

Research

Genetic variability of anti-nutritional factors among soybean (*Glycine max.*) germplasm

Norberto Jose Palange^{1,2,3}  · Tonny Obua¹  · Julius Pyton Sserumaga⁴  · Enoch Wembabazi⁵ · Mildred Ochwo-Ssemakula¹ · Arfang Badji^{1,2}  · Ephraim Nuwamanya^{1,5}  · Isaac Onziga Dramadri^{1,2} · Richard Edema^{1,2} · Moses Matovu⁴ · Sharon Valerie Kweyu² · Aurelio Macaringue¹ · Antonio Maquil¹ · Phinehas Tukamuhabwa¹

Received: 24 October 2024 / Accepted: 27 March 2025

Published online: 28 April 2025

© The Author(s) 2025 

Abstract

Soybean is a vital source of vegetable protein, carbohydrates, oil, and minerals. However, anti-nutritional factors (ANFs) can diminish its nutritional value for monogastric animals, including humans. This study assessed the variability of phytate and trypsin inhibitors in Ugandan soybean germplasm. A total of 308 soybean genotypes were analyzed for yield, phytate, and total trypsin inhibitors (TTI) using spectrophotometric methods. Significant variation was observed for yield, phytate, and TTI levels ($p < 0.001$). Phytate levels ranged from 14.8 mg kg⁻¹ (BSPS 48A-6-3) to 6928.8 mg kg⁻¹ (NGDT 2.15-7), while TTI varied from 14.7 mg kg⁻¹ (DN 16_N) to 1534.8 mg kg⁻¹ (Duiker). Using a weighted selection index, the best-performing genotypes, Ux 990-044, Ux 990-102, and S 6.22B, combined high yield with low ANF levels. Broad-sense heritability (H^2) was 0.68 for phytate, 0.84 for TTI, and 0.08 for yield. The phenotypic coefficient of variation (PCV) for phytate (203.1%) was significantly higher than the genetic coefficient of variation (GCV) (137.6%), indicating substantial environmental influence, while moderate for TTI (PCV = 34.6%; GCV = 29.2%). Genetic advance (GA) values were 1667.6 mg kg⁻¹ for phytate and 504.3 mg kg⁻¹ for TTI, highlighting considerable genetic variability and heritability. These findings have practical implications for breeding programs; the identified genotypes that combine high yield and reduced ANFs can reduce processing costs for food and feed industries while enhancing nutritional security, especially in regions dependent on soybean as a protein source.

Keywords Soybean · Anti-nutritional factors · Phytate · Trypsin inhibitors · Variability

1 Introduction

Soybean (*Glycine max.*, $2n = 40$) is rich in protein (40–50%), lipids (20–30%), carbohydrates (26–30%) [1–4] and micronutrients (3.93–6.15%) [5], and can be consumed in various forms, including roasted whole beans, baked beans, soy paste, soymilk, and fermented soybean curd [6, 7]. Due to its rich nutritional profile, soybean is widely used to combat malnutrition globally, particularly in developing countries [3, 8]. However, the soybean nutritional value is negatively

✉ Julius Pyton Sserumaga, j.serumaga@gmail.com | ¹Department of Crop Science and Horticulture, College of Agriculture and Environmental Science, Makerere University, P. O. Box 7062, Kampala, Uganda. ²Makerere Regional Centre for Crop Improvement (MaRCCI), College of Agriculture and Environmental Science, Makerere University, P. O. Box 7062, Kampala, Uganda. ³Faculty of Food and Agriculture Sciences, Rovuma University, P. O. Box 544, Nampula, Mozambique. ⁴National Agricultural Research Organization, National Livestock Resources Research Institute (NaLIRRI), P.O. Box 5704, Kampala, Uganda. ⁵National Agricultural Research Organization, National Crops Resources Research Institute (NaCRRRI), P. O. Box 7084, Kampala, Uganda.



affected by the presence of secondary metabolism products known as anti-nutritional factors (ANFs). While ANFs play a crucial role in plant defense [9] and have significant beneficial effects in animals [10], if present at high concentrations, they can hinder nutrient absorption in animals and humans by forming complexes with nutrients such as proteins, sugars, and minerals [11, 12]. Studies on other legumes, such as common beans and lentils, have shown similar challenges with ANFs like tannins and oligosaccharides affecting mineral absorption [11, 13]. These studies emphasize the broader relevance of addressing ANFs in legumes to improve nutritional quality.

The main ANFs in soybean seeds include protease inhibitors, metal chelators such as phytate, oligosaccharides, antigenic factors [14], saponins, and tannins [15]. The significance of phytate and TTI lies in their impact on the nutritional value of soybeans, as they influence two key groups of biomolecules – proteins and minerals. These molecules are crucial for the structural and functional processes in both animals and humans. Phytate (PA; *myo-inositol 1,2,3,4,5,6 hexakisphosphate*) chelates minerals such as Cu^{2+} , Ca^{2+} , Zn^{2+} , and Fe^{3+} , resulting in insoluble phytate-metal complexes [12], thereby limiting mineral bioavailability. Phytate can also form extremely persistent compounds with proteins preventing their digestion and absorption [13]. Trypsin inhibitors are the major proteinase inhibitors in soybean, accounting for about 6% of the total protein [16]. They form an inhibitor-enzyme complex that blocks the enzyme's active site, thereby impairing the digestion and absorption of proteins and consequently reducing animal and human growth [17]. To minimize nutrient losses, ideally, phytate should be less than 0.25 mg per 100 g (2.500 mg kg^{-1}) [18] while trypsin inhibitors should be between 1.750 and 2.500 mg kg^{-1} [19].

To address the negative effect of anti-nutritional factors on animals and humans, various processing methods have been proposed, including physical, chemical and enzymatic approaches. Traditional physical and chemical techniques such as soaking, cooking, autoclaving, microwave cooking, extrusion, germination, irradiation [20], debranning [11], dehulling [21], roasting, sprouting [22] and enzymatic treatments [23] can be used alone or in combination. However, the methods are often expensive, some requiring high expertise and can denature proteins [24]. Thus, the most practical and cost-effective approach is to breed genotypes that have reduced, well-documented levels of anti-nutritional factors for animal and human consumption.

Different breeding strategies are available for the development of soybean varieties with low anti-nutritional content, including conventional breeding [25, 26], mutational breeding [27], and modern plant breeding techniques [27–30] such as marker-assisted selection [31], CRISPR-Cas9 [27, 32]. Although most of these approaches still require technical expertise, and ethical or regulatory considerations, for instance challenges in breeding genetically modified crops; genotypes with low ANFs are crucial for nutrient utilization [14], reducing processing costs [33], and improving nutritional security and safety [34].

Several studies have extensively focused on morphological and agronomic traits [35], diseases [36–38]; yield and nutritional parameters [38, 39]; adaptability [40], assessing genetic diversity using single sequence repeats (SSR) molecular markers [41], genotype identification using expressed sequence tag derived simple sequence repeat (EST-SSR) [42], and SNP markers [37, 43]. However, anti-nutritional factors have not been extensively studied.

The lack of a thorough biochemical characterization of germplasm could result in the deployment of soybean varieties with high levels of anti-nutritional factors, potentially compromising the quality of consumed products. Uganda's soybean germplasm represents diverse genetic resources adapted to local environmental conditions. This genetic diversity provides a foundation for developing nutritionally superior and regionally adapted soybean varieties. Understanding the variability in anti-nutritional factors (ANFs) within this germplasm allows breeders to enhance the nutritional quality of soybean for human and animal consumption while reducing the dependency on expensive processing methods. Therefore, this study aimed to assess the variability of phytate and total trypsin inhibitors in Ugandan soybean germplasm. It sought to provide a basis for the improvement of soybean nutritional value while maintaining desirable traits such as yield for feed and food.

2 Material and methods

2.1 Germplasm and experimental design

A collection of 308 soybeans, maintained by the Makerere University Centre for Soybean Improvement and Development (MAKCSID), composed of lines sourced from Uganda (120), Taiwan (27), the USA (132), Japan (19), Nigeria (1), and Zimbabwe (13) was used in this study. To standardize the conditions and ensure consistent seed multiplication, genotypes were planted at the Makerere University Agricultural Research Institute Kabanyolo (MUARIK) in 2023. Kabanyolo is

located in the Central Region of Uganda, at coordinates 0°28' N, 32°36' E and an altitude of 1,180 m above sea level. The mean annual temperature is 21.4 °C, and the mean annual rainfall is 1,234 mm [39]. The experiment was laid out in an augmented design with 31 blocks, each containing 10 plots, and this design was chosen to ensure robust evaluations of 308 genotypes. This setup was implemented at Makerere University Agricultural Research Institute Kabanyolo, and the results reflect consistent experimental replication across all genotypes. Planting was done with *inter* and *intra*-row spacing of 60 cm × 5 cm, respectively. The study was conducted in a single environment.

