Received: 9 February 2024 | Accepted: 16 August 2024 | Published online: 29 September 2024

DOI: 10.1002/csc2.21356

Crop Science

ORIGINAL ARTICLE

Crop Breeding & Genetics

Evaluating genetic diversity and seed composition stability within Pan-African Soybean Variety Trials

Elizabeth De Meyer[1] | **Elizabeth Prenger**[1,2] | **Anser Mahmood**[1] | **Michelle da Fonseca Santos**[3] | **Godfree Chigeza**[4] | **Qijian Song**[5] | **Learnmore Mwadzingeni**[6] | **Ronica Mukaro**[7] | **Mwila Chibanda**[8] | **Gorden Mabuyaye**[6] | **Brian Diers**[3] | **Andrew Scaboo**[1]

[1]College of Agriculture, Food & Natural Resources, Division of Plant Science & Technology, University of Missouri, Columbia, Missouri, USA

[2]The Center for Agroforestry, University of Missouri, Columbia, Missouri, USA

[3]USAID Feed the Future Soybean Innovation Lab, University of Illinois Urbana–Champaign, Urbana, Illinois, USA

[4]CGIAR International Institute of Tropical Agriculture, Ibadan, Oyo State, Nigeria

[5]USDA-ARS, Soybean Genomics and Improvement Lab., Beltsville Agricultural Research Center, Beltsville, Maryland, USA

[6]Seed Co Limited, Arcturus, Harare, Zimbabwe

[7]Crop Breeding Institute (CBI), Causeway, Harare, Zimbabwe

[8]Seed Co Limited, Makeni, Lusaka, Zambia

Correspondence
Andrew Scaboo, College of Agriculture, Food & Natural Resources, Division of Plant Science & Technology, University of Missouri, Columbia, MO 65211, USA.
Email: scabooa@missouri.edu

Assigned to Associate Editor Bo Zhang.

Funding information
USAID Feed the Future Soybean Innovation Lab; Missouri Soybean Merchandising Council, Grant/Award Number: Award ID #16-340

Abstract

Given high animal protein costs, protein deficiency is a prevalent form of malnutrition in sub-Saharan Africa (SSA). Soybean [*Glycine max* (L.) Merr.] can provide a cheaper high-quality protein source and fortify lysine-limited cereal-based diets. Breeding soybean for seed composition in SSA requires understanding genotype by environment interactions (G × E). African breeding programs submit cultivars for evaluation in the Pan-African Soybean Variety Trials (PATs), providing the opportunity to examine G × E across diverse environments. With PAT data, we conducted additive main effects and multiplicative interaction (AMMI) and genotype plus genotype-by-environment (GGE) biplot analyses on seed protein and oil content of 17 cultivars grown with two replications in nine environments across Zimbabwe. Across environments, protein ranged from 322.8 to 445.1 g kg^{-1} and oil ranged from 164.8 to 242.7 g kg^{-1}. For protein AMMI analysis, MAKWACHA performed

Abbreviations: AMMI, additive main effects and multiplicative interaction; ANOVA, analysis of variance; CV, coefficient of variation; G × E, genotype by environment interaction; GGE, genotype plus genotype-by-environment; GRIN, Germplasm Resources Information Network; IITA, International Institute for Tropical Agriculture; MET, multi-environment trial; MG, maturity group; NIR, near-infrared spectroscopy; OrWAAS, ranked Weighted Average of Absolute Scores; OrWAASY, ranked Weighted Average of Absolute Scores and Response Variable; PAT, Pan-African Soybean Variety Trial; PC, principal component; PCA, principal component analysis; SIL, Soybean Innovation Lab; SNP, single-nucleotide polymorphism; SSA, sub-Saharan Africa; USDA, United States Department of Agriculture; WAAS, Weighted Average of Absolute Scores; WAASY, Weighted Average of Absolute Scores and Response Variable.

best for both stability and a combination of stability plus content. For oil AMMI analysis, SC SPIKE performed best for stability and KALEYA performed best for a combination of stability plus content. GGE biplot analyses identified three different mega-environments for protein and oil, with SC EXPT2, KALEYA, and SC EXPT1 having highest protein content and TGX 2002-9FM, LUKANGA, and SC EXPT3 having highest oil content. We also evaluated genetic diversity of 19 PAT cultivars through phylogenetic analyses with 1059 USDA Germplasm Resources Information Network (GRIN) diversity panel accessions. We recommend stable and adaptable PAT cultivars to breeders and producers while highlighting genetically distinct accessions with valuable traits as a resource for breeding programs.

Plain Language Summary

Soybean is currently promoted in sub-Saharan Africa as a high-quality protein source. Breeding soybean for improved seed composition requires understanding how cultivars interact with these new environments. This study evaluated genotype by environment interactions and the genetic diversity of cultivars grown in the Pan-African Soybean Variety Trials. We used stability analyses to evaluate how the environment, cultivar, and their interaction impacted protein and oil across Zimbabwe. We identified cultivars with increased protein or oil in specific environments or stable performance across environments. We also compared the genetic diversity of the African cultivars to over 1000 genetically diverse accessions from around the world. Our analyses showed that the African cultivars have relatively low genetic diversity and are most closely related to American accessions. Our results provide valuable resources for improving soybean seed composition in sub-Saharan Africa.

1 | INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is a valuable crop worldwide, given its high protein and oil content and diverse utility across food, feed, and industrial applications (Masuda & Goldsmith, 2009; Wilson, 2004). Soybean production worldwide in 2022 totaled 348.9 million metric tons, increasing over four times from the 81 million metric tons produced in 1980 (FAOSTAT, 2024). The top producing countries in 2022 were Brazil at 120.7 million metric tons, the United States at 116.4 million metric tons, and Argentina at 43.9 million metric tons, combining to produce 80.5% of the world's total soybean (FAOSTAT, 2024). Soybean production in Africa made up 1.3% of the world's total production in 2022 (FAOSTAT, 2024). South Africa, Nigeria, and Zambia were the top three soybean-producing countries in Africa in 2022, which when combined produced 59.1% of Africa's total soybean (FAOSTAT, 2024). However, even at current yield and production levels, soybean makes up a valuable component of smallholder cropping systems in sub-Saharan Africa (SSA) (Khojely et al., 2018; Marcillo et al., 2021; Masuda & Goldsmith, 2009; Pratap et al., 2012; Siamabele, 2021; Tesfaye et al., 2017). Further, soybean has the potential to enhance the food and nutritional security of households, given its range of uses, including as a protein and cooking oil source for humans, as feed for animals, and as a source of raw material for biofuel production (Khojely et al., 2018; Marcillo et al., 2021; Masuda & Goldsmith, 2009; Pratap et al., 2012; Siamabele, 2021; Tesfaye et al., 2017). Malnutrition disproportionately impacts women, adolescents, children, and infants, causing more than a third of the child mortality in SSA (Bain et al., 2013). Due to the high cost of animal protein in SSA, protein deficiency is a prevalent form of malnutrition (Asodina et al., 2020). Soybean can provide a cheaper alternative source of high-quality protein, and due to its relatively high protein and lysine content, it can help fortify lysine-limited cereal-based diets (Asodina et al., 2020; Engelbrecht et al., 2020; Foyer et al., 2019; Hartman et al., 2011; Kalumbi et al., 2019; Mangani et al., 2015).

SSA has been a net importer of soybean products and oil, due to slow agricultural gains for smallholder farmers causing an inability to meet the needs of rapidly growing populations (Acevedo-Siaca & Goldsmith, 2020; Engelbrecht et al., 2020). Meeting increased market and consumer demand for soybean

products in SSA requires rapid genetic gains in cultivar performance and increased production (Foyer et al., 2019). Research on SSA soybean production has taken place through different organizations, including the United States Agency for International Development (USAID) Feed the Future Soybean Innovation Lab (SIL), which partly focuses on supporting soybean cultivar development programs in SSA to improve yields and other important traits (Santos, 2019). A component of SIL research has been the Pan-African Soybean Variety Trials (PATs), which are cross-continent field experiments evaluating replicated panels of 20–40 soybean cultivars at each environment (Santos, 2019). As of 2021, these panels were evaluated at over 100 locations in 24 SSA countries (Marcillo et al., 2021). These trials aim to identify the best-performing cultivars based on yield, agronomic characteristics such as maturity date and disease resistance, and seed composition (Pawlowski et al., 2021). There are currently several public and private soybean breeding programs in Africa submitting cultivars to the PATs, including the International Institute for Tropical Agriculture (IITA) with headquarters in Nigeria, Makerere University in Uganda, and Seed Co in Zimbabwe (Diers & Scaboo, 2019). Given the cultivar-specific nature of yield and seed composition and environmental influences on these traits (Assefa et al., 2019), breeders should evaluate the performance and stability of their cultivars across broad environments such as the PATs to determine the best cultivars for SSA production.

Seed protein and oil content and yield are strongly influenced by genetics, environment, and genotype by environment interaction (G × E) (Assefa et al., 2019; Bellaloui et al., 2009; Yaklich et al., 2002). Conventional soybean cultivars contain 40% protein and 20% oil on a dry-weight basis (Wilson, 2004). However, an evaluation of diverse accessions in the United States Department of Agriculture (USDA) Germplasm Resources Information Network (GRIN) Soybean Germplasm Collection found a wide range of seed composition, from 8% to 27% for oil content and from 34% to 56% for protein content (Wilson, 2004). This demonstrates the wide base of genetic diversity for these traits available in the germplasm collection. Both seed protein and oil content are complex, highly polygenic traits that show significant interactions with environmental factors (Chung et al., 2003; Lee et al., 2007; Warrington et al., 2015). The nature of these seed composition traits reinforces the need for incorporating genetic diversity and evaluating cultivars across a broad range of environments. Additionally, the negative correlation between seed protein and yield and the negative correlation between seed protein and oil make it difficult to breed a high-yielding cultivar with enhanced protein and oil content (Brzostowski & Diers, 2017). Therefore, while breeding for increased yields, seed protein and oil content should also be monitored to ensure seed quality is maintained to meet commodity market and end-user demands.

Core Ideas

- This is a comprehensive report on seed composition stability within Pan-African Soybean Variety Trials (PATs).
- Genotype by environment interaction explained a significant amount of variation in seed protein and oil content.
- Additive main effects and multiplicative interaction (AMMI) and genotype plus genotype-by-environment (GGE) biplot analyses identified PAT cultivars with the highest stability and adaptability in Zimbabwe.
- We highlight the genetic relatedness of PAT cultivars relative to USDA Germplasm Resources Information Network (GRIN) diversity panel accessions.

To evaluate yield and seed composition in the later stages of breeding pipelines, the best-performing lines are evaluated in many different environments across locations, seasons, and years. These are referred to as multi-environment trials (METs) and can inform which new lines have the most stable or best performance within or across environments while allowing for the delineation of mega-environments (Olivoto et al., 2019; Smith et al., 2015; Vaezi et al., 2019). The PATs are an example of METs, providing breeders in Africa with the resources to evaluate the performance of their lines across diverse environments. With the data collected in these METs, both the additive effects of genotype and environment and the G × E effect on genotype performance can be determined (Burgueño et al., 2011). The evaluation of late-stage breeding program lines in multiple environments is essential to ensure released cultivars perform well in their intended environments (Yau, 1995).

To fully examine G × E, more sophisticated methods are needed than analysis of variance (ANOVA) or principal component analysis (PCA) on their own (Eberhart & Russell, 1966; Finlay & Wilkinson, 1963). Commonly used methods for stability analyses include additive main effects and multiplicative interaction (AMMI) (Gauch, 1992a) and genotype plus genotype-by-environment (GGE) (Yan & Hunt, 2001). Both AMMI and GGE biplot analyses combine ANOVA and PCA to evaluate G × E (Gauch, 2006). While the AMMI PCA only includes the variance estimates from G × E, the GGE PCA includes both the variance from genotype and G × E (Gauch, 2006). AMMI results can also be easily visualized with biplots, providing stratification of genotypes by environment (Zobel et al., 1988). Overall, AMMI can help improve understanding of G × E, improve the accuracy of phenotype estimates such as yield, increase the flexibility of

experimental design, help impute missing data, and identify mega-environments (Gauch, 1992b; Gauch et al., 2008). In addition to creating AMMI biplots, stability values can be calculated to compare genotypes and environments. Weighted Average of Absolute Scores (WAAS) values are calculated from the significant principal component (PCs) ($p < 0.05$) in an AMMI analysis (Olivoto et al., 2019) and can be used to rank stability. Additionally, Weighted Average of Absolute Scores and Response Variable (WAASY) values can be used to select cultivars with both high stability and response variable values (Olivoto et al., 2019). The GGE biplot method aids in interpreting G × E in MET experiments with graphical outputs (Yan, 2001). In this method, genotype and environment information can be visually inspected through a set of biplot interpretation models (Yan & Tinker, 2006). GGE biplots can be used to identify mega-environments, which are groupings of environments with similarly ranked genotypes (Yan et al., 2023). Furthermore, GGE biplot analysis has several other applications, such as identifying the best genotype in each environment, evaluating genotypes by both stability and performance, and evaluating the representativeness and discriminating ability of environments (Pour-Aboughadareh et al., 2022).

