique. Eur. J. Plant Pathol. 132, 17–21. https://doi.org/10.1007/s10658-011-9859-2.
Mallikarjuna, N., Senthilvel, S., Hoisington, D., 2011b. Development of new sources of tetraploid Arachis to broaden the genetic base of cultivated groundnut (Arachis hypogaea L.). Genet. Resour. Crop Evol. 58, 889–907. https://doi.org/10.1007/s10722-010-9627-8.
Mozingo, R.W., Coffelt, T.A., Wynne, J.C., 1988. Market grade effects on fatty acid composition of five peanut cultivars. Agron. J. 80, 73–75. https://doi.org/10.2134/agronj1988.00021962008000010017x.
Nawade, B., Mishra, G.P., Radhakrishnan, T., Dodia, S.M., Ahmad, S., Kumar, Abhay, Kumar, Atul, Kundu, R., 2018. High oleic peanut breeding: achievements, perspectives, and prospects. Trends Food Sci. Technol. 78, 107–119. https://doi.org/10.1016/j.tifs.2018.05.022.
Nguepjop, J.R., Tossim, H.-A., Bell, J.M., Rami, J.-F., Sharma, S., Courtois, B., Mallikarjuna, N., Sane, D., Fonceka, D., 2016. Evidence of genomic exchanges between homeologous chromosomes in a cross of peanut with newly synthetized allotetraploid hybrids. Front. Plant Sci. 7, 1635. https://doi.org/10.3389/fpls.2016.01635.
Norden, A.J., Gorbet, D.W., Knauft, D.A., Young, C.T., 1987. Variability in oil quality among peanut genotypes in the Florida breeding program 1. Peanut Sci. 14, 7–11.
Panero, F.D.S., Smiderle, O., Panero, J.S., Faria, F.S.D.V., Panero, P.D.S., Rodriguez, A.F.R., 2022. Non-destructive genotyping of cultivars and strains of sesame through NIR spectroscopy and chemometrics. Biosensors 12, 69. https://doi.org/10.3390/bios12020069.
Panero, J.S., Silva, H.E.B. da, Panero, P.S., Smiderle, O.J., Panero, F.S., Faria, F.S.E.D.V., Rodriguez, A.F.R., 2018. Separation of cultivars of soybeans by chemometric methods using near infrared spectroscopy. J. Agric. Sci. 10, p351. https://doi.org/10.5539/jas.v10n4p351.
Parilli-Moser, I., Hurtado-Barroso, S., Guasch-Ferré, M., Lamuela-Raventós, R.M., 2022. Effect of peanut consumption on cardiovascular risk factors: a randomized clinical trial and meta-analysis. Front. Nutr. 9.
Parmar, S., Sharma, V., Bomireddy, D., Soni, P., Joshi, P., Gangurde, S., Wang, J., Bera, S.K., Bhat, R., Desmae, H., Shirasawa, K., Varshney, R., Pandey, M., 2022. Recent Advances in Genetics, Genomics, and Breeding for Nutritional Quality in Groundnut, pp. 111–137. https://doi.org/10.1007/978-3-030-81107-5_4.
R Core Team, 2021. R: the R Project for Statistical Computing [WWW Document]. https://www.r-project.org/. (Accessed 8 January 2023).
Rabanera, J.D., Guzman, J.D., Yaptenco, K.F., 2021. Rapid and Non-destructive measurement of moisture content of peanut (Arachis hypogaea L.) kernel using a near-infrared hyperspectral imaging technique. J. Food Meas. Char. 15, 3069–3078. https://doi.org/10.1007/s11694-021-00894-x.
Sampaio, P.S., Brites, C.M., Sampaio, P.S., Brites, C.M., 2021. Near-infrared spectroscopy and machine learning: analysis and classification methods of rice. Integrative Advances in Rice Research. IntechOpen. https://doi.org/10.5772/intechopen.99017.
Simpson, C.E., 2001. Use of wild Arachis species/introgression of genes into A. hypogaea L. Peanut Sci. 28, 114–116. https://doi.org/10.3146/i0095-3679-28-2-12.
Simpson, C.E., Krapovickas, A., Valls, J.F.M., 2001. History of Arachis including evidence of A. hypogaea L. Progenitors. Peanut Sci. 28, 78–80. https://doi.org/10.3146/i0095-3679-28-2-7.
Singh, T., Garg, N.M., Iyengar, S.R.S., Singh, V., 2023. Near-infrared hyperspectral imaging for determination of protein content in barley samples using convolutional neural network. J. Food Meas. Char. 1–13.
Snee, R.D., 1977. Validation of regression models: methods and examples. Technometrics 19, 415–428. https://doi.org/10.1080/00401706.1977.10489581.
Song, W., Wang, H., Maguire, P., Nibouche, O., 2018. Nearest clusters based partial least squares discriminant analysis for the classification of spectral data. Anal. Chim. Acta 1009, 27–38. https://doi.org/10.1016/j.aca.2018.01.023.