2.2 Biochemical analysis

Genotypes were harvested separately at maturity, threshed and corrected to 10% moisture content by sun-drying before the biochemical analysis. Dried soybean seeds were ground using a food processor grinder (FOSS, Brook Crompton Laboratory mill, type 2-TDAB03J, England, 2014) at the Nutritional and Biochemical Laboratory of the National Crops Resources Research Institute (NaCRRI) in Uganda. The ground samples were then subjected to analysis to determine levels of phytate and total trypsin inhibitors.

2.2.1 Quantitative determination of phytate

Phytate extraction was performed through an acidic method as described by [44], with slight modifications on the sample's initial volume and refrigeration during centrifugation. Twenty milliliters of 0.5 M HCl was added to 0.5 g of finely powdered triplicate soybean samples, vortexed, and shaken on an orbital shaker for 1 h. The homogenized mixture was centrifuged at 4000 rpm for 30 min at 4 °C using a HERMLE 300 K centrifuge (HERMLE Labortechnik GmbH; Siemenstr. 25, 78564 Wehingen, Germany). For each sample, 5 ml of supernatant containing the soluble fraction was transferred into three new Falcon tubes, and the pH was adjusted by adding an equal volume of 0.5 M NaOH. The reaction was stopped by adding 5 ml of NaCl. The sample was then filtered, and 2 ml of a chromogenic solution (ferric chloride III, FeCl₃) was added for spectrophotometric (Anthos Mikrosysteme GmbH; Burkamp 22, 26169 Friesoythe, Germany) reading at 492 nm. The optical densities (OD) were used to calculate the concentration of phytate in the samples as equivalents of the external calibration standard of sodium phytate, using a linear regression calibration standard prepared from serial dilutions. Controls, prepared similarly to the testing samples, but without the analyte were subtracted from the final concentration.

2.2.2 Determination of total trypsin inhibitors

Total trypsin inhibitors were extracted following the method described by [19] with slight modifications on the sample size, volume, and refrigeration during centrifugation. Powdered triplicate soybean sample (0.05 g) was weighed into clean centrifuge tubes, and homogenized with 5 ml of phosphate-buffered saline (PBS). The mixtures were vortexed and shaken on an orbital shaker for 1 h, then centrifuged at 10,000 rpm for 10 min at 4 °C using a HERMLE 300 K centrifuge (HERMLE Labortechnik GmbH; Siemenstr. 25, 78564 Wehingen, Germany). A 0.1 ml aliquot of supernatant was transferred into new micro-centrifuge tubes, and mixed with 0.1 ml of 1 mg/ml trypsin solution, followed by incubation at 0 °C for 10 min. Then, 0.3 ml of 2% casein substrate was added [19, 45], and the mixture was incubated in a water bath at 37 °C for 20 min. The reaction was stopped by adding 0.2 ml of 10% Trichloroacetic acid (TCA), and the mixture was centrifuged at 10,000 rpm for 5 min to remove undigested casein, larger inhibitor fragments, and enzyme. All steps were performed in triplicate. Blank 1 was a control without any inhibitor addition, and Blank 2 was a control with casein only. After reading the absorbance at 410 nm using a spectrophotometer, trypsin inhibition activity (TIA) from the OD readings was calculated based on the following formula [19] and converted into mg kg⁻¹ to obtain the final concentration:

$$\% \text{ inhibition} = \frac{\text{test sample} - \text{blank2}}{\text{blank1} - \text{blank2}} \times 100$$

In both methods for the determination of phytate and trypsin inhibitors, were validated by running calibration curves using standard solutions of sodium phytate and trypsin inhibitors, ensuring linearity ($R^2 > 0.98$) and sensitivity across expected concentration ranges.

2.3 Genotypic data

Genotypic data was obtained using Genotyping-by-Sequencing (GBS) through the DArTseq™ technology. This method was selected for its high throughput, cost-effectiveness, and ability to provide both SilicoDART and SNP markers, enabling comprehensive genotypic characterization [46]. Alternative techniques, such as SSR and GBS, were considered but lacked the required marker density. In summary, DNA extraction was performed using the Nucleomag Plant Kit, yielding genomic DNA concentrations of 50–100 ng/μL. DNA quality was verified using 0.8% agarose gel [47]. The DArTseq complexity reduction method involved digesting genomic DNA with PstI and MseI enzymes, followed by adapter ligation and PCR amplification. Sequencing was conducted on the Illumina HiSeq 2500 platform, with 77-base single-read runs. DArTseq markers, including SilicoDART and SNP markers, were scored using DArTsoft14, an in-house marker scoring pipeline. Both marker types were binary scored (1 = presence, 0 = absence), and aligned to the soybean Wm82.a4.v1 reference genome to determine chromosome positions [46, 48, 49].

2.4 Morphometric characterization

Soybean seeds dried to a 10% moisture content were used to determine plot yield. Yield was measured by weighing the seeds from each plot for every genotype (in grams). Yield was considered a secondary trait and was used solely for calculating the linear selection index (SLI), alongside the primary traits including phytate and trypsin inhibitors. This approach aligns with the main objective of the study, which was to assess the variability of those anti-nutritional factors [50].

2.5 Data analysis

The *lme4* package in R was used to generate best linear unbiased predictors (BLUPs), which were then used to rank phytate, TTI and preliminary yield-related parameters. BLUPs were computed considering genotype and block as random effects as follow:

$Y_{ij} = \mu + G_i + B_j + G : B_{ij} + \varepsilon_{ij}$; where: Y_{ij} = phenotypic observation for a trait, μ = grand mean, B = random effect of block (j), G = random effect of genotype (i), $G:B$ = interaction effect between genotype (i), and ε = random residual term. The resulting BLUPs were then used for computation of weighted linear selection index (LSI) with weights of 0.5, 0.2 and 0.3 for yield, phytate, and trypsin inhibitors, respectively. These weights were assigned based on the agro-economic importance of yield, and the greater nutritional emphasis on proteins (affected by trypsin inhibitors), compared to minerals impacted by phytate in soybean. Pearson correlation analysis was conducted using the residuals versus fitted values for both phytate and trypsin inhibitors.

2.5.1 Phenotypic and genotypic clustering

Clustering of phenotypic and genotypic data was conducted using the *cluster* package in R. The optimal number of clusters was determined based on the silhouette criterion. Hierarchical clustering was performed using a hierarchical clustering algorithm, Euclidian distance metric, and the *ward.D2* method, which was used to plot the dendrogram. Principal component analysis (PCA) of phenotypic data was carried out using the *factoextra* package in R. The structure analysis of genotypic data was performed using the *ade4* package in R. Discriminant analysis of principal components (DAPC) was conducted to assess the genetic structure and group differentiation within the dataset.

2.5.2 Estimation of genetic parameters

Phenotypic and genotypic coefficients of variation (PCV and GCV) were estimated using R software, following the formula proposed by Johnson et al. (1955): phenotypic coefficients of variation (PCV) $PVC = \frac{\sigma_P}{\bar{x}} \times 100$ and genotypic coefficients of variation (GCV); $GCV = \frac{\sigma_g}{\bar{x}} \times 100$; where σ_P = phenotypic standard deviation ($\sigma_g + \sigma_e$), σ_g = genotypic standard deviation, σ_e = environmental standard deviation and $\bar{x}$ = mean for phytate and TTI. Broad-sense heritability (H^2) for yield, phytate, and trypsin was calculated using best linear unbiased predictors as follow: $H^2 = V_g / (V_g + V_e)$; where: H^2 is the broad sense heritability, V_g is the genotypic variance, and V_e is the error (residual) variance. Genetic advance in absolute unit (GA) and genetic advance percent of the mean (GAM), assuming selection of the superior 5% of the genotypes, was estimated as: $GA = k \times \sigma_P \times H^2$ and $GAM = GA / \bar{x} \times 100$; where k = the standardized selection differential at 5% selection

intensity ($k=2.063$), σ_p = phenotypic standard deviation, H^2 = Heritability and $\bar{x}$ = Grand mean of the trait [51]. Admixture analysis was performed using *snmf* function in LEA R package with $k=1:10$ and 10 repetitions [52].

3 Results

3.1 Variability for Phytate, total trypsin inhibitors, and yield related parameters

There was a highly significant difference among genotypes in respect to phytate, total trypsin inhibitors and yield-related parameters ($p < 0.001$). The mean value for phytate was $1756.9 \text{ mg kg}^{-1}$ (min. 14.8 mg kg^{-1} for BSPS 48A-6-3 to $6928.8 \text{ mg kg}^{-1}$ for NGDT 2.15-7), while for total trypsin inhibitors the mean value was 850.3 mg kg^{-1} (min. 14.7 mg kg^{-1} for DN 16_N and max. $1534.8 \text{ mg kg}^{-1}$ for Duiker). The broad sense heritability (H^2) was 0.68 for phytate, 0.84 for TTI and yield, 0.08. Genotypes sourced from Japan exhibited the lowest levels of phytate, while genotypes from Nigeria and Taiwan showed low levels of trypsin inhibitors (Fig. 1A). The top 10 genotypes with the highest linear selection index (LSI) scores, in descending order, are: Ux 990-044, Ux 990-102, S 6.22B, AGS 329, Ux 990-63, S 13.2B, NGDT 8.10-9, Ux 990-062, NGDT 4.11-4, and G 457021-26-3-1 (Table 1).