While several stability analyses of soybean yield (Arega et al., 2018; Cheelo et al., 2017; Jandong et al., 2011, 2019; Mushoriwa et al., 2022; Mwiinga et al., 2020; Nachilima et al., 2021; Tolorunse et al., 2018) and/or soybean seed composition (Hampango et al., 2017; Mukuze et al., 2020; Popović et al., 2013) have been conducted across Africa, none have examined both protein and oil content within the PATs. Additionally, some phylogenetic analyses of breeding lines in Africa have been conducted (Chander et al., 2021; Chigeza et al., 2019; Tsindi et al., 2023). However, to our knowledge, a study has yet to examine the genetic diversity of these lines in reference to a wide range of geographic origins and relative maturity groups (MGs). Therefore, the objectives of our study were to conduct AMMI and GGE biplot stability analyses of seed protein and oil content for selected PAT cultivars grown at nine environments across Zimbabwe and to evaluate the genetic diversity of selected PAT cultivars compared to a diversity panel of USDA GRIN Soybean Germplasm Collection accessions. Our study identified PAT cultivars and mega-environments with stable and high seed protein and oil content, and we highlight the genetic relatedness of PAT cultivars relative to USDA GRIN diversity panel accessions.

2 | MATERIALS AND METHODS

2.1 | Trial locations and experimental design

The datasets used for this study included cultivars submitted to the PATs that were grown in field experiments at nine environments in Zimbabwe during the summer 2019–2020 season. Table 1 details the field experiment sites used and their characteristics; Figure 1 shows maps of the sites across Zimbabwe with elevation, soil type, average daily temperature, and cumulative precipitation; and Figure S1 shows photosynthetically active radiation across the growing season. Two replications of each cultivar were grown at each site in an Alpha (0,1) lattice design. Each plot consisted of four 5-m rows, with interrow spacing of 0.45 m and in-row plant spacing of 0.05 m. The center two rows were harvested and used for phenotyping. Phenotypic data for yield and agronomic traits are available from the "Pan-African Soybean Variety Trials (PAT) Database" (https://soybean-innovation-lab.github.io/PAT-trials/main/).

2.2 | Plant materials

Nineteen soybean cultivars submitted to the PATs were used in this study. Cultivar names, country of origin, and other characteristics of the cultivars are detailed in Table 2. These cultivars originate from four different African countries and have varying commercial status.

2.3 | Data collection for environmental traits

Maps of Zimbabwe were created using QGIS. Elevation was plotted using "1 Arc-Minute digital elevation model" (DEM) data downloaded from The Climate Data Tool (Dinku et al., 2022). Soil type data were obtained from the FAO Digital Soil Map of the World (Sanchez et al., 2009). "Temperature at 2 Meters," "Precipitation," and "All Sky Surface Photosynthetically Active Radiation (PAR) Total" were downloaded from the NASA Prediction of Worldwide Energy Resource (POWER) project for the growing season, between the first environment planting date of December 16, 2019 and the last environment harvest date of June 5, 2020. The daily temperature and daily photosynthetically active radiation data were averaged and the daily precipitation data were summed before plotting.

Elevation, soil test data, and weather data were collected from each PAT environment. Soil samples were collected following a set protocol across the PAT environments. Briefly, a soil probe was used to collect at least five to 10 soil samples down to a depth of 20 cm in an M-shaped zigzag pattern across each block in the PAT environment. Samples from each block were combined to create one composite sample for each environment, and the composite sample was airdried overnight. The composite samples from each PAT environment were sent to Brookside Laboratories, Inc. (New Bremen, Ohio, USA) for soil testing. Weather data for each PAT environment was obtained from the QuickTrials platform (https://www.quicktrials.com) using the Meteoblue add-on (https://content.meteoblue.com/). Elevation, soil pH, and soil

TABLE 1 Characteristics of nine Pan-African Soybean Variety Trial (PAT) environments used in this study.

Environment (ENV) code	Location code	Location name	Planting date	Harvest date	Latitude, longitude	Altitude (m)	Soil type	Soil pH	Soil Na (mg kg^{-1})	Average temperature (°C)	Cumulative precipitation (mm)
ENV1	CBI-Site 4	Chisumbanje	January 22, 2020	May 29, 2020	−20.47, 32.14	421	Chromic Luvisols	8.3	156	23.3	64.1
ENV2	CBI-Site 5	Panmure	January 8, 2020	June 5, 2020	−17.07, 31.11	962	Ferric Luvisols	6.5	26	19.8	231.0
ENV3	CBI-Site 2	Kadoma	January 20, 2020	June 3, 2020	−18.19, 29.52	1142	Ferric Luvisols	5.8	40	21.2	95.1
ENV4	SEEDCO-Site 6	Mutare	December 17, 2019	May 12, 2020	−18.89, 32.60	1200	Lithosols	5.9	30	21.2	326.4
ENV5	SEEDCO-Site 5	Banket	December 18, 2019	May 14, 2020	−17.64, 30.40	1260	Ferric Luvisols	6.2	26	20.8	325.8
ENV6	SEEDCO-Site 1	RARS	December 16, 2019	May 18, 2020	−17.65, 31.24	1341	Ferric Luvisols	5.3	12	20.4	369.6
ENV7	SEEDCO-Site 3	Stapleford	December 17, 2019	April 16, 2020	−17.73, 30.88	1448	Eutric Nitosols	6.1	24	20.4	340.8
ENV8	SEEDCO-Site 2	Harare (Seed Co)	December 18, 2019	May 12, 2020	−17.72, 31.08	1480	Eutric Nitosols	6.1	19	19.4	392.8
ENV9	CBI-Site 1	Harare (CBI)	December 16, 2019	May 26, 2020	−17.48, 31.03	1504	Eutric Nitosols	6.4	21	20.6	307.5

Note: All trials were grown in Zimbabwe during the summer 2019–2020 season. Soil pH and Na were obtained from PAT environment soil test data, and average temperature and cumulative precipitation were calculated from daily measurements across the growing season at each environment.

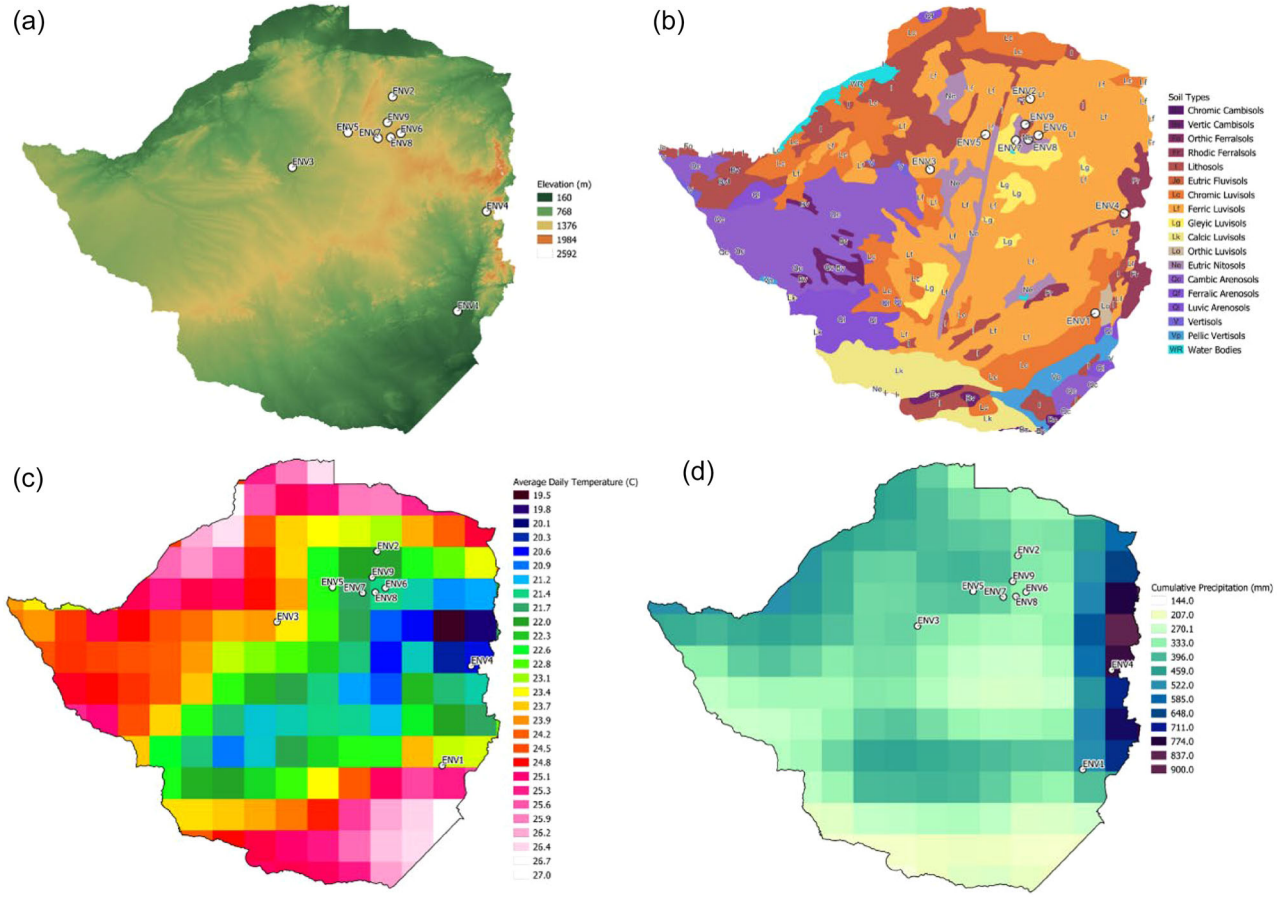

FIGURE 1 Maps of Zimbabwe with (a) elevation in meters, (b) soil type, (c) average daily temperature in degrees Celsius, and (d) cumulative precipitation in millimeters. Pan-African Soybean Variety Trial (PAT) environments are indicated with white markers, and the key for the environment codes can be found in Table 1. All maps were created using QGIS. Elevation was plotted using "1 Arc-Minute digital elevation model" (DEM) data downloaded from the Climate Data Tool (Dinku et al., 2022). Soil type data were obtained from the FAO Digital Soil Map of the World (Sanchez et al., 2009). "Temperature at 2 Meters" and "Precipitation" were downloaded from the NASA Prediction of Worldwide Energy Resource (POWER) project for the growing season, between the first environment planting date of December 16, 2019 and the last environment harvest date of June 5, 2020. The daily temperature data were averaged, and the daily precipitation data were summed before plotting.

Na for each environment were plotted against seed dry weight protein and oil content. Within the growing season dates for each environment, the daily mean temperature was averaged and the daily precipitation was summed before plotting against protein and oil content.

2.4 | Data collection for protein and oil content

Samples of 50 g of seed from each plot were analyzed at the Bay Farm Research Facility by the University of Missouri Northern Soybean Breeding Program in Columbia, Missouri, for seed composition. Seed samples were visually assessed as a measure of quality control. Hilum color and seed quality for each plot sample were noted, and samples with significant visual differences from the majority of the same cultivar were excluded from the analysis. This resulted in 19 cultivars at nine environments across Zimbabwe for further analysis. Near-infrared spectroscopy (NIR) was used to predict seed composition of whole seed samples on a Perten DA 7250 machine (PerkinElmer, Inc.). Calibrations are updated annually by the Soybean NIR Consortium, and the 2019 and 2020 models were based on a minimum of 3800 samples with known composition values. This study used the NIR calibration for 2019 whole soybeans. NIR prediction accuracies for oil and protein were high, with R^2 values ranging from 0.79 to 0.90. Moisture is reported as a percentage of the seed as is, and the remaining composition values were reported as percent on a dry matter basis and were converted to grams per kilogram before use in stability analyses.

2.5 | Genotyping and quality control

Within the dataset containing two replications of 19 cultivars at nine environments, we conducted genotyping to ensure genetic integrity of the cultivars between replications and

TABLE 2 Characteristics of 19 soybean cultivars submitted to the Pan-African Soybean Variety Trials (PATs) and used in this study.