Sundaram, J., Kandala, C.V., Butts, C.L., 2009. Application of near infrared spectroscopy to peanut grading and quality analysis: overview. Sens. Instrum. Food Qual. Saf. 3, 156–164. https://doi.org/10.1007/s11694-009-9081-5.
Tang, Y., Qiu, X., Hu, C., Li, J., Wu, L., Wang, W., Li, Xin, Li, Xiaoting, Zhu, H., Sui, J., Wang, J., Qiao, L., 2022. Breeding of a new variety of peanut with high-oleic-acid content and high-yield by marker-assisted backcrossing. Mol. Breed. New Strateg. Plant Improv. 42, 42. https://doi.org/10.1007/s11032-022-01313-9.
Tao, F., Yao, H., Hruska, Z., Liu, Y., Rajasekaran, K., Bhatnagar, D., 2019. Use of visible–near-infrared (Vis-NIR) spectroscopy to detect aflatoxin B1 on peanut kernels. Appl. Spectrosc. 73, 415–423. https://doi.org/10.1177/0003702819829725.
Temga, P., Sababa, E., Mamdem, L., Bijeck, M., Azinwi Tamfuh, P., Tehna, N., Zame, P., Onana, V.L., Jean Pierre, N., Bitom, L., Ndjigui, P.-D., 2021. Rare earth elements in tropical soils, Cameroon soils (Central Africa). Geoderma Reg. 25, e00369. https://doi.org/10.1016/j.geodrs.2021.e00369.
Tian, W., Zang, L., Nie, L., Li, L., Zhong, L., Guo, X., Huang, S., Zang, H., 2023. Structural analysis and classification of low-molecular-weight hyaluronic acid by near-infrared spectroscopy: a comparison between traditional machine learning and deep learning. Molecules 28, 809. https://doi.org/10.3390/molecules28020809.
Tossim, H.-A., Nguepjop, J.R., Diatta, C., Sambou, A., Seye, M., Sane, D., Rami, J.-F., Fonceka, D., 2020. Assessment of 16 peanut (Arachis hypogaea L.) CSSLs derived from an interspecific cross for yield and yield component traits: QTL validation. Agronomy 10, 583. https://doi.org/10.3390/agronomy10040583.
Upadhyaya, H.D., 2005. Variability for drought resistance related traits in the mini core collection of peanut. Crop Sci. 45, 1432–1440. https://doi.org/10.2135/cropsci2004.0389.
Upadhyaya, H.D., Dwivedi, S.L., Nadaf, H.L., Singh, S., 2011. Phenotypic diversity and identification of wild Arachis accessions with useful agronomic and nutritional traits. Euphytica. https://doi.org/10.1007/s10681-011-0518-7.
Wang, Ling, Wang, Lei, Tan, M., Yu, H., Wang, J., Li, Y., Wang, W., Yan, X., Wang, Lijun, 2022. Rapid identification and preliminary evaluation of quality characters of oilseed sunflower by near infrared spectroscopy. Oil Crop Sci 7, 142–148. https://doi.org/10.1016/j.ocsci.2022.08.003.
Wang, Y., Song, S., 2023. Variety identification of sweet maize seeds based on hyperspectral imaging combined with deep learning. Infrared Phys. Technol. 130, 104611. https://doi.org/10.1016/j.infrared.2023.104611.
Wang, Z., Zhang, Y., Yan, L., Chen, Y., Kang, Y., Huai, D., Wang, X., Liu, K., Jiang, H., Lei, Y., Liao, B., 2023. Correlation and variability analysis of yield and quality related traits in different peanut varieties across various ecological zones of China. Oil Crop Sci 8, 236–242. https://doi.org/10.1016/j.ocsci.2023.12.001.
Worthington, R.E., Hammons, R.O., 1971. Genotypic Variation in Fatty Acid Composition and Stability of Arachis hypogaea L. Oil. Oleagineux.

Xu, P., Sun, W., Xu, K., Zhang, Y., Tan, Q., Qing, Y., Yang, R., 2023. Identification of defective maize seeds using hyperspectral imaging combined with deep learning. Foods 12, 144. https://doi.org/10.3390/foods12010144.

Young, C.T., Worthington, R.E., Hammons, R.O., Matlock, R.S., Waller, G.R., Morrison, R.D., 1974. Fatty acid composition of Spanish peanut oils as influenced by planting location, soil moisture conditions, variety, and season. J. Am. Oil Chem. Soc. 51, 312–315. https://doi.org/10.1007/BF02633004.

Yu, H., Liu, H., Erasmus, S.W., Zhao, S., Wang, Q., van Ruth, S.M., 2020. Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics. Ind. Crops Prod. 158, 112956. https://doi.org/10.1016/j.indcrop.2020.112956.