Yield averaged per plot was 674.5 g, with the highest yield being 1800 g and the lowest being 77.0 g. Residuals are clustered linearly for phytate and trypsin inhibitors (Fig. 1-B). Correlation analysis among the BLUPs versus observed values showed strong positive linear relationships, r^2 (phytate): 0.97; r^2 (TTI) = 0.98 and r^2 (yield): 0.98.

3.2 Cluster analysis among soybean germplasm

The cluster analysis showed different patterns within the population. The optimal number of clusters determined by the Silhouette method indicated three clusters for the phenotypic (Fig. 2A) and four for genotypic data (Fig. 2B). The resulting

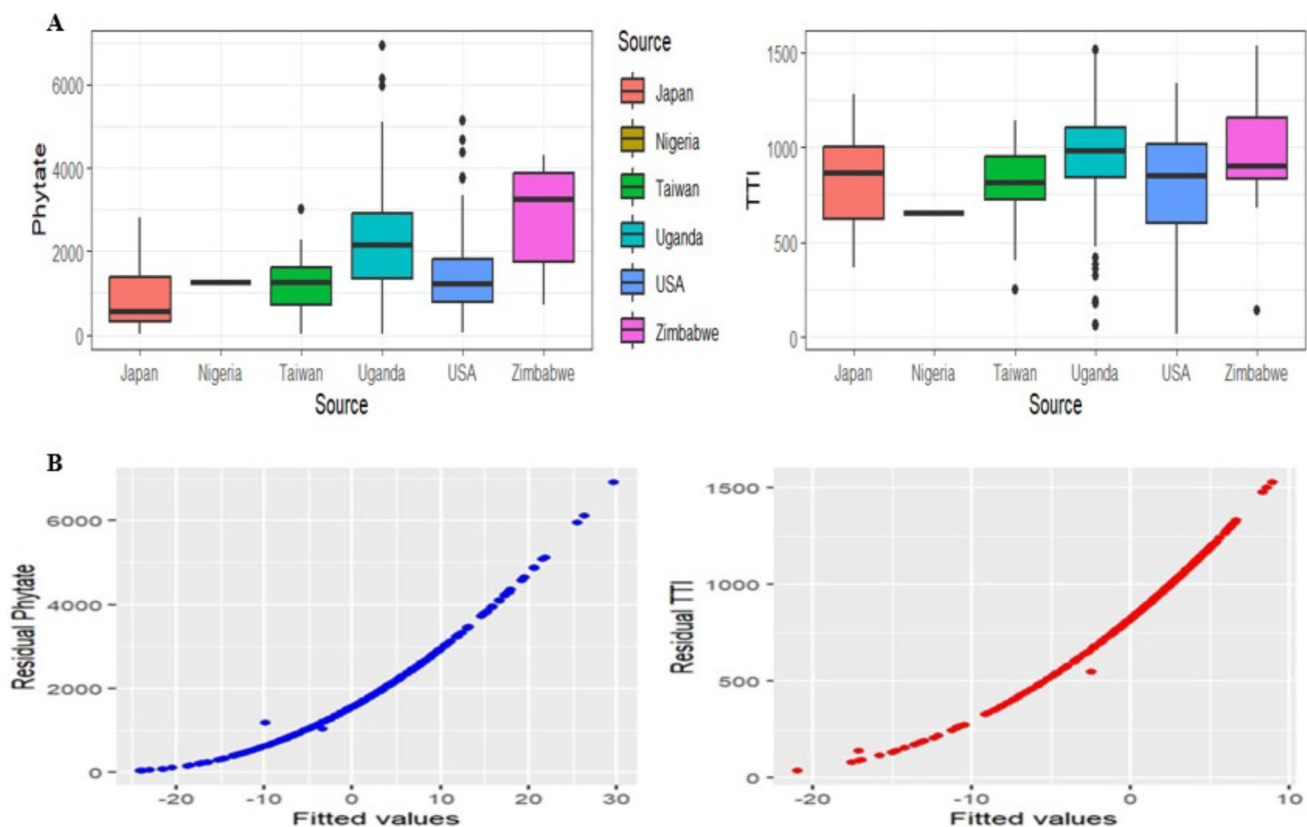
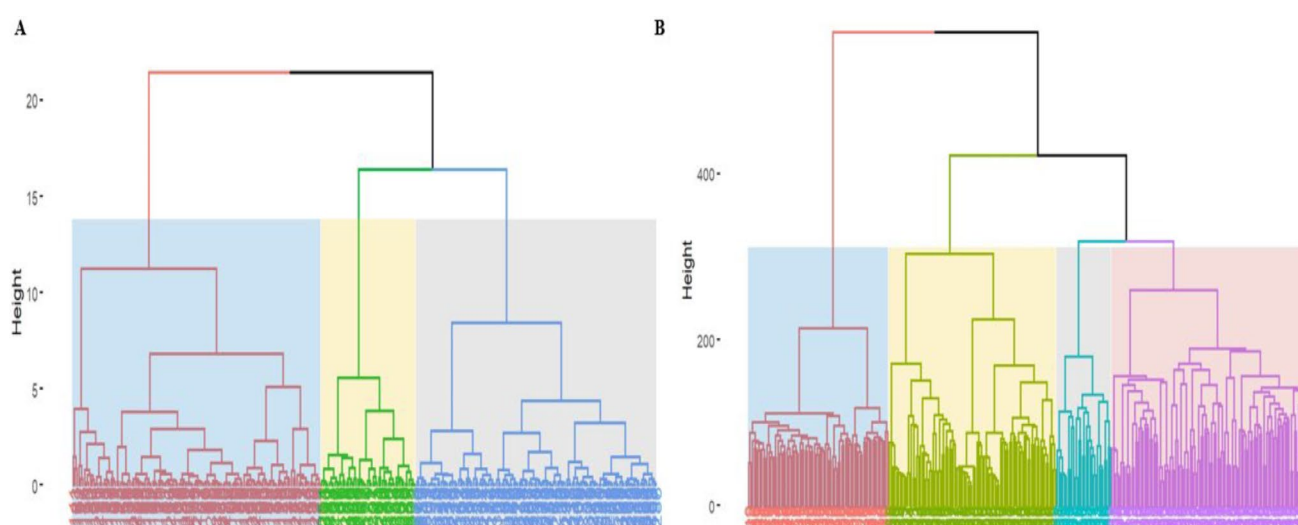


Fig. 1 Levels of phytate, trypsin inhibitors (A), and residuals versus fitted plots for of phytate and TTI (B). Genotypes from Japan showed the lowest levels of phytate, while those from Nigeria and Taiwan had the lowest TTI

Table 1 BLUPs of the top 10 lower and higher levels of phytate and trypsin inhibitors

Lower Phytate (Genotype)	Higher Phytate (Genotype)	Lower TTI ^a (Genotype)	Higher TTI (Genotype)
–24.14 (BSPS 48A–6–3)	29.65 (NGDT 2.15–7)	–20.91 (DN 16_N)	8.92 (Duiker)
–23.75 (G 84037–31–1–23)	26.34 (NGDT 8.10–9)	–17.48 (Maksoy 2N * 5N-1)	8.64 (NGDT 15.16–3)
–23.02 (S 11.1B)	25.58 (NGDT 6.10–3)	–17.10 (BSPS 48A– 27–6)	8.36 (Wonder * 4 M)
–21.58 (Pan 33)	21.82 (Ux 990–100)	–17.04 (Pan 3)	6.72 (Ux 990 – 072)
–21.42 (Pan 1)	21.62 (BSPS 48A-16)	–16.91 (DN 22)	6.59 (6N * Signal)
–20.41 (Pan 9)	20.67 (BSPS 48A-17)	–15.74 (Ux 990-079A)	6.46 (DN 12)
–18.75 (S 13.2B)	20.67 (BSPS 48A-24)	–15.00 (Pan 34)	6.29 (BSPS 48A–28–1)
–18.46 (DN 8)	20.59 (BSPS 48A-30)	–14.72 (Ux 990–091)	6.25 (NGDT 2.13–3)
–17.45 (S 7.11A)	19.49 (Pan 22)	–14.18 (Saga)	6.075 (NGDT 3.15–10)
–17.29 (BSPS 48A-38)	19.18 (NGDT 2.15-2A)	–13.55 (DN 5)	6.07 (S 11.1A)

^a TTI total trypsin inhibitors**Fig. 2** Dendrogram of phenotypic (A) and genotypic (B) data showing clustering pattern of the soybean genotypes. Based on the silhouette method, the phenotypic and genotypic data have three and four clusters, respectively

dendrogram visualized the hierarchical relationships between the clusters. Regardless of their source, the population is divided into three clusters. In cluster 1, there are 50 individuals in total (16.23%). The breakdown by source is as follows: 5 individuals from Japan, 3 from Taiwan, 30 from the USA, 11 from Uganda, and 1 from Zimbabwe. In cluster 2, there are 127 individuals (41.23%). Specifically, 14 are from Japan, 1 from Nigeria, 19 from Taiwan, 67 from the USA, 25 from Uganda, and 1 from Zimbabwe. Cluster 3 has the largest number of individuals, with 131 in total (42.53%), but the fewest sources represented. It includes 5 individuals from Taiwan, 35 from the USA, 81 from Uganda, and 10 from Zimbabwe. In clusters 1 and 2, the majority of the genotypes are from the USA, while in cluster 3, the majority are from Uganda (Table 2). For the genotypic data, four clusters were identified. In cluster 1, there are 77 individuals, accounting for 25.25% of the total. Cluster 2 contains 106 individuals, representing 34.75%. Cluster 3 comprises 92 individuals, or 30.16%, while cluster 4 includes 30 individuals, making up 9.84% of the total. Both the phenotypic and genotypic data indicate that clusters 3 and 2 comprise the largest number of individuals. The principal component analysis using genotypic data has revealed clustering pattern into distinct groups and some admixture genotypes (Fig. 3).