Cultivar	Country of origin	Seed supplier	Flower color	Pubescence color	Hilum color	Maturity	Commercial status
BIMHA	Zimbabwe	CBI	Purple	Gray	Yellow	Medium	Released in Zimbabwe
KALEYA	Zambia	ZamSeed	Purple	Tawny	Brown	Late	Released in Zambia
LUKANGA	Zambia	ZamSeed	Purple	Gray	Yellow	Medium	Released in Zambia
MAKWACHA	Malawi	DARS	Purple	Tawny	Yellow	Medium	Released in Malawi
MHEMBWE	Zimbabwe	CBI	White	Tawny	Brown	Early	Released in Zimbabwe
MHOFU	Zimbabwe	CBI	White	Gray	Yellow	Medium	Released in Zimbabwe
NASOKO	Malawi	DARS	Purple	Gray	Yellow	Medium	Released in Malawi
SC EXPT1	Zimbabwe	Seed Co	White	Gray	Gray/black	Medium	Experimental
SC EXPT2	Zimbabwe	Seed Co	Purple	Tawny	Dark brown	Medium	Experimental
SC EXPT3	Zimbabwe	Seed Co	Purple	Tawny	Brown	Late	Experimental
SC SAGA	Zimbabwe	Seed Co	Purple	Gray	Buff	Medium	Released in Zimbabwe
SC SIGNAL	Zimbabwe	Seed Co	Purple	Gray	Brown	Late	Released in Zimbabwe
SC SL01	Zimbabwe	Seed Co	Purple	Gray	Imperfect black	Medium	Released in Nigeria
SC SPIKE	Zimbabwe	Seed Co	Purple	Tawny	Brown	Medium	Released in Zimbabwe and Zambia
TGX 1987-62F	Nigeria	IITA	Purple	Tawny	Brown	Early	Experimental
TGX 1991-22F	Malawi	DARS	Purple	Tawny	Brown	Early	Released in Malawi
TGX 2002-9FM	Zambia	IITA	White	Gray	Black	Medium	Released in Zimbabwe (AKA: Ngongoni)
TGX 2014-16FM	Zambia	IITA	White	Tawny	Brown	Medium	Experimental
TIKOLORE	Malawi	DARS	Purple	Tawny	Brown	Early	Released in Malawi, Mozambique, and Zambia (AKA: TGx 1740-2F, Kafue)

Note: Maturity designations were based on the relative maturity of this set of regional varieties within field evaluations.
Abbreviations: CBI, Crop Breeding Institute; DARS, Department of Agricultural Research Services; IITA, International Institute of Tropical Agriculture.

environments. We grew 15–30 seeds from each plot in the greenhouse; leaf tips from the first trifoliates of six different plants were tissue sampled for each plot; and we extracted total genomic DNA using a modified cetyltrimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987). SC SL01 and TGX 2002-9FM had consistent and unique hilum colors, so only six random plots were sampled for these cultivars compared to all 18 plots for the rest of the cultivars. DNA samples were sent to the Soybean Genomics and Improvement Laboratory, USDA-ARS, for genotyping with the Illumina Infinium BARCSoySNP3K BeadChip, which contains a subset of the BARCSoySNP6K BeadChip (Song et al., 2020). Single-nucleotide polymorphism (SNP) alleles were called using GenomeStudio Genotyping Module 2.0 (Illumina, Inc.). We conducted genetic similarity analysis to quality control all field plots associated with each of the 19 cultivars. Genetic similarity between the cultivars using the 3K SNP data was determined using the Distance Matrix function in TASSEL (Bradbury et al., 2007). The distance matrix was visualized using the "superheat" package in R (Barter & Yu, 2018), and plots with low genetic similarity compared to the rest of the plots for the same cultivar were discarded. This was done to ensure that the genetic integrity of all experimental data was of the highest quality. This approach did not result in the discarding of cultivars; rather, it resulted in the discarding of plot data that were clearly compromised. To further identify outliers and quality control the data, ANOVA for each environment was completed using the "aov" function in R with genotype, replication, and block nested within replication as fixed effects for percent protein dry weight (Tables S1–S9) and percent oil dry weight (Tables S10–S18). Three times the overall coefficient of variation (CV) for each ANOVA was used as a threshold for quality control of replications. Out of the total 342 plots, 36 plots were discarded based on low within-cultivar relative genetic similarity and outlying protein and/or oil CV values.

2.6 | Stability analyses

Since balanced data allow for unbiased estimates of interaction effects, two cultivars with both replications missing

from at least one environment after quality control were removed from downstream analyses. This resulted in 17 cultivars grown at nine environments across Zimbabwe. To evaluate G × E, stability analyses were conducted for protein and oil content in grams per kilogram using the "metan" package in R (Olivoto & Lúcio, 2020). AMMI analyses (Gauch, 1992b) were conducted on protein and oil data to obtain AMMI tables, WAAS (Olivoto et al., 2019), ranked Weighted Average of Absolute Scores and Response Variable (OrWAASY), and AMMI1 and content × WAAS biplots. GGE biplot analyses were also implemented for protein and oil to produce which-won-where biplots. These biplots delineate mega-environments and show which cultivars performed best, that is, had the highest protein or oil content, in which mega-environments (Yan & Tinker, 2006).

2.7 | GGE biplot analysis with additional environments

In addition to the nine summer 2019/2020 growing season environments, two additional environments from winter 2019 were also available corresponding to location codes CBI-Site 4 and CBI-Site 5 in Table 1. We conducted quality control across all 11 environments, as described in Section 2.5, resulting in 14 of the original 19 cultivars present across all environments. ANOVA tables for the two winter environments are presented in Tables S19–S22. We repeated the GGE biplot analyses with these two additional environments and plotted the which-won-where biplots.

2.8 | USDA GRIN diversity panel and PAT cultivar consensus genotypes

A list of 1417 *G. max* accessions was downloaded from the USDA GRIN Soybean Germplasm Collection core set "SOYBEAN.CORE.MOLECULAR.QS.2021" (https://npgsweb.ars-grin.gov/gringlobal/method?id=496607, accessed April 21, 2023). The core set was developed at the USDA-ARS Soybean Genomics and Improvement Laboratory in Beltsville, Maryland, based on genetic distances among >18,000 *G. max* accessions genotyped using the SoySNP50K array containing >50,000 SNP markers (Song et al., 2015). Briefly, *G. max* accessions were divided into 1417 clusters based on the distance matrix, and from each cluster, the accession with the largest distance to accessions in other clusters was selected. These accessions were further subset by MG II through IX and improvement status of "cultivated" and "cultivar," resulting in a total of 1059 accessions, which will be referred to as the diversity panel from here onward. The full SoySNP50K dataset (Song et al., 2015) was downloaded in version Wm82.a2.v1 from SoyBase (Grant et al., 2010) and subset by our list of 1059 accessions.

After conducting quality control on the PAT cultivar plots based on genetic similarity and CV, we still observed some genetic variability within the same cultivar. To get a representative genetic sequence for each cultivar, we created a consensus genotype from the remaining plots using a custom python script (Mahmood, 2024). For this, we parsed through each SNP position for all the plots belonging to the same cultivar. We assigned the corresponding allele (A, T, C, G, H) when the frequency of the allele was more than 80% between the plots. Further, if no allele exceeded 80% frequency for a SNP position, we checked to see if the SNP position contained two homozygous alleles. If the frequency difference between the two alleles was >15%, we assigned the most prevalent homozygous allele. However, when the frequency difference between the two homozygous alleles was <15%, we assigned the heterozygous allele (H). Finally, the resulting consensus genotypes from the above analysis were converted into VCF format and merged with the SoySNP50K genotyped diversity panel accessions using the isec function in bcftools (Danecek et al., 2021). This resulted in a VCF file containing 1078 accessions and 2527 SNP positions to be used for downstream analysis.

2.9 | Phylogenetic analyses

With the resulting merged SNP dataset, we conducted phylogenetic analyses between the 19 PAT cultivars in our study and the 1059 diversity panel accessions. PCA was conducted on all genotypes using the –pca function in PLINK (Purcell et al., 2007) to examine the genetic structure and variation between all genotypes. The "ggplot2" package in R (Wickham, 2011) was used to visualize the PCA, which was colored according to MG and geographic origin. To find relationships among the PAT cultivars and diversity panel accessions, a neighbor-joining phylogenetic tree was created using the "Create Tree" function with the "Neighbor_Joining Clustering Method" in TASSEL (Bradbury et al., 2007) and visualized using the package "ggtree" in R (Yu et al., 2017). To further examine the genetic relationships among the genotypes, we used the Bayesian clustering program ADMIXTURE (Alexander & Lange, 2011) with 10-fold cross-validation to calculate the appropriate *K* value and identify the model complexity maximizing marginal likelihood. Based on the cross-validation, a *K* value of 4 was selected, and the Admixture genomic structure barplot was visualized using base R.

3 | RESULTS

3.1 | Variation in protein and oil content

In this study, we evaluated nine PAT environments (Table 1; Figure 1) and 17 cultivars (Table 2) for seed protein and

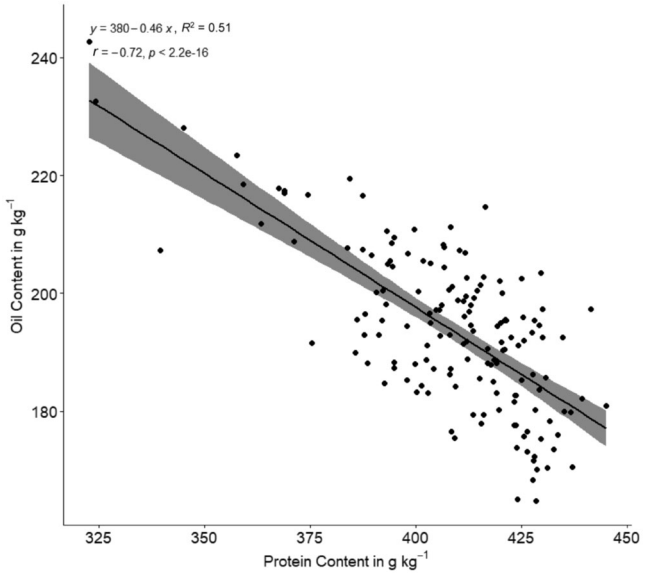

FIGURE 2 Scatterplot of mean protein content (g kg^{-1}) versus mean oil content (g kg^{-1}) for 17 cultivars grown in two replications across nine environments in Zimbabwe during the summer 2019–2020 season. The regression equation, R^2 value, and Pearson's correlation coefficient (r) are listed in the top left corner, and the 95% confidence interval is indicated in gray shading around the trendline.

oil content. We observed a significant negative correlation between seed protein and oil content (dry matter basis), with a Pearson's correlation coefficient of −0.72, and oil content decreasing 0.46 g kg^{-1} for every 1 g kg^{-1} increase in protein content (Figure 2). Mean seed protein content for cultivars at each environment ranged from 322.8 to 445.1 g kg^{-1}, with an overall average of 408.2 g kg^{-1} (Figure 3a). Mean seed oil content for cultivars at each environment ranged from 164.8 to 242.7 g kg^{-1}, with an overall average of 193.9 g kg^{-1} (Figure 3b). From the AMMI analyses, the ANOVAs across all environments for protein content (Table 3) and oil content (Table 4) showed that genotype, environment, and G × E were all significant ($p < 0.05$), and replication in environment was only significant for protein content.

Given the variation in elevation between PAT environments in our study, from 421 to 1504 m (Table 1), we examined correlations between seed composition and elevation. We observed a significant positive correlation between protein content and elevation, with a Pearson's correlation coefficient of 0.69 (Figure 3c). We also found that oil content and elevation had a significant negative correlation, with a Pearson's correlation coefficient of −0.58 (Figure 3d). From these correlations, ENV1 at 421 m appeared to be a potential outlier. When we evaluated these correlations without ENV1, the Pearson's correlation coefficients decreased from 0.69 to 0.35 for protein content and from −0.58 to −0.42 for oil content (Figure S2).

Since elevation does not fully explain these patterns in seed composition, we examined correlations between seed composition and other environmental variables. We mapped PAT environments against soil type (Figure 1b), average daily temperature (Figure 1c), cumulative precipitation (Figure 1d), and photosynthetically active radiation across the growing season (Figure S1). From site-specific weather data for each PAT environment (Table 1), we initially observed correlations of protein and oil with temperature (Figure S3a,c) and precipitation (Figure S4a,c). When ENV1 was removed, the correlations diminished, indicating ENV1 was causing spurious correlations for temperature (Figure S3b,d) and precipitation (Figure S4b,d). With soil test data from each PAT environment used in this study, we also examined correlations between seed composition and soil pH (Figure S5a,c) and soil Na content (Figure S6a,c). Again, when we removed ENV1 for soil pH (Figure S5b,d) and soil Na content (Figure S6b,d), the correlations with seed composition diminished. Given these results, ENV1 appears to be an outlier compared to the rest of the environments for the environmental variables we examined.

3.2 | Seed protein content stability analyses

For our seed protein content AMMI analysis of 17 cultivars in nine environments, genotype, environment, replication in environment, and G × E were significant ($p < 0.05$) (Table 3). Environment explained the most variation out of the total sum of squares at 56.2%, followed by G × E at 19.2%, genotype at 18.7%, block in replication in environment at 4.4%, and replication in environment at 1.5%. The first two PCs were significant for this analysis, with PC1 explaining the majority of the variation in G × E at 59%, and PC2 explaining 15.4% (Table 3; Figure 4a). In the AMMI1 biplot, with PC1 on the *Y*-axis, cultivars and environments with values near zero for PC1 are considered most stable with low G × E, while those with higher absolute PC1 scores are considered unstable (Gauch et al., 2008). Additionally, seed composition content is plotted on the *X*-axis, with cultivars further to the right having a higher mean content. For protein content, MAKWACHA, SC SL01, and MHOFU have PC1 scores closest to zero and are considered most stable (Figure 4a). SC EXPT2, SC SAGA, and TGX 2002-9FM have the highest absolute PC1 scores and are considered least stable. SC EXPT1 had the highest mean protein content, and LUKANGA had the lowest. Additionally, most of the environments clustered together in the lower right quadrant, indicating generally higher protein content, while ENV1 and ENV2 fell in the upper left quadrant with generally lower protein content.

We also calculated WAAS values and created a protein content × WAAS biplot (Figure 4c). In this biplot, similar to an AMMI1 biplot, content is plotted on the *X*-axis;

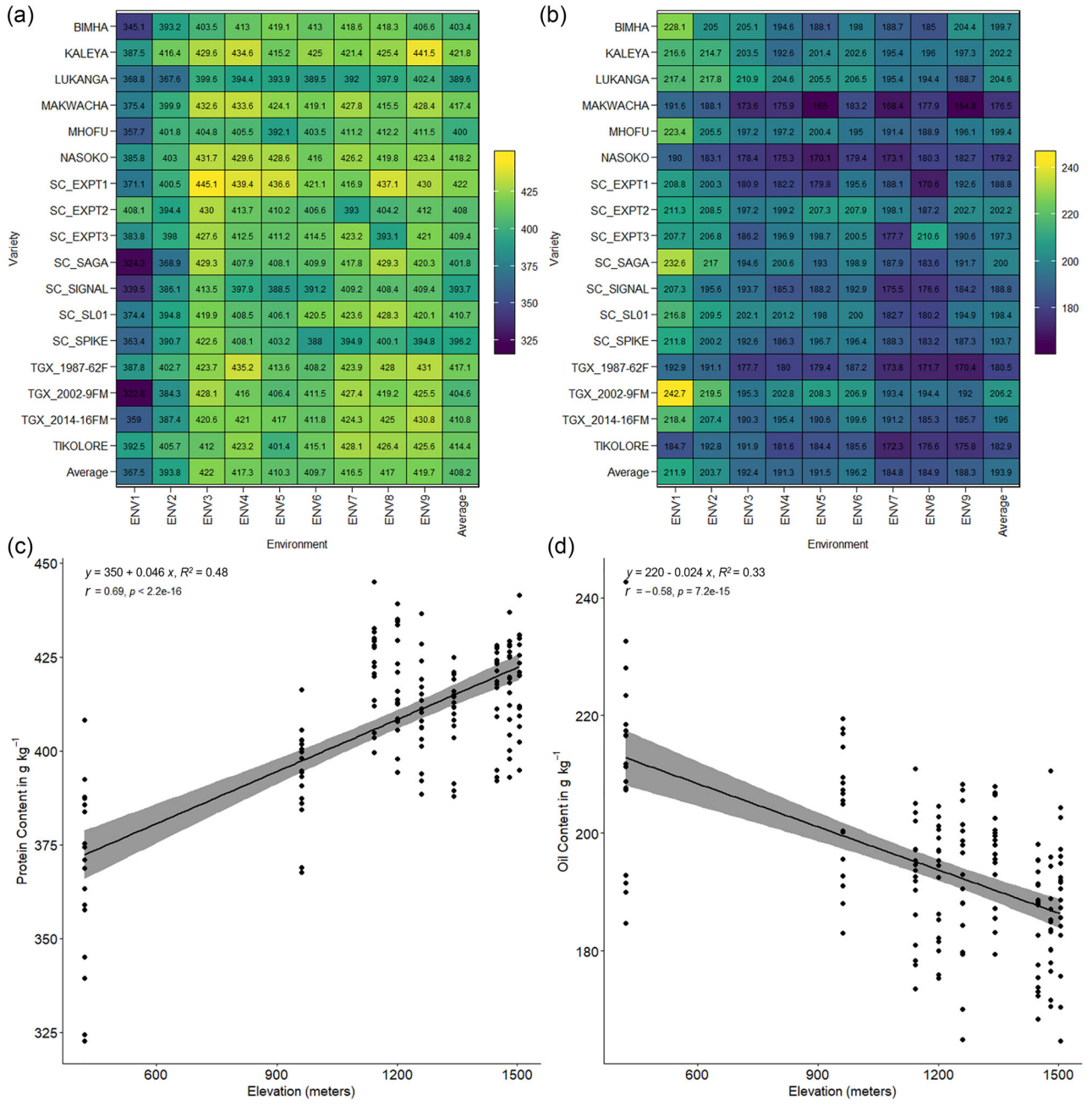

FIGURE 3 Heatmaps of (a) mean protein content (g kg^{-1}) and (b) mean oil content (g kg^{-1}) by cultivar and environment. Environments are organized by increasing elevation from left to right, and the key for the environment codes can be found in Table 1. Scatterplots of (c) elevation (meters) versus mean protein content (g kg^{-1}) and (d) elevation (meters) versus mean oil content (g kg^{-1}). The regression equations, R^2 values, and Pearson's correlation coefficients (r) are listed in the top left corners, and 95% confidence intervals are indicated in gray shading around the trendlines.

however, WAAS values are plotted on the *Y*-axis. WAAS values are a measure of stability calculated from the significant PCs ($p < 0.05$) in an AMMI analysis (Olivoto et al., 2019). For protein content, PC1 and PC2 were significant and used in calculation of the WAAS values. In the content × WAAS biplot, cultivars and environments with lower WAAS values are considered more stable than those with higher WAAS values (Olivoto et al., 2019). This format allows for interpretation of quadrants: quadrant I in the upper left corner indicates low content and low stability, quadrant II in the upper right indicates high content and low stability, quadrant III in the lower left indicates low content and high stability, and quadrant IV in the bottom right indicates high content and high stability (Olivoto et al., 2019). Quadrant IV in the protein content × WAAS biplot contained several cultivars: MAKWACHA, SC SL01, NASOKO, TGX 1987-62F,

TABLE 3 Additive main effects and multiplicative interaction (AMMI) analysis table for seed protein content in g kg^{-1} of 17 cultivars grown in two replications across nine environments in Zimbabwe during the summer 2019–2020 season.

Source	Df	Sum Sq	Mean Sq	*F*-value	Pr(>*F*)	PVE	Accumulated PVE
Environment (E)	8	73,840.3	9230	130.1	2.10×10^{-29}	56.2	NA
Replication (R) in E	9	1948.8	216.5	3.1	0.005871	1.5	NA
Block in R in E	69	5828.2	84.5	1.2	0.264992	4.4	NA
Genotype (G)	16	24,564.7	1535.3	21.6	2.07×10^{-16}	18.7	NA
G × E	128	25,270.6	197.4	2.8	6.37×10^{-5}	19.2	NA
PC1	23	17,502.5	761	10.7	0	59	59
PC2	21	4581.7	218.2	3.1	7.00×10^{-4}	15.4	74.4
PC3	19	2160.9	113.7	1.6	0.0962	7.3	81.7
PC4	17	1840.1	108.2	1.5	0.1253	6.2	87.9
PC5	15	1625.9	108.4	1.5	0.1331	5.5	93.3
PC6	13	1209.3	93	1.3	0.2412	4.1	97.4
PC7	11	603.8	54.9	0.8	0.6673	2	99.4
PC8	9	165.3	18.4	0.3	0.9822	0.6	100
Residuals	47	3334.8	71	NA	NA	NA	NA

Abbreviation: PVE, percent variation explained from the sum of squares.

TABLE 4 Additive main effects and multiplicative interaction (AMMI) analysis table for seed oil content in grams per kilogram of 17 cultivars grown in two replications across nine environments in Zimbabwe during the summer 2019–2020 season.

Source	Df	Sum Sq	Mean Sq	*F*-value	Pr(>*F*)	PVE	Accumulated PVE
Environment (E)	8	19,792.5	2474.1	60.1	4.35×10^{-22}	37.6	NA
Replication (R) in E	9	224.1	24.9	0.6	0.786339	0.4	NA
Block in R in E	69	1349.7	19.6	0.5	0.997598	2.6	NA
Genotype (G)	16	23,075.6	1442.2	35	1.02×10^{-20}	43.8	NA
G × E	128	8208.5	64.1	1.6	4.18×10^{-2}	15.6	NA
PC1	23	3515.8	152.9	3.7	1.00×10^{-4}	36.8	36.8
PC2	21	2111.4	100.5	2.4	5.60×10^{-3}	22.1	58.9
PC3	19	1485.8	78.2	1.9	0.0379	15.5	74.4
PC4	17	974.4	57.3	1.4	0.1845	10.2	84.6
PC5	15	758.8	50.6	1.2	0.2843	7.9	92.5
PC6	13	406.3	31.3	0.8	0.696	4.3	96.8
PC7	11	166.2	15.1	0.4	0.9615	1.7	98.5
PC8	9	140.7	15.6	0.4	0.9389	1.5	100
Residuals	47	1934.4	41.2	NA	NA	NA	NA

Abbreviation: PVE, percent variation explained from the sum of squares.

KALEYA, TGX 2014-16FM, SC EXPT1, and SC EXPT3. Quadrant IV also contained ENV4, ENV6, ENV9, ENV5, and ENV3. From our WAAS analysis, we compiled WAAS, ranked Weighted Average of Absolute Scores (OrWAAS), and OrWAASY for the protein content of each cultivar (Table 5). For protein content stability, MAKWACHA ranked first by both OrWAAS and OrWAASY, while SC EXPT2 ranked last for OrWAAS and SC SAGA ranked last for OrWAASY.

The which-won-where GGE biplot for protein content is shown in Figure 5a, delineating mega-environments and highlighting cultivars with the highest protein content in each mega-environment. In GGE biplots, the PCs contain variation from both genotype and genotype × environment interaction (Yan & Tinker, 2006), as opposed to only G × E in the AMMI PCs. PC1 in this analysis explained the majority of the variation from G + G × E at 50.75%, with PC2 explaining 28.87%. In this plot, SC EXPT2 won in the ENV1 mega-environment, KALEYA won in the ENV2 mega-environment, and SC EXPT1 won in the ENV3-9 mega-environment. SC SL01 fell closest to zero for both PCs and would be considered the most stable cultivar for protein content in this analysis.