3.3 Structure analysis of genotypic data

The discriminant analysis of principal components (DAPC) identified four distinct groups within the dataset with overall classification accuracy of 98.05%. In the first and fourth populations, 100% of individuals were correctly

Table 2 Statistics for the individuals per cluster and per source

Cluster	Ind_cluster	%_cl	Mean_Phyt_cl	SD_Phyt_cl	SD_Phy_cl	Mean_TTI_cl	SD_TTI.cl	Source	Ind_source	%_sour	Mean_Phy_sour	SD_Phy_sour	Mean_TTI_sour	SD_TTI_sour
1	50	16.23	1331.42	665.77	338.54	172.90	172.90	Japan	5	100.00	2280.08	454.07	482.05	90.60
								Taiwan	3		636.05	489.62	350.43	84.59
								USA	30		1242.24	608.48	322.22	176.76
								Uganda	11		1311.75	585.61	332.67	189.77
								Zimbabwe	1		1566.21	NA	139.20	NA
2	127	41.23	888.11	464.86	888.75	182.83	182.83	Japan	14		485.91	323.98	957.35	172.96
								Nigeria	1		1249.50	NA	648.35	NA
								Taiwan	19		1058.17	462.28	840.76	143.48
								USA	67		902.99	472.03	877.54	181.56
								Uganda	25		936.13	421.72	914.14	203.02
3	131	42.53	2758.88	1090.43	1013.45	158.35	158.35	Zimbabwe	1		729.50	NA	1197.80	NA
								Taiwan	5		2221.45	520.92	934.80	93.30
								USA	35		2503.31	1003.25	998.22	121.75
								Uganda	81		2848.28	1142.00	1028.76	161.46
								Zimbabwe	10		3197.89	979.30	982.05	251.09

ind_cluster – total number of individuals in each cluster; %_cl – percentage of individuals in the respective cluster relative to the total dataset; Mean_Phy.cl – mean value of phytate for the cluster as a whole; SD_Phy.cl – the standard deviation of the phytate trait within the cluster; Mean_TTI.cl – mean value of the TTI trait for the cluster as a whole; SD_TTI.cl – standard deviation of the TTI trait within the cluster. Ind_source – number of individuals from each source within the cluster; %_sour – percentage of individuals from the respective source within that cluster; Mean_Phy_sour: mean phytate value for the individuals from a specific source in the cluster; SD_Phy_sour – standard deviation of the phytate value for the source within the cluster; Mean_TTI_sour: mean TTI value for the individuals from a specific source in the cluster and SD_TTI_sour – standard deviation of the TTI value for the source within the cluster

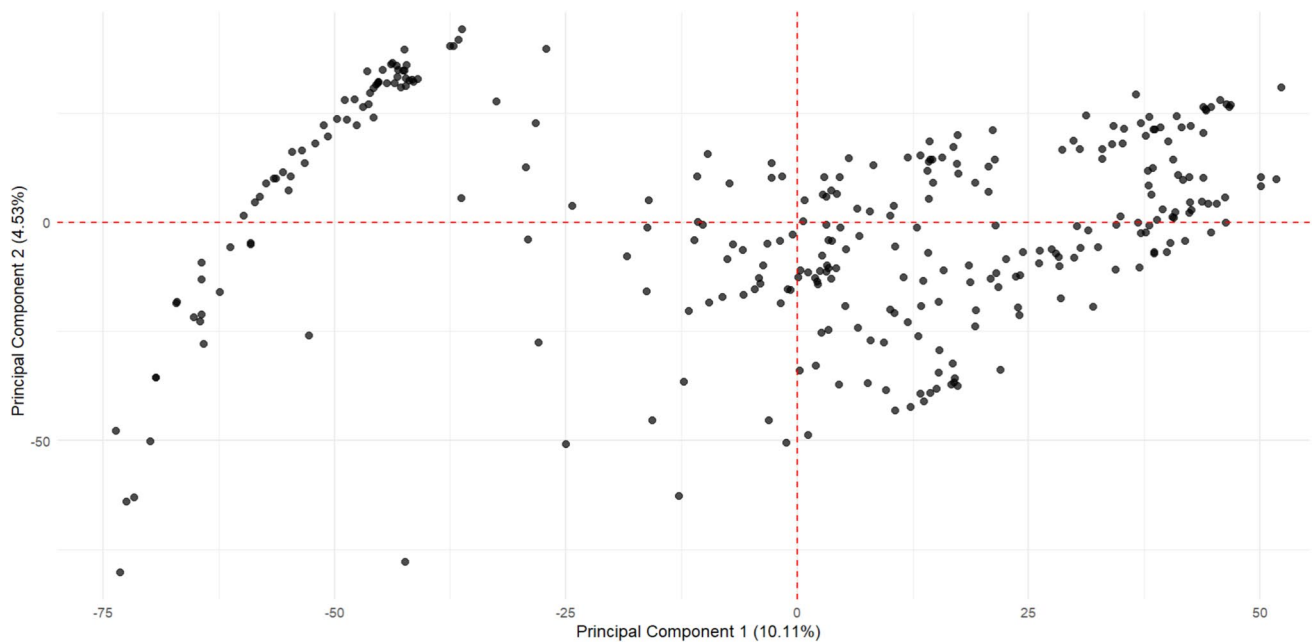


Fig. 3 Principal component analysis of the phenotypic data. The traits analyzed are spread out along the four quadrants, and the two principal components accounts for up to approximately 15% of total variation in the data

classified. In the second, third and fourth populations, 96.67%, 96.67%, and 82.61% of individuals were correctly classified, respectively, with a small margin of misclassification. The first cluster shows the highest diversity based on source representation, while the other clusters are more homogeneous, particularly clusters 2 and 4, which are dominated by single sources. Regarding cluster association, cluster 1 shows a mix of sources, with USA, Uganda, Taiwan and Japan being the most predominant. Cluster 2 and 3 are strongly associated with the USA, with minimal representation from Uganda. Cluster 4 is dominated by Uganda, with only a small contribution from the USA and Zimbabwe. The clustering appears to reflect some geographical grouping, with clusters showing dominance by specific countries such as USA in clusters 2 and 3, Uganda in cluster 4 (Fig. 4).

3.4 Genetic parameters analysis for phytate and total trypsin inhibitors (TTI)

68% of the observed variation in phytate concentration is due to genetic differences among the genotypes while 84% of the variation in TTI is attributed to genetic factors (Table 3). The GCV of 137.58% for phytate indicates a high level of genetic variability and, thus a greater genetic diversity for the trait. While for TTI, the GCV of 29.21% shows a moderate level of genetic variability for the trait. A large difference between PCV and GCV was found (Table 3).

3.5 Genetic composition of each individual

The ancestry matrix shows individuals in six groups with most of the lines having multiple ancestral populations. The majority are derived from multiple ancestral populations (multi-color bars) representing admixture genotypes, or genetic contributions from different ancestral populations (Fig. 5). Thus, the population can be described as admixed, as most genotypes show multiple colors, representing contributions from different ancestral populations.