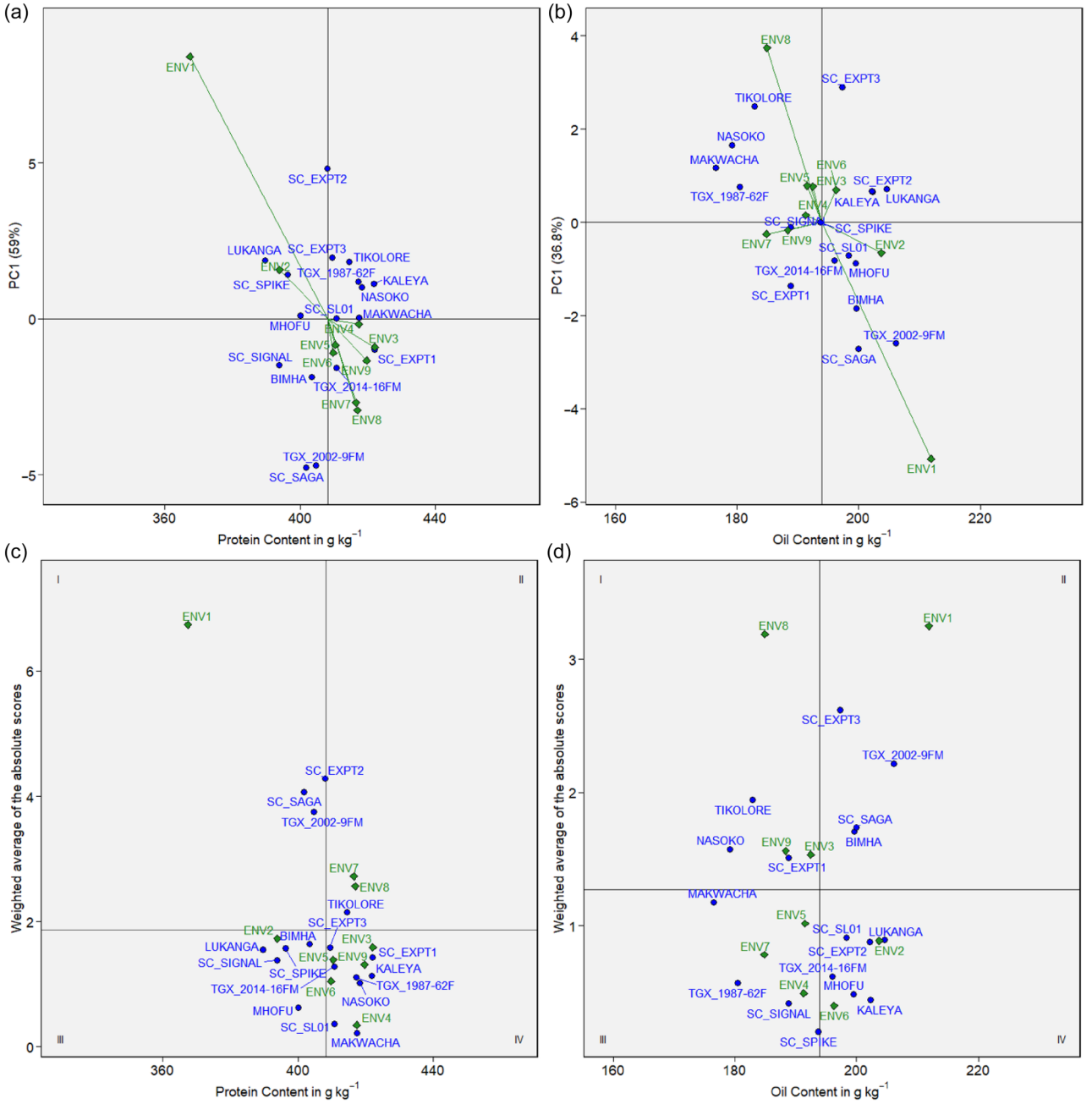

FIGURE 4 Additive main effects and multiplicative interaction 1 (AMMI1) biplots for (a) protein content (g kg^{-1}) and (b) oil content (g kg^{-1}), and content × Weighted Average of Absolute Scores (WAAS) biplots for (c) protein content (g kg^{-1}) and (d) oil content (g kg^{-1}). Biplots include 17 cultivars and nine environments, with green diamonds representing environments, and blue circles representing cultivars. The key for the environment codes can be found in Table 1. For the AMMI1 biplots, the *X* axes represent content (main effects), while the *Y* axes represent PC1 of the corresponding AMMI analysis (interaction effects). For the content × WAAS biplots, the *X* axes represent content (main effects), while the *Y* axes represent WAAS, a measure of stability based on all significant ($p < 0.05$) PCs from the AMMI analysis.

3.3 | Seed oil content stability analyses

For our seed oil content AMMI analysis of 17 cultivars in nine environments, genotype, environment, and G × E were significant ($p < 0.05$) (Table 4). Genotype explained the most variation out of the total sum of squares at 43.8%, followed by environment at 37.6%, G × E at 15.6%, block in replication in environment at 2.6%, and replication in environment at 0.4%. The first three PCs were significant for this analysis, with PC1 explaining the majority of the variation in G × E at 36.8%, PC2 explaining 22.1%, and PC3 explaining 15.5% (Table 4; Figure 4b). In the AMMI1 biplot for oil content, SC SPIKE and SC SIGNAL have PC1 scores closest to zero and are considered most stable (Figure 4b). SC EXPT3, TIKOLORE, SC

TABLE 5 Mean protein and oil content, Weighted Average of Absolute Scores (WAAS), ranked WAAS (OrWAAS), and ranked Weighted Average of Absolute Scores and Response Variable (OrWAASY) of 17 soybean cultivars across nine environments in Zimbabwe.

Cultivar	Mean protein (g kg^{-1})	Protein WAAS	Protein OrWAAS	Protein OrWAASY	Mean oil (g kg^{-1})	Oil WAAS	Oil OrWAAS	Oil OrWAASY
BIMHA	403.4	1.6	13	11	199.7	1.7	13	10
KALEYA	421.8	1.1	6	2	202.2	0.4	3	1
LUKANGA	389.5	1.5	10	14	204.6	0.9	8	2
MAKWACHA	417.4	0.2	1	1	176.5	1.2	10	15
MHOFU	400.0	0.6	3	10	199.4	0.5	4	3
NASOKO	418.2	1.0	4	4	179.2	1.6	12	16
SC EXPT1	422.0	1.4	9	3	188.8	1.5	11	13
SC EXPT2	408.0	4.3	17	16	202.2	0.9	7	4
SC EXPT3	409.4	1.6	12	9	197.3	2.6	17	14
SC SAGA	401.8	4.1	16	17	200.0	1.7	14	11
SC SIGNAL	393.7	1.4	8	13	188.8	0.4	2	8
SC SL01	410.7	0.4	2	6	198.4	0.9	9	7
SC SPIKE	396.2	1.6	11	12	193.7	0.2	1	5
TGX 1987-62F	417.1	1.1	5	5	180.5	0.6	5	12
TGX 2002-9FM	404.6	3.7	15	15	206.2	2.2	16	9
TGX 2014-16FM	410.8	1.3	7	7	196.0	0.6	6	6
TIKOLORE	414.4	2.1	14	8	182.9	1.9	15	17

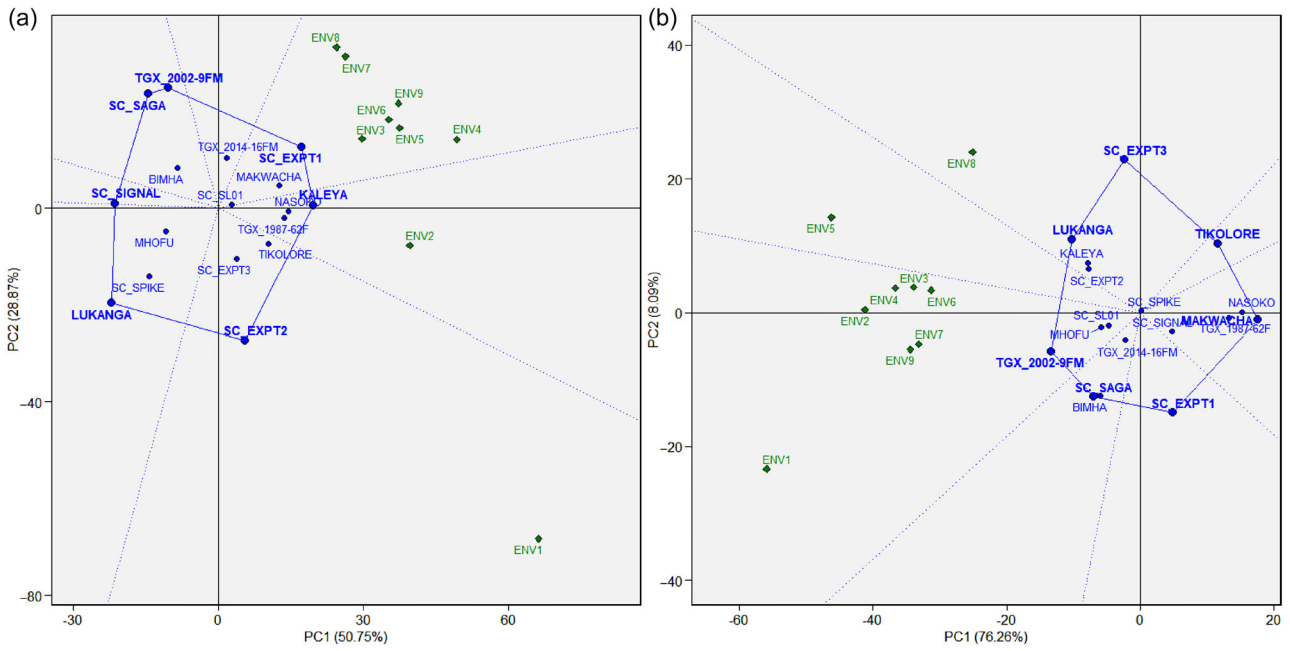

FIGURE 5 Which-won-where genotype plus genotype-by-environment (GGE) biplots of 17 cultivars and nine environments for (a) protein content (g kg^{-1}) and (b) oil content (g kg^{-1}). Green diamonds represent environments, while blue circles represent cultivars. The key for the environment codes can be found in Table 1. The *X* axes represent PC1, while the *Y* axes represent PC2 of the corresponding GGE analysis. The polygons connect genotypes furthest from the origin of the biplots, with dotted lines delineating mega-environments, and genotypes at the polygon vertices performing best in those respective mega-environments.

SAGA, and TGX 2002-9FM have the highest absolute PC1 scores and are considered least stable. TGX 2002-9FM had the highest mean oil content, and MAKWACHA had the lowest. Additionally, ENV1, ENV2, ENV3, and ENV6 fell to the right of the oil content mean, indicating generally higher oil content, while ENV4, ENV5, ENV7, ENV8, and ENV9 fell to the left with generally lower oil content.

For the WAAS analysis of oil content, we created an oil content × WAAS biplot (Figure 4d). For oil content, PC1, PC2, and PC3 were significant and used in calculation of the WAAS values. Quadrant IV in the oil content × WAAS biplot contained several cultivars: KALEYA, MHOFU, TGX 2014-16FM, SC EXPT2, LUKANGA, and SC SL01. Quadrant IV also contained ENV6 and ENV2. From our WAAS analysis, we compiled WAAS, OrWAAS, and OrWAASY for oil content of each cultivar (Table 5). For oil content stability, SC SPIKE ranked first by OrWAAS and KALEYA ranked first by OrWAASY, while SC EXPT3 ranked last for OrWAAS and TIKOLORE ranked last for OrWAASY.

The which-won-where GGE biplot for oil content is shown in Figure 5b, delineating mega-environments and highlighting cultivars with the highest oil content in each mega-environment. PC1 in this analysis explained the majority of the variation from G + G × E at 76.26%, with PC2 explaining 8.09%. In this plot, LUKANGA and SC EXPT3 won in the ENV8 mega-environment, no cultivars present in our study won in the ENV5 mega-environment, and TGX 2002-9FM won in the mega-environment containing the rest of the environments. SC SPIKE fell closest to zero for both PCs and would be considered the most stable cultivar for oil content in this analysis.

3.4 | Evaluation of mega-environments across growing seasons

While all the above analyses were conducted on the nine summer 2019/2020 growing season environments, two additional environments from winter 2019 were also available. These environments correspond to location codes CBI-Site 4 and CBI-Site 5 in Table 1. After quality control across all 11 environments, 14 of the original 19 cultivars remained. To compare mega-environment groupings within and across growing seasons, we repeated the GGE biplot analyses with these two additional environments.

The which-won-where GGE biplot for protein content including two additional winter environments is shown in Figure S7a, delineating mega-environments and highlighting cultivars with the highest protein content in each mega-environment. PC1 in this analysis explained the majority of the variation from G + G × E at 49.99%, with PC2 explaining 27.93%. In this plot, SC EXPT2 won in the ENV1 mega-environment, TIKOLORE won in the ENV2 and ENV10 mega-environment, TGX 2002-9FM won in the ENV8 mega-environment, and MAKWACHA won in the mega-environment containing the rest of the environments. MHUFO fell closest to zero for both PCs and would be considered the most stable cultivar for protein content across all 11 environments in this analysis.

The which-won-where GGE biplot for oil content including two additional winter environments is shown in Figure S7b. PC1 in this analysis explained the majority of the variation from G + G × E at 77.68%, with PC2 explaining 6.18%. In this plot, no cultivars present in our study won in the ENV3,5,6,8,9 mega-environment, and TGX 2002-9FM won in the mega-environment containing the rest of the environments. SC SPIKE fell closest to zero for both PCs and would be considered the most stable cultivar for oil content across all 11 environments in this analysis.

3.5 | Phylogenetic analyses of PAT cultivars and diversity panel accessions

In addition to the stability analyses of protein and oil content for PAT cultivars, we conducted phylogenetic analyses on 19 PAT cultivars and a diversity panel of 1059 USDA GRIN accessions. The accessions were subset from a core collection to only contain cultivars of MG II through MG IX. We used a total of 2527 genome-wide SNP markers to conduct PCA and create a neighbor-joining phylogenetic tree of the 1078 accessions (Figure 6). The genetic variation explained by PC1 was 21.7% and PC2 was 16.3% (Figure 6a,b). The PCA analysis with accessions colored according to MG revealed a general trend of lower MG accessions falling in the upper left and higher MG accessions falling in the lower right, with the PAT cultivars clustering in the middle left (Figure 6a). For the PCA plot colored by geographic origin of accessions, a pattern was less apparent; however, North American accessions were grouped most to the left, while Asian accessions were distributed more to the middle and right (Figure 6b). The PAT accessions grouped closer to the North American accessions, along with some accessions from South America and Africa.

A similar grouping by geographic origin appeared in the neighbor-joining phylogenetic tree (Figure 6c). Most of the branches on the lower half of the tree contained Asian accessions, while most of the North American accessions were within the branches on the top of the tree. One branch contained 14 of the PAT cultivars, along with a few other accessions from Africa and one South American accession. KALEYA fell nearby, grouping with mostly North American accessions. TGX 2014-16FM grouped with other African accessions among Asian and North American accessions on the top right. Lastly, TIKOLORE, TGX 1987-62F, and TGX

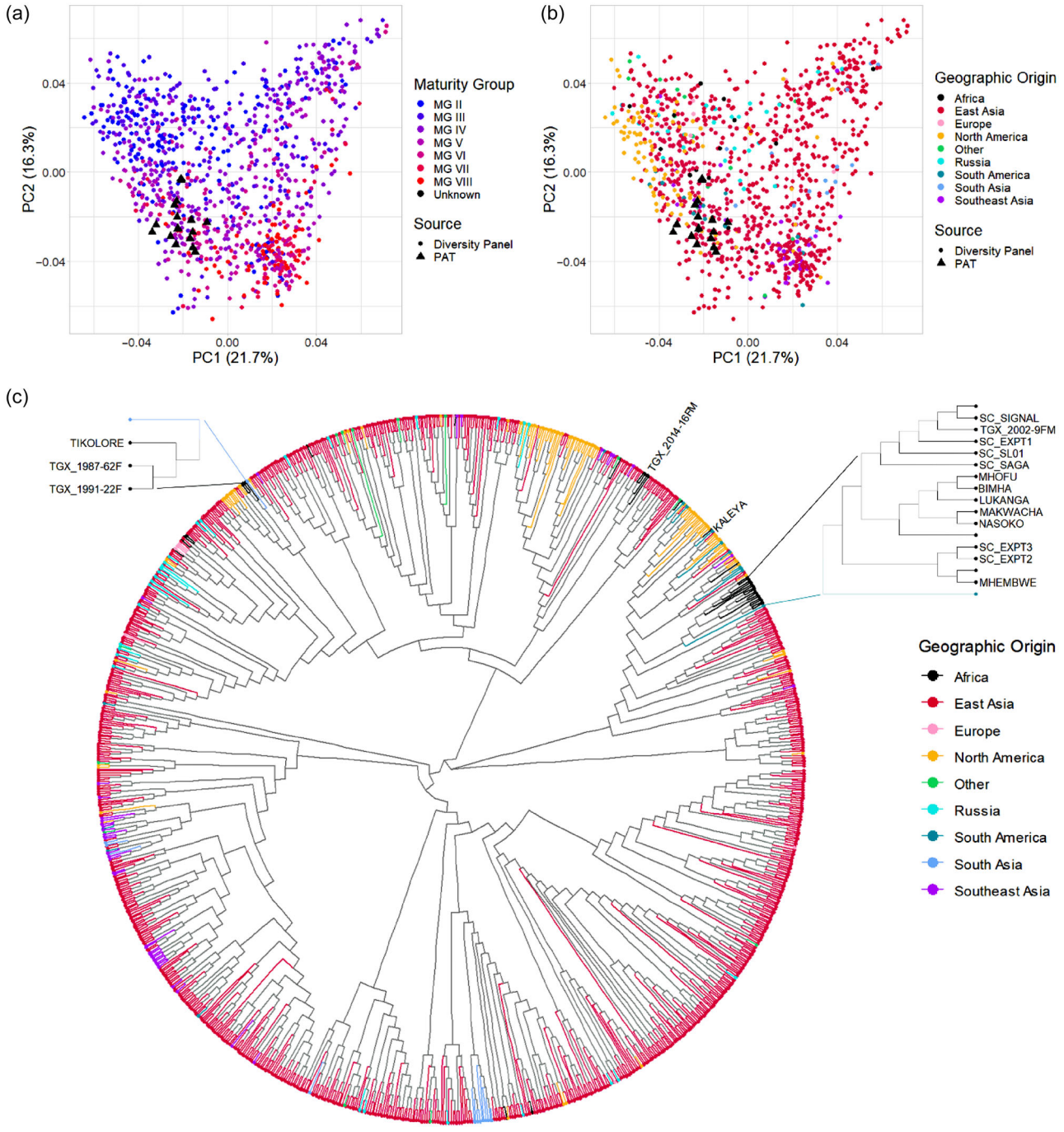

FIGURE 6 Principal component analysis by (a) maturity group and (b) geographic origin, and (c) a neighbor-joining phylogenetic tree of 19 Pan-African Soybean Variety Trial (PAT) cultivars with 1059 diversity panel accessions. Analyses were completed using 2527 genome-wide single-nucleotide polymorphism markers.

1991-22F grouped with Asian and North American accessions on the top left. We also conducted an Admixture analysis with a *K* value of 4 on the PAT cultivars and diversity panel accessions to examine their genomic structure (Figure S8). Based on the patterns of ancestral proportions, the PAT cultivars appeared most similar to the North American and South American accessions. For reference, we created the neighbor-joining phylogenetic tree with all accession names (Figure S9), along with a table of the diversity panel accessions with USDA GRIN phenotypes (Table S23).

4 | DISCUSSION

Through these stability analyses, we identified PAT cultivars that exhibited the best performance in specific environments

and the highest overall stability across environments for seed protein and oil content. We also identified protein and oil content mega-environments within the nine environments across Zimbabwe. Through creating a consensus genotype for each of the 19 PAT cultivars in this study and conducting PCA and phylogenetic analysis, we revealed patterns in MG and geographic origin relative to USDA GRIN diversity panel accessions.

In this study, we observed a range of protein and oil contents across 17 cultivars and nine environments, with a highly significant negative correlation between protein and oil content (Figure 2), consistent with previous reports of soybean grown in North America (Chung et al., 2003; La et al., 2019; Leamy et al., 2017). It has also been reported that both protein and oil are complex and highly polygenic traits with significant G × E effects (Chung et al., 2003; Lee et al., 2007; Warrington et al., 2015). We found G × E, genotype, and environment each explained significant ($p < 0.05$) amounts of variation in protein (Table 3) and oil (Table 4) across the nine PAT environments. A previous study examining protein and oil stability of soybean in Uganda found only environment to be significant for both protein and oil (Mukuze et al., 2020), while a similar study in Zambia found that G, E, and G × E were all significant (Hampango et al., 2017). Some studies of soybean protein and oil content stability have reported a larger influence of environment on protein content (Carrera et al., 2009; Šarčević et al., 2022; Sudarić et al., 2006), while other studies have reported the opposite pattern with environment having more influence on oil content (Flajšman et al., 2019; Gurmu et al., 2009). In our study, environment explained the most variation for protein, while genotype explained the most variation for oil. G × E explained a similar level of variation for both protein and oil. Additionally, replication in environment explained a significant amount of variation for protein, indicating that environmental factors impacted protein both within and across environments.

Given the correlations of protein (Figure 3c) and oil with elevation (Figure 3d), we examined other environmental factors as potential explanations for the relationships we observed. Upon examination of soil data from the PAT environments, we found significant negative correlations of protein with soil pH and soil Na content and significant positive correlations of oil with soil pH and soil Na content (Figures S5 and S6). The environment with the lowest average protein and highest average oil, ENV1, also had the highest soil pH (8.3) and Na content (156 mg kg^{-1}). For the individual environment ANOVAs, ENV1 was the only environment with significant ($p < 0.05$) replication and block in replication for protein and oil (Tables S1 and S10). These ANOVA results can likely be explained by variability in soil characteristics across the field at ENV1. The lowest soil pH environment, ENV6, at pH 5.3, also had lower protein content and higher oil content compared to the rest of the environments above 1000 m. High and low soil pH have been previously reported to impact nodule formation in legumes, thereby impacting nitrogen fixation (Miransari et al., 2013). High soil salinity has also been reported to impact nodule formation of soybean (Miransari et al., 2013). Therefore, both soil pH and Na should be evaluated and considered in soybean trials and production in Zimbabwe and the whole of SSA, especially for maintaining protein levels and yield.

For our AMMI stability analyses of 17 PAT cultivars in nine environments, the first two PCs representing the G × E for protein were significant, while the first three PCs were significant for oil. The first two PCs explained >50% of the variation in both cases. For the protein AMMI1 biplot, the most stable cultivars were MAKWACHA, SC SL01, and MHOFU, and the highest mean protein content cultivars were SC EXPT1, KALEYA, and NASOKO (Figure 4a). Additionally, ENV4 appeared to be the most stable protein environment, and ENV1 was the most unstable. For the oil AMMI1 biplot, the most stable cultivars were SC SPIKE and SC SIGNAL, and the highest mean oil content cultivars were TGX 2002-9FM, LUKANGA, and SC EXPT2 (Figure 4b). Similar to the protein results, ENV4 appeared to be the most stable oil environment, and ENV1 was the most unstable. Overall, our AMMI stability analyses inform breeders and producers about which PAT cultivars and environments provide the best protein and oil performance.

WAAS values are a measure of stability based on all significant PCs from an AMMI analysis, compared to AMMI1 biplots for which stability rankings are based only on PC1 (Olivoto et al., 2019). Consequently, WAAS values can result in different rankings than those obtained from AMMI1 biplots depending on how much variance is captured in the different significant PCs (Olivoto et al., 2019). However, the most stable cultivars identified in our AMMI1 biplots corresponded with those identified by WAAS values (Table 5). This indicates that the first PC captured enough variation to identify the most stable cultivars. The content × WAAS biplots provide a clear visualization of cultivar performance, with the most stable and highest content cultivars in quadrant IV in the lower right corner. For protein, these cultivars were MAKWACHA, SC SL01, NASOKO, TGX 1987-62F, KALEYA, TGX 2014-16FM, SC EXPT1, and SC EXPT3 (Figure 4c). For oil, these cultivars were KALEYA, MHOFO, TGX 2014-16FM, SC EXPT2, LUKANGA, and SC SL01 (Figure 4d). The cultivars that overlap between these two lists are recommended for use by breeders and producers in Zimbabwe for their high and stable protein and oil content: SC SL01, KALEYA, and TGX 2014-16FM.

From our which-won-where GGE biplots of nine summer environments, we identified stable cultivars, mega-environments, and the best-performing cultivars in each mega-environment. For high protein, SC EXPT1 is recommended in the ENV3-9 mega-environment, KALEYA

is recommended in the ENV2 mega-environment, and SC EXPT2 is recommended in the ENV1 mega-environment (Figure 5a). SC EXPT2 performed well compared to the other cultivars in ENV1 for protein despite both high soil pH and salinity and is recommended for use in similar environments. For high oil, TGX 2002-9FM is recommended in the ENV1,2,3,4,6,7,9 mega-environment, and LUKANGA and SC EXPT3 are recommended in the ENV8 mega-environment (Figure 5b). Since protein and oil are negatively correlated in this study, one might expect to see the same mega-environment groupings with opposite cultivars winning for protein and oil; however, this was not observed. The differences in mega-environment groupings between protein and oil suggest that the environmental factors influenced protein and oil differently in the PAT cultivars. This is apparent from the overall ANOVAs where the proportion of variation explained by environment for protein and oil differed, with E explaining the most variation for protein and G explaining the most variation for oil.

The addition of two winter environments to the which-won-where GGE biplots resulted in shifts of mega-environment groupings and winning cultivars for both protein and oil content (Figure S7). While these two winter environments corresponded to two of the summer environment growing locations, these corresponding environments did not fall within the same mega-environments for protein content. However, all four of these environments fell within the same mega-environment for oil content. These results suggest that the winter and summer growing seasons in Zimbabwe pose slightly different environmental conditions that could impact protein and oil content differently. Additionally, the mega-environment groupings of 17 cultivars in nine environments (Figure 5) differed compared to our analysis with 14 cultivars in 11 environments (Figure S7). These differences could be explained by the exclusion of certain cultivars and the inclusion of additional environments impacting the variance captured within the PCs in the biplots. Furthermore, mega-environment groupings should be based on repeatable patterns of G × E from multiple years of data (Yan et al., 2023). Therefore, additional research will be needed to conclusively determine mega-environments for seed composition in Zimbabwe.

The phylogenetic analyses revealed limited genetic diversity of the PAT cultivars compared to USDA GRIN diversity panel accessions (Figure 6). In our PCA, the 19 PAT cultivars generally grouped with accessions ranging from MG IV to VII (Figure 6a). A previous study demonstrated that the most appropriate MG for East Africa (covering latitude 5°N–20°S and longitude 30°E–39°E) would be MG VII, and for West Africa (covering latitude 15°N–6°N and longitude 17°W–10°E), it would be MG VIII, with more southern regions of East Africa potentially requiring lower MGs (Sinclair et al., 2014). We also found that the PAT cultivars grouped mostly with accessions from North America and South America (Figure 6b,c). This is consistent with the sources of lines introduced for breeding in Africa, with a majority coming from the southern United States and Brazil (Khojely et al., 2018).