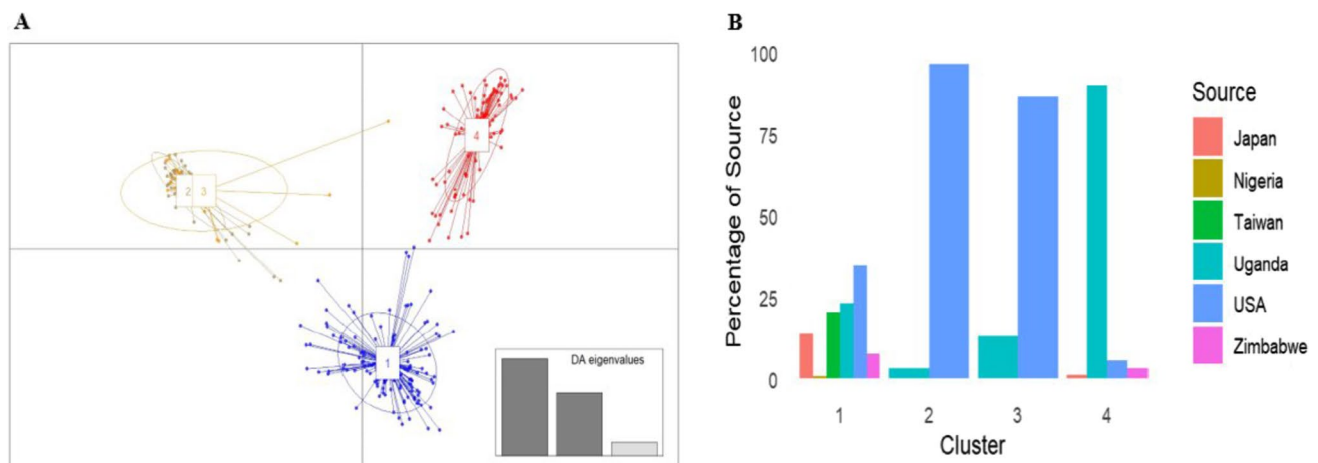


Fig. 4 Discriminant analysis of principal components (DAPC). Figure A represents the structure of the population and B, the distribution of source by cluster. There are 4 distinct groups. Better separation of clusters is seen between clusters 1 and 4 indicating strong genetic differentiation. The overlap between clusters 2 and 3 suggests an admixture or less differentiation

Table 3 Estimates of variance components for phytate and total trypsin inhibitors of the soybean genotypes

Trait	Min	Max	Mean SD	H^2	GCV	PCV	GA	GAM	SD	CV (%)
Phytate (mg.kg ⁻¹)	14.8	6928.8	1756.9	0.68	137.58	203.07	1667.6	94.92	11.7	0.68
TTI (mg.kg ⁻¹)	14.7	1538.5	850.3	0.84	29.21	34.62	504.32	59.33	5.4	0.34
Yield (g)	74	1800	675.18	0.08	3.92	47.25	7.8	1.16	1.98	0.53

H^2 broad-sense heritability, GCV genetic coefficient of variation, PCV phenotypic coefficient of variation, GA genetic advance at 5% selection intensity ($k=2.063$); GAM genetic advance as a function as percent of mean, SD standard variation, CV coefficient of variation

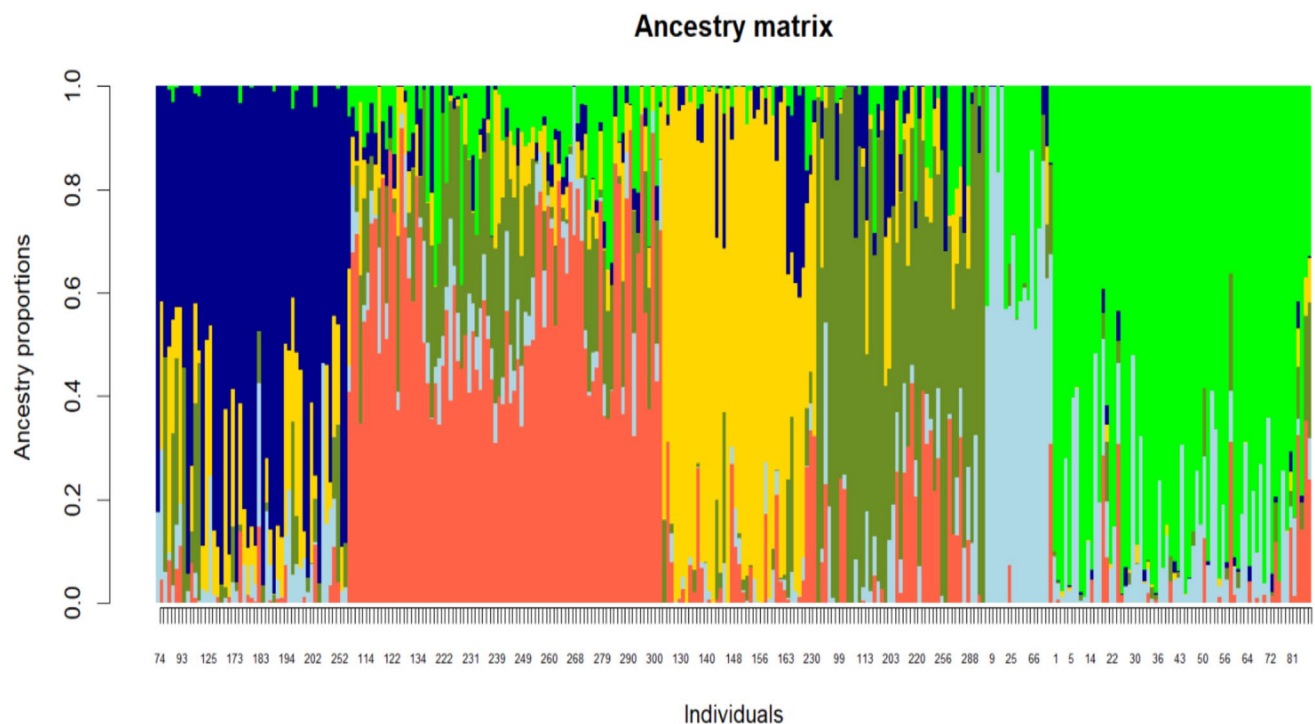


Fig. 5 Ancestry coefficients obtained from Sparse non-negative matrix factorization (SNMF). Individuals are split into multiple segments, indicating that they have significant genetic contributions from several populations

4 Discussion

In this study, we sought to understand the variability of anti-nutritional factors including phytate and trypsin inhibitors in Ugandan germplasm. Highly significant variation regarding phytate, trypsin inhibitors, and yield levels ($p < 0.001$) was found among the genotypes, indicating a promising opportunity for soybean breeders to select genotypes based on phytate and trypsin inhibitor levels and high yield potential as parents in breeding programs to develop new cultivars as desired. The evaluated genotypes have lower phytate levels than the reported range of 10,000–22,200 mg kg⁻¹ [53]. Conversely, phytic acid phosphorus levels of 3,770 mg kg⁻¹ to 5070 mg kg⁻¹ were reported in soybean mutants and natural breeding lines using HPLC method [44]. In selected soybean genotypes classified as low phytate lines the maximum concentration found was 962 mg kg⁻¹ [54]. The total trypsin inhibitors in the evaluated genotypes are lower than the 20 to 540 mg kg⁻¹ reported in different soybean products using processed soybean products [55]. On the contrary, higher levels of trypsin inhibitors (1,200 to 3,540 mg kg⁻¹) were found in soybean meals using three heat treatment processes and spectrometric method [19]. Kunitz trypsin inhibitors (KTI), a fraction of trypsin inhibitors, was reported to range from 800 mg kg⁻¹ to 8020 mg kg⁻¹ using HPLC method when developing and validating markers linked to low concentrations of KTI [14]. Another study reported levels of the two isoforms of KTI and BBI (Bowman-Birk Inhibitor) ranging from 5,250 to 1,4600 mg kg⁻¹ and 1,810 to 5,740 mg kg⁻¹, respectively, using sodium dodecyl sulfate polyacrylamide gel electrophoresis [56]. Significant differences were similarly found in evaluation of soybean genotypes for yield and other economically important traits [57, 58]. The observed variability among the genotypes might be attributed to their genetic background and geographical origins.

The higher broad-sense heritability for TTI compared to phytate and yield indicate that the phenotypic variance observed for TTI is mostly due to genetic factors rather than environmental factors. These genetic component include additive, dominance, and epistatic effects. Total trypsin inhibitors and phytate are likely to be more predictable and consistent across different environments compared to yield. Though the selection based on phytate can be still effective, the influence of environmental factors is more pronounced than TTI. Thus, selecting individuals for improvement based on the levels of TTI will likely result in a faster and more reliable genetic gain compared to phytate as it has moderate genetic control. Among the three traits, yield was found to be largely controlled by environmental factors, making selection for improvement based on this trait less effective or slower. Conversely, higher heritability of 0.97 was reported for phytate [59], while another study reported heritability of ~0.7 for phosphorus [60]. Phytate is known to be the main storage form of seed phosphorus in living systems [60]. Thus, phosphorus is often used as an indirect measure of phytate and vice versa. Phosphorus has been found to have significant positive correlation with phytate in legume seeds [61]. Lower heritability of 0.54 has been reported for KTI [14]. While low broad-sense heritability was found in this study, a higher heritability for yield of 0.66 has been reported [62]. Overall, emphasis should be given to TTI, as it offers the most genetic gain potential through selection. Phytate also is promising for improvement, though requires environment management. For yield, breeding alone might not be sufficient to achieve significant gains. Other measures such as environmental management and innovative breeding techniques are required to address the low heritability. Regarding the environmental factors influencing these traits, it has been reported that phytate levels in soybeans can be influenced by availability of potassium, phosphorus, hydrolyzable nitrogen, and sulfur content in the soil, diurnal temperature range during blooming and podding, and mean temperature during seed filling and maturity [63]. Additionally, growing location has been found to significantly impact phytate content [64]. Drained paddy fields versus upland fields, resulted in different phytate contents within the same soybean variety [65]. This highlight the complex interaction between genetic and environmental factors in determining soybean phytate levels. On the other hand, the smaller difference between phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) for TTI indicates that environmental influence is less significant compared to phytate. Consistently, a study has reported no significant difference between locations, and the parent-by-location interaction was not significant either [14]. The low heritability ($H^2 = 0.08$) for yield suggests limited genetic control, emphasizing the need to prioritize traits like phytate and TTI in early generations while improving yield through hybrid vigor or better agronomic practices.