Although 14 of the PAT cultivars examined grouped in the same branch of the phylogenetic tree with other African accessions, five cultivars appeared to be more diverse, including KALEYA, TGX 2014-16FM, TIKOLORE, TGX 1987-62F, and TGX 1991-22F. These last three cultivars clustered together on the same branch, and upon closer inspection, they had >99% genetic similarity. This contradicts a previous study on genetic diversity among IITA soybean breeding program lines where TIKOLORE and TGX 1987-62F were reported as genetically distinct (Chigeza et al., 2019). That study also discussed the measures taken to increase genetic diversity in the IITA program, and our results confirm that the TGX cultivars examined here show wider genetic diversity compared to the other PAT cultivars. Other studies on genetic diversity of African soybean cultivars have not included a wide range of accessions for comparison (Chander et al., 2021; Chigeza et al., 2019; Tsindi et al., 2023). Therefore, our results provide unique insight into the genetic relatedness between African cultivars and diverse accessions from around the world. Many USDA GRIN Soybean Germplasm Collection accessions have been screened for various traits, including seed composition and biotic and abiotic stress resistance (https://npgsweb.ars-grin.gov/gringlobal/descriptors, accessed April 27, 2023). The supplemental file showing the phylogenetic tree with all accession names (Figure S9) and a table of diversity panel accessions with USDA GRIN phenotypes (Table S23) can be used together to identify accessions of interest with valuable traits.

Overall, our study presented comprehensive seed composition stability analyses within PATs and the first examination of PAT cultivar genetic diversity compared to diverse accessions. Through AMMI and GGE biplot analyses, we identified PAT cultivars with the highest stability and adaptability in nine diverse environments across Zimbabwe. We examined PAT cultivars in the context of diverse soybean accessions through PCA and phylogenetic analysis and presented genetic relatedness by both MG and geographic origin. The stable and adaptable PAT cultivars identified in our study are recommended for utilization by breeders and producers in Africa, while genetically distinct accessions with valuable traits are highlighted as a resource for integration into breeding programs.

AUTHOR CONTRIBUTIONS

Elizabeth De Meyer: Conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Elizabeth Prenger**: Conceptualization; data curation;

investigation; writing—original draft; writing—review and editing. **Anser Mahmood**: Conceptualization; data curation; investigation; methodology; writing—original draft; writing—review and editing. **Michelle da Fonseca Santos**: Project administration; resources; writing—review and editing. **Godfree Chigeza**: Project administration; resources; writing—review and editing. **Qijian Song**: Data curation; writing—review and editing. **Learnmore Mwadzingeni**: Resources; writing—review and editing. **Ronica Mukaro**: Resources; writing—review and editing. **Mwila Chibanda**: Resources; writing—review and editing. **Gorden Mabuyaye**: Resources; writing—review and editing. **Brian Diers**: Project administration; supervision; writing—review and editing. **Andrew Scaboo**: Conceptualization; funding acquisition; project administration; resources; supervision; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

We would like to thank the coordinators and participants of the United States Agency for International Development (USAID) Feed the Future Soybean Innovation Lab Pan-African Soybean Variety Trials for providing data used in this study, particularly Erica Leles for assistance accessing PAT data. We also thank the personnel in the University of Missouri Northern Soybean Breeding Program for their efforts in collecting data for this study. We declare financial support from the USAID Feed the Future Soybean Innovation Lab and the Missouri Soybean Merchandising Council (Award ID #16-340). E.D.M. is partially funded via the University of Missouri Life Sciences Fellowship.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the Supporting Information. Further inquiries can be directed to the corresponding author.

ORCID

Elizabeth De Meyer https://orcid.org/0009-0003-0312-6127
Brian Diers https://orcid.org/0000-0003-3584-5495
Andrew Scaboo https://orcid.org/0000-0002-9670-8446

REFERENCES

Acevedo-Siaca, L., & Goldsmith, P. D. (2020). Soy-maize crop rotations in sub-Saharan Africa: A literature review. *International Journal of Agronomy*, *2020*, Article 8833872. https://doi.org/10.1155/2020/8833872

Adane, A., Alemayehu, D., Meseret, T., & Chala, D. (2018). Genotype and genotype by environment interaction and grain yield stability of medium maturity groups of soybean [*Glycine max* (L.) Merrill] cultivars in Western Oromia, Ethiopia. *African Journal of Plant Science*, *12*(9), 227–237. https://doi.org/10.5897/AJPS2018.1674

Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*, Article 246. https://doi.org/10.1186/1471-2105-12-246

Asodina, F. A., Adams, F., Nimoh, F., Weyori, E. A., Wongnaa, C. A., & Bakang, J. E. A. (2020). Are non-market benefits of soybean production significant? An extended economic analysis of smallholder soybean farming in Upper West region of northern Ghana. *Agriculture & Food Security*, *9*, Article 13. https://doi.org/10.1186/s40066-020-00265-7

Assefa, Y., Purcell, L. C., Salmeron, M., Naeve, S., Casteel, S. N., Kovács, P., Archontoulis, S., Licht, M., Below, F., Kandel, H., Lindsey, L. E., Gaska, J., Conley, S., Shapiro, C., Orlowski, J. M., Golden, B. R., Kaur, G., Singh, M., Thelen, K., … Ciampitti, I. A. (2019). Assessing variation in US soybean seed composition (protein and oil). *Frontiers in Plant Science*, *10*, Article 298. https://doi.org/10.3389/fpls.2019.00298

Bain, L. E., Awah, P. K., Geraldine, N., Kindong, N. P., Sigal, Y., Bernard, N., & Tanjeko, A. T. (2013). Malnutrition in Sub–Saharan Africa: Burden, causes and prospects. *Pan African Medical Journal*, *15*(1), Article 120. https://doi.org/10.11604/pamj.2013.15.120.2535

Barter, R. L., & Yu, B. (2018). Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. *Journal of Computational and Graphical Statistics*, *27*(4), 910–922. https://doi.org/10.1080/10618600.2018.1473780

Bellaloui, N., Smith, J. R., Ray, J. D., & Gillen, A. M. (2009). Effect of maturity on seed composition in the early soybean production system as measured on near-isogenic soybean lines. *Crop Science*, *49*(2), 608–620. https://doi.org/10.2135/cropsci2008.04.0192

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brzostowski, L. F., & Diers, B. W. (2017). Agronomic evaluation of a high protein allele from PI407788A on chromosome 15 across two soybean backgrounds. *Crop Science*, *57*(6), 2972–2978. https://doi.org/10.2135/cropsci2017.02.0083

Burgueño, J., Crossa, J., Cotes, J. M., Vicente, F. S., & Das, B. (2011). Prediction assessment of linear mixed models for multienvironment trials. *Crop Science*, *51*(3), 944–954. https://doi.org/10.2135/cropsci2010.07.0403

Carrera, C., Martínez, M. J., Dardanelli, J., & Balzarini, M. (2009). Water deficit effect on the relationship between temperature during the seed fill period and soybean seed oil and protein concentrations. *Crop Science*, *49*(3), 990–998. https://doi.org/10.2135/cropsci2008.06.0361

Chander, S., Garcia-Oliveira, A. L., Gedil, M., Shah, T., Otusanya, G. O., Asiedu, R., & Chigeza, G. (2021). Genetic diversity and population structure of soybean lines adapted to sub-Saharan Africa using single nucleotide polymorphism (SNP) markers. *Agronomy*, *11*(3), Article 604. https://doi.org/10.3390/agronomy11030604

Cheelo, P., Lungu, D., & Mwala, M. (2017). GGE biplot analysis for identification of ideal soybean [*Glycine max* L. Merrill] test and production locations in Zambia. *Journal of Experimental Agriculture International*, *15*(3), 1–15. https://doi.org/10.9734/JEAI/2017/30154

Chigeza, G., Boahen, S., Gedil, M., Agoyi, E., Mushoriwa, H., Denwar, N., Gondwe, T., Tesfaye, A., Kamara, A., Alamu, O. E., & Chikoye, D.

(2019). Public sector soybean (*Glycine max*) breeding: Advances in cultivar development in the African tropics. *Plant Breeding*, *138*(4), 455–464. https://doi.org/10.1111/pbr.12682

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., Shoemaker, R. C., & Specht, J. E. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Science*, *43*(3), 1053–1067. https://doi.org/10.2135/cropsci2003.1053

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., Mccarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), Article giab008. https://doi.org/10.1093/gigascience/giab008

Diers, B., & Scaboo, A. (2019). Soybean breeding in Africa. *African Journal of Food, Agriculture, Nutrition and Development*, *19*(5), 15121–15125. https://doi.org/10.18697/ajfand.88.SILFarmDoc03

Dinku, T., Faniriantsoa, R., Islam, S., Nsengiyumva, G., & Grossi, A. (2022). The climate data tool: Enhancing climate services across Africa. *Frontiers in Climate*, *3*, Article 787519. https://doi.org/10.3389/fclim.2021.787519

Doyle, J. J., & Doyle, J. L. (1987). A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Eberhart, S. T., & Russell, W. A. (1966). Stability parameters for comparing cultivars 1. *Crop Science*, *6*(1), 36–40. https://doi.org/10.2135/cropsci1966.0011183x000600010011x

Engelbrecht, G., Claassens, S., Mienie, C. M., & Fourie, H. (2020). South Africa: An important soybean producer in sub-Saharan Africa and the quest for managing nematode pests of the crop. *Agriculture*, *10*(6), Article 242. https://doi.org/10.3390/agriculture10060242.[CrossRef]

FAOSTAT. (2024). *Crops and livestock products*. https://www.fao.org/faostat/en/#data/QCL/visualize

Finlay, K. W., & Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Australian Journal of Agricultural Research*, *14*(6), 742–754. https://doi.org/10.1071/AR9630742

Flajsman, M., Santavec, I., Kolmanic, A., Kosmelj, K., & Kocjan-Acko, D. (2019). Agronomic performance and stability of seed, protein and oil yields of seven soybean cultivars determined in field experiments in Slovenia. *Genetika*, *51*(1), 31–46. https://doi.org/10.2298/GENSR1901031F

Foyer, C. H., Siddique, K. H., Tai, A. P., Anders, S., Fodor, N., Wong, F. L., Ludidi, N., Chapman, M. A., Ferguson, B. J., Considine, M. J., Zabel, F., Prasad, P. V. V., Varshney, R. K., Nguyen, H. T., & Lam, H. M. (2019). Modelling predicts that soybean is poised to dominate crop production across Africa. *Plant, Cell & Environment*, *42*(1), 373–385. https://doi.org/10.1111/pce.13466

Gauch, H. G. (1992a). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier Science Publishers.

Gauch, H. G. (1992b). AMMI analysis of yield trials. In M. S. Kang & H. G. Gauch (Eds.), *Genotype-by-environment interaction* (pp. 1–40). CRC Press.

Gauch, H. G. (2006). Statistical analysis of yield trials by AMMI and GGE. *Crop Science*, *46*(4), 1488–1500. https://doi.org/10.2135/cropsci2005.07-0193

Gauch, H. G., Jr., Piepho, H. P., & Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Science*, *48*(3), 866–889. https://doi.org/10.2135/cropsci2007.09.0513

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research*, *38*(Suppl_1), D843–D846. https://doi.org/10.1093/nar/gkp798

Gurmu, F., Mohammed, H., & Alemaw, G. (2009). Genotype x environment interactions and stability of soybean for grain yield and nutrition quality. *African Crop Science Journal*, *17*(2), 87–99. https://doi.org/10.4314/acsj.v17i2.54202

Hampango, M., Cheelo, P., Lungu, M., Mwala, M., & Agrama, H. (2017). AMMI determination of stability for oil and protein content in soybean (*Glycine max* L. Merrill) seed in Zambia. *Journal of Experimental Agriculture International*, *17*(5), 1–16. https://doi.org/10.9734/JEAI/2017/35583

Hartman, G. L., West, E. D., & Herman, T. K. (2011). Crops that feed the World 2. Soybean—Worldwide production, use, and constraints caused by pathogens and pests. *Food Security*, *3*, 5–17. https://doi.org/10.1007/s12571-010-0108-x

Jandong, E. A., Uguru, M. I., & Okechukwu, E. C. (2019). Genotype-by-environment interaction and stability analysis of soybean genotypes for yield and yield components across two locations in Nigeria. *African Journal of Agricultural Research*, *14*(34), 1943–1949. https://doi.org/10.5897/AJAR2019.14251

Jandong, E. A., Uguru, M. I., & Oyiga, B. C. (2011). Determination of yield stability of seven soybean (*Glycine max*) genotypes across diverse soil pH levels using GGE biplot analysis. *Journal of Applied Biosciences*, *43*, 2924–2941.

Kalumbi, M., Matumba, L., Mtimuni, B., Mwangwela, A., & Gama, A. P. (2019). Hydrothermally treated soybeans can enrich maize stiff porridge (Africa's main staple) without negating sensory acceptability. *Foods*, *8*(12), Article 650. https://doi.org/10.3390/foods8120650

Khojely, D. M., Ibrahim, S. E., Sapey, E., & Han, T. (2018). History, current status, and prospects of soybean production and research in sub-Saharan Africa. *The Crop Journal*, *6*(3), 226–235. https://doi.org/10.1016/j.cj.2018.03.006

La, T., Large, E., Taliercio, E., Song, Q., Gillman, J. D., Xu, D., Nguyen, H. T., Shannon, G., & Scaboo, A. (2019). Characterization of select wild soybean accessions in the USDA germplasm collection for seed composition and agronomic traits. *Crop Science*, *59*(1), 233–251. https://doi.org/10.2135/cropsci2017.08.0514

Leamy, L. J., Zhang, H., Li, C., Chen, C. Y., & Song, B. H. (2017). A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). *BMC Genomics*, *18*, Article 18. https://doi.org/10.1186/s12864-016-3397-4

Lee, J. D., Bilyeu, K. D., & Shannon, J. G. (2007). Genetics and breeding for modified fatty acid profile in soybean seed oil. *Journal of Crop Science and Biotechnology*, *10*(4), 201–210.

Mahmood, A. (2024). Consensus-cultivar-genotypes. GitHub. https://github.com/AnserM/Consensus-Cultivar-Genotypes

Mangani, C., Maleta, K., Phuka, J., Cheung, Y. B., Thakwalakwa, C., Dewey, K., Manary, M., Puumalainen, T., & Ashorn, P. (2015). Effect of complementary feeding with lipid-based nutrient supplements and corn–soy blend on the incidence of stunting and linear growth among 6-to 18-month-old infants and children in rural Malawi. *Maternal & Child Nutrition*, *11*, 132–143. https://doi.org/10.1111/mcn.12068

Marcillo, G. S., Martin, N. F., Diers, B. W., Da Fonseca Santos, M., Leles, E. P., Chigeza, G., & Francischini, J. H. (2021). Implementation of a generalized additive model (Gam) for soybean maturity prediction in African environments. *Agronomy*, *11*(6), Article 1043. https://doi.org/10.3390/agronomy11061043

Masuda, T., & Goldsmith, P. D. (2009). World soybean production: Area harvested, yield, and long-term projections. *International Food and Agribusiness Management Review*, *12*(4), 1–20. https://doi.org/10.22004/ag.econ.92573

Miransari, M., Riahi, H., Eftekhar, F., Minaie, A., & Smith, D. L. (2013). Improving soybean (*Glycine max* L.) N 2 fixation under stress. *Journal of Plant Growth Regulation*, *32*, 909–921. https://doi.org/10.1007/s00344-013-9335-7

Mukuze, C., Tukamuhabwa, P., Maphosa, M., Dari, S., Obua, T., Kongai, H., & Rubaihayo, P. (2020). Evaluation of the performance of advanced generation soybean [*Glycine max* (L.) Merr.] genotypes using GGE biplot. *Journal of Plant Breeding and Crop Science*, *12*(3), 246–257. https://doi.org/10.5897/JPBCS2020.0905

Mushoriwa, H., Mathew, I., Gwata, E. T., Tongoona, P., & Derera, J. (2022). Grain yield potential and stability of soybean genotypes of different ages across diverse environments in southern Africa. *Agronomy*, *12*(5), Article 1147. https://doi.org/10.3390/agronomy12051147

Mwiinga, B., Sibiya, J., Kondwakwenda, A., Musvosvi, C., & Chigeza, G. (2020). Genotype x environment interaction analysis of soybean (*Glycine max* (L.) Merrill) grain yield across production environments in Southern Africa. *Field Crops Research*, *256*, Article 107922. https://doi.org/10.1016/j.fcr.2020.107922

Nachilima, C., Cordoba-Novoa, H., Hoyos-Villegas, V., Chigeza, G., Manda, N., Lungu, D., & Kamfwa, K. (2021). Evaluation of soybean (*Glycine max* (L.) Merr.) breeding lines and genotype-by-environment interaction across production environments in Southern Africa. *Journal of Plant Breeding and Crop Science*, *13*(4), 203–215. https://doi.org/10.5897/JPBCS2021.0977

Olivoto, T., & Lúcio, A. D. C. (2020). metan: An R package for multi-environment trial analysis. *Methods in Ecology and Evolution*, *11*(6), 783–789. https://doi.org/10.1111/2041-210X.13384

Olivoto, T., Lúcio, A. D. C., da Silva, J. A., Marchioro, V. S., de Souza, V. Q., & Jost, E. (2019). Mean performance and stability in multi-environment trials I: Combining features of AMMI and BLUP techniques. *Agronomy Journal*, *111*(6), 2949–2960. https://doi.org/10.2134/agronj2019.03.0220

Pawlowski, M. L., Lagos-Kutz, D. M., Da Fonseca Santos, M., Lee, N., Chigeza, G., Nachilima, C., Francischini, J. H. M. B., & Hartman, G. L. (2021). Potential threat of bruchids on soybean production in Sub-Saharan Africa. *Plant Health Progress*, *22*(2), 86–91. https://doi.org/10.1094/PHP-10-20-0093-MR

Popović, V., Miladinović, J., Tatić, M., Đekić, V., Dozet, G., Đukić, V., & Grahovac, N. (2013). Stability of soybean yield and quality components. *African Journal of Agricultural Research*, *8*(45), 5651–5658. https://doi.org/10.5897/AJAR12.1146

Pour-Aboughadareh, A., Khalili, M., Poczai, P., & Olivoto, T. (2022). Stability indices to deciphering the genotype-by-environment interaction (G × E) effect: An applicable review for use in plant breeding programs. *Plants*, *11*(3), Article 414. https://doi.org/10.3390/plants11030414

Pratap, A., Gupta, S. K., Kumar, J., & Solanki, R. K. (2012). Soybean. In S. Gupta (Ed.), *Technological innovations in major world oil crops, Volume 1: Breeding* (pp. 293–321). Springer. https://doi.org/10.1007/978-1-4614-0356-2_12

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J., Lagacherie, P., Mcbratney, A. B., Mckenzie, N. J., Mendonça-Santos, M. D. L., Minasny, B., Montanarella, L., Okoth, P., Palm, C. A., Sachs, J. D., Shepherd, K. D., Vågen, T.-G., Vanlauwe, B., Walsh, M. G., … Zhang, G.-L. (2009). Digital soil map of the world. *Science*, *325*(5941), 680–681. https://doi.org/10.1126/science.1175084

Santos, M. (2019). Soybean cultivars in sub-Saharan Africa. *African Journal of Food, Agriculture, Nutrition and Development*, *19*(5), 15136–5139. https://doi.org/10.18697/ajfand.88.SILFarmDoc06

Šarčević, H., Iljkić, D., Andrijanić, Z., Sudarić, A., Gunjača, J., Varnica, I., Rastija, M., Godena, S., Maričević, M., Stepinac, D., & Pejić, I. (2022). Stability of protein and oil content in soybean across dry and normal environments—A case study in Croatia. *Agronomy*, *12*(4), Article 915. https://doi.org/10.3390/agronomy12040915

Siamabele, B. (2021). The significance of soybean production in the face of changing climates in Africa. *Cogent Food & Agriculture*, *7*(1), Article 1933745. https://doi.org/10.1080/23311932.2021.1933745

Sinclair, T. R., Marrou, H., Soltani, A., Vadez, V., & Chandolu, K. C. (2014). Soybean production potential in Africa. *Global Food Security*, *3*(1), 31–40. https://doi.org/10.1016/j.gfs.2013.12.001

Smith, A. B., Ganesalingam, A., Kuchel, H., & Cullis, B. R. (2015). Factor analytic mixed models for the provision of grower information from national crop cultivar testing programs. *Theoretical and Applied Genetics*, *128*, 55–72. https://doi.org/10.1007/s00122-014-2412-x

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3: Genes, Genomes, Genetics*, *5*(10), 1999–2006. https://doi.org/10.1534/g3.115.019000

Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., Dong, F., Araya, S., Liu, J., Hyten, D., Pantalone, V., & Nelson, R. L. (2020). Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research. *The Plant Journal*, *104*(3), 800–811. https://doi.org/10.1111/tpj.14960

Sudarić, A., Šimić, D., & Vratarić, M. (2006). Characterization of genotype by environment interactions in soybean breeding programmes of southeast Europe. *Plant Breeding*, *125*(2), 191–194. https://doi.org/10.1111/j.1439-0523.2006.01185.x

Tesfaye, A., Githiri, M., Derera, J., & Debele, T. (2017). Genetic variability in soybean (*Glycine max* L.) for low soil phosphorus tolerance. *Ethiopian Journal of Agricultural Sciences*, *27*(2), 1–15.

Tolorunse, K. D., Gana, A. S., Bala, A., & Sangodele, E. A. (2018). Yield stability studies of soybean (*Glycine max* (L.) Merrill) under rhizobia inoculation in the savanna region of Nigeria. *Plant Breeding*, *137*(3), 262–270. https://doi.org/10.1111/pbr.12583

Tsindi, A., Eleblu, J. S. Y., Gasura, E., Mushoriwa, H., Tongoona, P., Danquah, E. Y., Mwadzingeni, L., Zikhali, M., Ziramba, E., Mabuyaye, G., & Derera, J. (2023). Analysis of population structure and genetic diversity in a Southern African soybean collection based on single nucleotide polymorphism markers. *CABI Agriculture and Bioscience*, *4*(1), Article 15. https://doi.org/10.1186/s43170-023-00158-2

Vaezi, B., Pour-Aboughadareh, A., Mohammadi, R., Mehraban, A., Hossein-Pour, T., Koohkan, E., Ghasemi, S., Moradkhani, H., & Siddique, K. H. M. (2019). Integrating different stability models to investigate genotype × environment interactions and identify stable and high-yielding barley genotypes. *Euphytica*, *215*, Article 63. https://doi.org/10.1007/s10681-019-2386-5

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., Bajjalieh, N., Li, Z., & Boerma, H. R. (2015). QTL for seed protein and amino acids in the Benning × Danbaekkong

soybean population. *Theoretical and Applied Genetics*, *128*, 839–850. https://doi.org/10.1007/s00122-015-2474-4

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180–185. https://doi.org/10.1002/wics.147

Wilson, R. F. (2004). Seed composition. *Soybeans: Improvement, Production, and Uses*, *16*, 621–677. https://doi.org/10.2134/agronmonogr16.3ed.c13

Yaklich, R. W., Vinyard, B., Camp, M., & Douglass, S. (2002). Analysis of seed protein and oil from soybean northern and southern region uniform tests. *Crop Science*, *42*(5), 1504–1515. https://doi.org/10.2135/cropsci2002.1504

Yan, W. (2001). GGEbiplot—A Windows application for graphical analysis of multienvironment trial data and other types of two-way data. *Agronomy Journal*, *93*(5), 1111–1118. https://doi.org/10.2134/agronj2001.9351111x

Yan, W., & Hunt, L. A. (2001). Interpretation of genotype × environment interaction for winter wheat yield in Ontario. *Crop Science*, *41*(1), 19–25. https://doi.org/10.2135/cropsci2001.41119x

Yan, W., Nilsen, K. T., & Beattie, A. (2023). Mega-environment analysis and breeding for specific adaptation. *Crop Science*, *63*(2), 480–494. https://doi.org/10.1002/csc2.20895

Yan, W., & Tinker, N. A. (2006). Biplot analysis of multi-environment trial data: Principles and applications. *Canadian Journal of Plant Science*, *86*(3), 623–645. https://doi.org/10.4141/P05-169

Yau, S. K. (1995). Regression and AMMI analyses of genotype × environment interactions: An empirical comparison. *Agronomy Journal*, *87*(1), 121–126. https://doi.org/10.2134/agronj1995.00021962008700010021x

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zobel, R. W., Wright, M. J., & Gauch, H. G., Jr. (1988). Statistical analysis of a yield trial. *Agronomy Journal*, *80*(3), 388–393. https://doi.org/10.2134/agronj1988.00021962008000030002x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: De Meyer, E., Prenger, E., Mahmood, A., da Fonseca Santos, M., Chigeza, G., Song, Q., Mwadzingeni, L., Mukaro, R., Chibanda, M., Mabuyaye, G., Diers, B., & Scaboo, A. (2024). Evaluating genetic diversity and seed composition stability within Pan-African Soybean Variety Trials. *Crop Science*, *64*, 3272–3292. https://doi.org/10.1002/csc2.21356