The GCV observed for TTI is relatively moderate, indicating lower level of genetic variation. This suggests that the potential for maximizing genetic gain through selection may be limited compared to phytate, which exhibits lower heritability but has a high GCV. The high GCV for phytate suggests considerable genetic diversity among individuals, offering a greater opportunity for improvement in terms of anti-nutrient level reduction. Considering both heritability and GCV, selection based on TTI can be effective, but the moderate genetic variability among individuals may restrict the extent of anti-nutrient level reduction. In contrast, phytate, with its high GCV, presents a more favorable option for prioritization in breeding programs, despite its

moderate heritability. In fact, the high GA for phytate reported in this study indicates an expected improvement in terms of levels reduction of $1667.6 \text{ mg kg}^{-1}$ in the next generation from the selected genotypes with low phytate levels. On the other hand, the expected improvement in terms of reduction for TTI is about $504.32 \text{ mg kg}^{-1}$. Both traits indicate a considerable reduction in phytate and total trypsin inhibitor content compared to the current population. In contrast to these findings, lower genetic gains (GA) have been reported for morphometric traits and heritability in soybean [66]. The reduction in levels of ANFs in subsequent generations will be more effective for phytate than for total trypsin inhibitors (TTI), as indicated by their genetic gain averages (GAM) of 93% and 59%, respectively. Different from what was found, GAM for phytate was lower (23.43%) in a study targeting nutritional potential in *Vigna angularis* [67]. Breeding should focus on traits with high GCV, like phytate while considering environmental adjustments to manage traits with low heritability such as yield. This strategy aligns with reported findings on the importance of balanced selection indices [60].

The first principal component analysis captured up to around 15% of the total variation in the dataset. Conversely, higher PC1 and PC2 explained the variance in a previous study [68]. The low variance explained by PCA (15%) reflects the complex nature of the dataset, possibly due to high environmental influence, as indicated by the broad PCV-GCV gap. The higher variance captured in this study might help breeders to easily select genotypes with desirable levels of the anti-nutritional factors as parental lines. The power shown by the discriminant analysis of principal components (98.05%) along with the principal components (PCs) hints potential for optimization of breeding programs by targeting distinct subpopulations. On the other hand, the extent of overlapping between the clusters might indicate an interplay of genetic material from different sources, opening a window for complementary studies.

Based on genotypic data, the genotypes are grouped into three main clusters, different from the two groups reported based on principal coordinate analysis [69]. The variability among the genotypes reflects their diverse backgrounds, which can be exploited for informed decisions in breeding programs. These results provide the phenotypic and genotypic background of Ugandan soybean germplasm that can be exploited for development of cultivars with desired level of phytate and trypsin inhibitors. Based on the selection index, top genotypes with low ANFs against yield are provided along with the genetic parameters for assessment of genetic gain. The population in this study is structured, and exhibit multiple ancestries, indicating a historical interbreeding between populations and that gene flow from various ancestral sources has contributed to the actual population's genetic diversity.

Insights from this study provide a valuable foundation for soybean improvement in Uganda. By identifying genotypes with lower levels of phytate and total trypsin inhibitors (TTI), breeders can target these traits to enhance the nutritional value of soybean for human and animal consumption. For example, genotypes with reduced phytate could enhance mineral bioavailability, reducing incidences of iron and zinc deficiencies in developing countries. Similarly, low TTI levels can enhance protein digestibility, crucial for malnutrition-prone populations [53]. The variability observed in anti-nutritional factors (ANFs) highlights the genetic diversity within Ugandan soybean germplasm, which can be leveraged for molecular marker development. Molecular markers linked to low phytate and TTI levels can accelerate breeding programs by enabling precise selection of desirable traits, reducing dependency on costly post-harvest processing methods.

Furthermore, understanding the genetic and environmental components affecting ANFs can guide breeding strategies to balance nutritional improvements with yield stability. For example, traits like TTI, with high heritability, offer greater predictability for selection, while traits influenced more by environmental factors, such as phytate, may require integrated breeding and agronomic management approaches. The potential to reduce ANFs aligns with global efforts to enhance the nutritional security of soybean-dependent populations [14, 53].

This study also underscores the importance of leveraging advanced genomic tools, such as Genotyping-by-Sequencing (GBS), to identify quantitative trait loci (QTL) associated with ANFs. Such insights could facilitate the adoption of marker-assisted selection (MAS) and genomic selection (GS) in soybean breeding. Future research should focus on multi-environment trials to validate these findings and further dissect genotype-by-environment interactions impacting ANFs.

5 Conclusion

This study highlights the extent of variability of ANFs including phytate and TTI in the Ugandan soybean germplasm. The observed variability in the population provides a valuable opportunity for genotype selection reconciling agro-economic traits of interest and levels of the anti-nutritional factors studied. The results might help in accelerating breeding for low anti-nutritional factors, enhance soybean nutritional value, and reduce processing costs. Furthermore, the genetic parameters discussed may provide insights into the directions for the efforts aiming to reduction the the ANFs evaluated. The significant genetic variability and heritability of the ANFs within the population suggest a strong potential for

improvement through selection, making this an important pathway for breeding advancements. Thus, future efforts should focus on testing these genotypes across environments. However, traits like phytate and TTI can be environment-dependent. Since the study was conducted at one location (MUARIK), suggest future multi-environment trials to validate findings.

Acknowledgements Acknowledgments to the Partnership for Applied Skills in Sciences, Engineering and Technology-Regional Scholarship and Innovation Fund (PASET-RSIF) and Carnegie Corporation of New York for funding this work. Gratitude to Makerere Regional Centre for Crops Improvement (MaRCCI) for providing the facilities, to Makerere University Centre for Soybean Improvement and Development (MAKCSID) staff for providing soybean materials used in this study. Extended gratitude goes to the National Agricultural Research Organization (NARO), and the technicians at the National Crops Resources Research Institute (NaCRRI) for their valuable support.

Author contributions Conceptualization: N.J.P, T.O and P.T; Data curation: N.J.P and E.W; Formal analysis: N.J.P, E.W and S.V.K; Funding acquisition: N.J.P; Investigation: N.J.P; Methodology: N.J.P, T.O, E.N and E.W; Project administration: M.O.S, I.O.D and R.E; Resources: R.E; Software: N.J.P and E.W; Supervision: P.T, T.O, M.M, J.P.S and E.N; Validation: J.P.S; Visualization: N.J.P and E.W; Writing—original draft: N.J.P; Writing—review & editing: N.J.P, J.P.S, S.V.K, A.M, An.M and A.B.

Funding This work was supported by the Partnership for Applied Skills in Sciences, Engineering and Technology-Regional Scholarship and Innovation Fund (PASET-RSIF) [B8501G30218]; and Carnegie Corporation of New York.

Data availability Dataset will be made available upon request.

Declarations

Ethics approval and consent to participate The plants used in the study were curated by breeders, from whom some of the plants were sourced, while others were obtained through collaboration with Professor Phinehas Tukamuhabwa (one of the authors). These materials are maintained by the Makerere University Soybean Improvement and Development (MAKCSID) center during trials and studies. The collection of plants used in the study adhered to local and national guidelines, and additional attribution was not required. All procedures for the collection, characterization, and documentation of landrace varieties or germplasm accessions comply with the University's research ethics.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Guo B, Sun L, Jiang S, Ren H, Sun R, Wei Z, Hong H, Luan X, Wang J, Wang X, et al. Soybean genetic resources contributing to sustainable protein production. *Theor Appl Genet*. 2022;135(11):4095–121. <https://doi.org/10.1007/s00122-022-04222-9>.
2. Malle S, Morrison M, Belzile F. Identification of loci controlling mineral element concentration in soybean seeds. *BMC Plant Biol*. 2020;20(419):1–14.
3. Murithi HM, Beed F, Tukamuhabwa P, Thomma BPHJ, Joosten MHAJ. Soybean production in eastern and southern Africa and threat of yield loss due to soybean rust caused by *Phakopsora pachyrhizi*. *Plant Pathol*. 2016;65(2):176–88. <https://doi.org/10.1111/ppa.12457>.
4. Rui X, Wang M, Zhang Y, Chen X, Li L, Liu Y. Optimization of soy solid-state fermentation with selected lactic acid bacteria and the effect on the anti-nutritional components. 2017;(September 2016). <https://doi.org/10.1111/jfpp.13290>
5. Chauhan D, Kumar K, Ahmed N, Thakur P, Rizvi QUEH, Jan S, Yadav AN. Impact of soaking, germination, fermentation, and roasting treatments on nutritional, anti-nutritional, and bioactive composition of black soybean (*Glycine max* L.). *Journal of Applied Biology and Biotechnology*. 2022;10(5):186–192. <https://doi.org/10.7324/JABB.2022.100523>
6. Agarwal DK, Billore SD, Sharma AN, Dupare BU, Srivastava SK. Soybean: Introduction, Improvement, and Utilization in India-Problems and Prospects. *Agricultural Research*. 2013;2(4):293–300. <https://doi.org/10.1007/s40003-013-0088-0>.
7. do Prado FG, Pagnoncelli MGB, de Melo Pereira GV, Karp SG, Soccol CR. Fermented soy products and their potential health benefits: a review. *microorganisms*. 2022;10(8):1–24. <https://doi.org/10.3390/microorganisms10081606>
8. Singh G, Dukariya G, Kumar A. Distribution, importance and diseases of soybean and common bean: a review. *Biotechnol J Int*. 2020;2020:86–98. <https://doi.org/10.9734/bji/2020/v24i630125>.
9. Sinha K, Khare V, Scientist J, Bharti L. Review on: Antinutritional factors in vegetable crops. *The Pharma Innovation Journal*. 2017;6(12):353–8.
10. Nieto-Veloza A, Wang Z, Zhong Q, D'Souza D, Krishnan HB, Dia VP. Lunasin protease inhibitor concentrate decreases pro-inflammatory cytokines and improves histopathological markers in dextran sodium sulfate-induced ulcerative colitis. *Food Sci Human Wellness*. 2022;11(6):1508–14. <https://doi.org/10.1016/j.fshw.2022.06.008>.

11. Samtiya M, Aluko RE, Dhewa T. Plant food anti-nutritional factors and their reduction strategies : an overview. 2020;5(2):1–14.
12. Jha R, Yadav HK, Raiya R, Singh RK, Jha UC, Sathee L, Singh P, Thudi M, Singh A, Chaturvedi SK, et al. Integrated breeding approaches to enhance the nutritional quality of food legumes. *Front Plant Sci.* 2022. <https://doi.org/10.3389/fpls.2022.984700>
13. Rosa-Sibakov N, Re M, Karsma A, Laitila A, Nordlund E. Phytic Acid Reduction by Bioprocessing as a Tool to Improve the in Vitro Digestibility of Faba Bean Protein. *J Agric Food Chem.* 2018;66(40):10394–9. <https://doi.org/10.1021/acs.jafc.8b02948>.
14. Rosso ML, Shang C, Song Q, Escamilla D, Gillenwater J, Zhang B. Development of Breeder-Friendly KASP Markers for Low Concentration of Kunitz Trypsin Inhibitor in Soybean Seeds. *Int J Mol Sci.* 2021;2675(22):1–16.
15. Maria John KM, Khan F, Luthria DL, Garrett W, Natarajan S. Proteomic analysis of anti-nutritional factors (ANFs) in soybean seeds as affected by environmental and genetic factors. *Food Chem.* 2017;218:321–9. <https://doi.org/10.1016/j.foodchem.2016.09.072>.
16. Hymowitz T. GENETICS AND BREEDING OF SOYBEANS LACKING THE KUNITZ TRYPSIN INHIBITOR Theodore Hymowitz Department of Agronomy, University of Illinois Urbana, Illinois 61801. *Nutr Toxicol Sign Enzy Inhib Foods.* 1986;18:291–2.
17. Rada V, Lichovnikova M, Safarik I. The effect of soybean meal replacement with raw full-fat soybean in diets for broiler chickens. *J Appl Anim Res.* 2017;45(1):112–7. <https://doi.org/10.1080/09712119.2015.1124337>.
18. Zaminipira S, Tahghighat O. High untreated phytic-acid in the diet. May Lead To Mineral Deficiencies, Specifically, During Pregnancy - Soroush Niknamian. 2016;3(10):5765–70.
19. Anozie AN, Salami OA, Babatunde DE, Babatunde OE. Comparative evaluation of processes for production of soybean meal for poultry feed in Nigeria Evaluación comparativa de procesos para la producción de harina. *Anim Sci.* 2018;52(2):193–202.
20. Suhag R, Dhiman A, Deswal G, Thakur D, Sharanagat VS, Kumar K, Kumar V. Microwave processing: a way to reduce the anti-nutritional factors (ANFs) in food grains. *Lwt.* 2020;2021(150): 111960.
21. Zhong Y, Wang Z, Zhao Y. Impact of radio frequency, microwaving, and high hydrostatic pressure at elevated temperature on the nutritional and antinutritional components in black soybeans. *J Food Sci.* 2015;80(12):C2732–9. <https://doi.org/10.1111/1750-3841.13131>.
22. Duraiswamy A, A NMS, K SJ, Selvaraj S, J LP, Selvaraj R, K IP, Sherif SK, Thinakaran J. Genetic manipulation of anti- nutritional factors in major crops for a sustainable diet in future. *Front Plant Sci.* 2023;13(February):1–26. <https://doi.org/10.3389/fpls.2022.1070398>
23. Huang L, Xu Y. Effective reduction of antinutritional factors in soybean meal by acetic acid - catalyzed processing. *J Food Process Preserv.* 2018;42(August):1–8. <https://doi.org/10.1111/jfpp.13775>.
24. Pacheco WJ, Stark CR, Ferket PR, Brake J. Effects of trypsin inhibitor and particle size of expeller-extracted soybean meal on broiler live performance and weight of gizzard and pancreas. *Poult Sci.* 2014;93(9):2245–52. <https://doi.org/10.3382/ps.2014-03986>.
25. Wang T, Zhang H, Zhu H. CRISPR technology is revolutionizing the improvement of tomato and other fruit crops. *Horticulture Research.* 2019;6: 1. <https://doi.org/10.1038/s41438-019-0159-x>. <https://doi.org/10.1038/s41438-019-0159-x>
26. Lamichhane S, Thapa S. Advances from conventional to modern plant breeding methodologies. *Plant Breed Biotech.* 2022;2022(1):1–14.
27. Songstad DD, Petolino JF, Voytas DF, Reichert NA. Genome editing of plants. *Crit Rev Plant Sci.* 2017;36(1):1–23. <https://doi.org/10.1080/07352689.2017.1281663>.
28. Qaim M. Role of new plant breeding technologies for food security and sustainable agricultural development. *Appl Econ Perspect Policy.* 2020;42(2):129–50. <https://doi.org/10.1002/aep.13044>.
29. Khatodia S, Bhatotia K, Passricha N, Khurana SMP, Tuteja N. The CRISPR/Cas genome-editing tool: application in improvement of crops. *Front Plant Sci.* 2016;7(APR2016):1–13. <https://doi.org/10.3389/fpls.2016.00506>.
30. Cai Y, Chen L, Sun S, Wu C, Yao W, Jiang B, Han T, Hou W. CRISPR/Cas9-mediated deletion of large genomic fragments in Soybean. *Int J Mol Sci.* 2018;19: 12. <https://doi.org/10.3390/ijms19123835>
31. Bahadur B, Rajam MV, Sahijram L, Krishnamurthy KV. Plant biology and biotechnology: Volume II: Plant genomics and biotechnology. *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology.* 2015;II:1–768. <https://doi.org/10.1007/978-81-322-2283-5>.
32. El-sappah AH, Yan K, Huang Q, Islam M. Comprehensive Mechanism of Gene Silencing and Its Role in Plant Growth and Development. 2021;12(September). <https://doi.org/10.3389/fpls.2021.705249>
33. Ertop MH, Bektaş M. Enhancement of bioavailable micronutrients and reduction of antinutrients in foods with some processes antinutrients in foods. 2018;4(3):159–65. <https://doi.org/10.3153/FH18016>.
34. Gvozdanović JM, Šumić ZM, Todorović VJ. Phytic acid content in different dry bean and faba bean landraces and cultivars. *ROMANIAN AGRICULTURAL RESEARCH.* 2012;29:79–84.
35. Mudibu J, Nkongolo KKC, Kalonji-Mbuyi A. Morphovariability and agronomic characteristics of soybean accessions from the Demfile:///C:/Users/Norberto/Downloads/Documents/754-Article Text-777-1-10-20170213.pdfocratic Republic of Congo (DR-Congo) gene pool. *J Plant Breed Crop Sci.* 2011;3(11):260–268.
36. Kawuki R, Adipala E, Lamo J, Tukamuhabwa P. Responding to the soybean rust epidemic in sub-Saharan Africa: a review. *Afr Crop Sci J.* 2004;11: 4. <https://doi.org/10.4314/acsj.v11i4.27580>
37. Lukanda MM, Dramadri IO, Adjei EA, Badji A, Arusei P, Gitonga HW, Wasswa P, Edema R, Ochwo-ssemakula M, Tukamuhabwa P, et al. Genome-Wide Association Analysis for Resistance to Coniothyrium glycinis Causing Red Leaf Blotch Disease in Soybean. *Genes.* 2023;14(1271):1–23.
38. Tukamuhabwa P, Kawuki R, Nanfumba D, Obua T, Bashaasha B. Over view of impact indicators of soybean rust resistant varieties on Uganda 's economy Résumé. 2012;(September):2155–2160.
39. Obua T, Nabasiyire M, Namara M, Tusiime G, Maphosa M, Tukamuhabwa P. Yield stability of tropical soybean genotypes in selected agro-ecologies in Uganda. *South Afr J Plant Soil.* 2020;37(2):168–73. <https://doi.org/10.1080/02571862.2019.1678687>.
40. Tsindi A, Kawuki R, Tukamuhabwa P. Adaptation and stability of vegetable soybean genotypes in Uganda. *Afr Crop Sci J.* 2019;27(2):267. <https://doi.org/10.4314/acsj.v27i2.11>.
41. Clever M, Phinehas T, Mcebisi M, Shorai D, Isaac OD, Tonny O, Hellen K, Patrick R. Genetic diversity analysis among soybean genotypes using SSR markers in Uganda. *Afr J Biotech.* 2020;19(7):439–48. <https://doi.org/10.5897/ajb2020.17152>.
42. Khanande AS, Jadhav PV, Kale PB, Madavi SM, Mohari MP, Nichal SS, Sonkamble PA, Wandhare MR, Walke RD, Varghese P, et al. Genetic diversity in vegetable and grain type soybean genotypes identified using morphological descriptor and EST-SSR markers. *Vegetos.* 2016;29(3):158–73. <https://doi.org/10.5958/2229-4473.2016.00085.9>.

43. Tsindi A, Eleblu JSY, Gasura E, Mushoriwa H, Tongoona P, Danquah EY, Mwadzingeni L, Zikhali M, Ziramba E, Mabuyaye G, et al. Analysis of population structure and genetic diversity in a Southern African soybean collection based on single nucleotide polymorphism markers. *CABI Agric Biosci*. 2023;4(1):1–14. <https://doi.org/10.1186/s43170-023-00158-2>.
44. Israel DW. Genetic variability for phytic acid phosphorus and inorganic phosphorus in seeds of soybeans in maturity groups V, VI, and VII. 2006;2013(46):67–71. <https://doi.org/10.2135/cropsci2005.0076>.
45. Rackis MLKJJ, Mcghee JE, Pusk and G. Determination of Trypsin Inhibitor Activity of Soy Products: A Collaborative Analysis of an Improved Procedure. *Cereal Chemistry*. 1974;51(3):376–382. <http://www.aaccnet.org/cerealchemistry/abstracts/1974/cc1974a45.asp>
46. Baloch FS, Alsaleh A, Shahid MQ, Çiftçi V, Sáenz De Miera LE, Aasim M, Nadeem MA, Aktaş H, Özkan H, Hatipoğlu R. A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. *PLoS ONE*. 2017;12(1):1–18. <https://doi.org/10.1371/journal.pone.0167821>
47. Macherey-Nagel. Genomic DNA from plant - User manual (NucleoSpin® Plant II, -Midi, -Maxi). 2018;(July):36.
48. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol*. 2012;888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5.
49. Egea LA, Mérida-García R, Kilian A, Hernandez P, Dorado G. Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology “genotyping-by-sequencing” platform (DArTseq). *Frontiers in Genetics*. 2017;8(JUL):1–9. <https://doi.org/10.3389/fgene.2017.00098>
50. FAO. Methodology for estimation of crop area and crop yield under mixed and continuous cropping. Improving agricultural & rural statistics, global strategy. 2017;(Technical Report series GO-21–2017):88. <http://gsars.org/wp-content/uploads/2017/03/TR-15.03.2017-Methodology-for-Estimation-of-Crop-Area-and-Crop-Yield-under-Mixed-and-Continuous-Cropping.pdf>
51. Yirga M, Sileshi Y, Tesfaye A, Hailemariam and M. Ethiopian Journal of Crop Science. *Ethiopian Journal of Crop Science*. 2018;6: 3.
52. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*. 2015;6(8):925–9. <https://doi.org/10.1111/2041-210X.12382>.
53. Gupta RK, Gangoliya SS, Singh NK. Reduction of phytic acid and enhancement of bioavailable micronutrients in food grains. *J Food Sci Technol*. 2015;52(2):676–84. <https://doi.org/10.1007/s13197-013-0978-y>.
54. Spear JD, Fehr WR. Genetic Improvement of Seedling Emergence of Soybean Lines with Low Phytate. *Crop Sci*. 2007;47(August):1354–60. <https://doi.org/10.2135/cropsci2006.09.0600>.
55. Spelbrink REJ, Gerrits PJ, Mooij C, Giuseppin MLF. Quantitative Determination of Trypsin Inhibitory Activity in Complex Matrices Quantitative Determination of Trypsin Inhibitory Activity in Complex Matrices. 2011;(November). <https://doi.org/10.2174/1874256401105010042>
56. Yang Y, Chang KS. Determination of protease inhibitors, glycinin, and beta-conglycinin in soybeans and their relationships. *Food Science*. 2022;87:1082–95. <https://doi.org/10.1111/1750-3841.16054>.
57. Baig D. Evaluation of Soybean genotypes for yield and other economically important traits under rainfed condition. *Pure and Applied Biology*. 2018;7(1):1–7. <https://doi.org/10.19045/bspab.2018.70001>.
58. Suresh Rao SS, Singh VJ, Gampala S, Rangare NR, Science S. Assess Gen Variab Main Yield. 2014;4(2):69–74.
59. Scaboo AM, Pantalone VR, Walker DR, Boerma HR, West DR, Walker FR, Sams CE. Confirmation of molecular markers and agronomic traits associated with seed phytate content in two soybean RIL populations. *Crop Sci*. 2009;49(April):426–32. <https://doi.org/10.2135/cropsci2007.11.0614>.
60. Maria C, Coelho M, Cesar J, Santos P, Tsai SM, Vitorello VA. Seed phytate content and phosphorus uptake and distribution in dry bean genotypes. *Braz J Plant Physiol*. 2002;14(1):51–8.
61. Cichy KA, Caldas GV, Snapp SS, Blair MW. QTL Analysis of Seed Iron, Zinc, and Phosphorus Levels in an Andean Bean Population. *Crop Sci*. 2009;49(October):1742–50. <https://doi.org/10.2135/cropsci2008.10.0605>.
62. Ravelombola W, Qin J, Id AS, Song Q, Yuan J, Wang F, Chen P, Yan L, Feng Y, Zhao T, et al. Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS ONE*. 2021;8:1–21.
63. Li H, Wang X, Ma Y, Wen Z, Li J, Zhang H, Wu Y, Lei C, Wang S, Wang J, et al. Ecophysiological Factors on Phytic Acid Concentration in Soybean Seed. *Crop science*. 2015;53(october 2013):2195–2201. <https://doi.org/10.2135/cropsci2012.02.0068>
64. Kumar V, Rani A, Rajpal S, Srivastava G, Ramesh A, Joshi OP. Phytic acid in Indian soybean: genotypic variability and influence of growing location. *J Sci Food Agric*. 2005;85(9):1523–6. <https://doi.org/10.1002/jsfa.2151>.
65. Shiguro TI, No TO, Ada TW, Sukamoto CT, Ono YK. Changes in Soybean Phytate Content as a Result of Field Growing Conditions and Influence on Tofu Texture. *Biosci Biotechnol Biochem*. 2006;70(4):874–80.
66. Obua T. Genetic Improvement of Nutritional Traits and Yield of Tropical Soybean in Uganda. 2022:1–121.
67. Padhi SR, Gore PG, Tripathi K, Katral A, Chandora R, Abhishek GJ, Kondal V, Singh R, Bharadwaj R, et al. Nutritional potential of adzuki bean germplasm and mining nutri-dense accessions through multivariate analysis. *Foods*. 2023;12:22. <https://doi.org/10.3390/foods12224159>.
68. Obua T, Sserumaga JP, Opiyo SO, Tukamuhabwa P, Odong TL, Mutuku J, Yao N. Genetic Diversity and Population Structure Analysis of Tropical Soybean (*Glycine max* (L.) Merrill) using Single Nucleotide Polymorphic Markers. *Global Journal of Science Frontier Research*. 2020;20(6):35–43. <https://doi.org/10.34257/gjsfrdvol20is6pg35>
69. Lukanda MM, Dramadri IO, Adjei EA, Arusei P, Gitonga HW, Wasswa P, Edema R, Ssemakula MO, Tukamuhabwa P, Tusiime G. Genetic Diversity and Population Structure of Ugandan Soybean (*Glycine max* L.) Germplasm Based on DArTseq. *Plant Mol Biol Reporter*. 2023;1: 0123456789. <https://doi.org/10.1007/s11105-023-01375-9>.