ARTICLE

Boosting soybean adaptation in Africa through multi-trait selection in multi-environment trials

Crop Breeding and Applied Biotechnology
26(1): e54982618, 2026
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332026v26n1a8

Yasmim Isabel Retore[1], Bruno Figueiró Fregonezi[1], João Paulo Silva Pavan[1], Godfree Chigeza[2], Erica Pontes Leles[3], Michelle F. Santos[3], Peter Goldsmith[3], Brian Diers[3], Mauricio dos Santos Araújo[1] and José Baldin Pinheiro[1]*

Abstract: *Soybean breeding in Africa is constrained by environmental heterogeneity and strong genotype-by-environment (G×E) interaction, which limits the identification of high-performing and stable genotypes. We evaluated 145 soybean genotypes across 24 environments in seven African countries for grain yield (GY), number of days to maturity (NDM), 100-grain weight (W100), oil content (OIL), and protein content (PROT). Likelihood ratio tests indicated significant genotype, environment, and G×E effects for all traits. GY ranged from 1.404 to 3.391 kg ha^{-1}, with variation in maturity and grain composition associated with altitude and cropping season. Genetic correlations revealed important trade-offs, including positive associations among GY, NDM, and W100, and antagonism between OIL and PROT. Multi-trait selection with 10% intensity identified 14 genotypes closest to the ideotype, with expected gains of 15.7% in GY and favorable responses in the other traits. These findings demonstrate the effectiveness of multi-trait, multi-environment strategies for soybean improvement in Africa.*

Keywords: Glycine max *(*L.*)* Merr.*, selection index, genotype × environment interaction, multi-environment trials*

INTRODUCTION

Soybean (*Glycine max* (L.) Merrill) is one of the world's leading plant-based protein sources (≈40%), playing a central role in global food security and in supplying protein for both human consumption and animal feed (Messina 2022, Qin et al. 2022). The rising demand for plant-derived protein ingredients has intensified the need to develop cultivars that combine high yield with elevated protein content, thereby supporting value chains that rely on grains with greater nutritional and industrial value (Assefa et al. 2018). Nevertheless, breeding for higher protein concentration remains challenging because of the well-known negative correlations among protein, grain yield, and oil content, a classic trade-off in soybean improvement (Patil et al. 2017, Taliercio et al. 2024).

Analyses of long-term datasets and extensive multi-location trials consistently show that high-yielding cultivars exhibit lower grain protein concentration (Assefa et al. 2018, Arce et al. 2025). As a result, breeding approaches that simultaneously improve both grain quantity and grain quality remain a strategic bottleneck in soybean improvement programs (Patil et al. 2017, Taliercio et

***Corresponding author:**
E-mail: jbaldin@usp.br

Received: 18 December 2025
Accepted: 23 December 2025
Published: 13 January 2026

[1] Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Departamento de Genética, Avenida Pádua Dias, 11, 13418-900, Piracicaba, SP, Brazil
[2] International Institute of Tropical Agriculture Breeding for Tomorrow, Plot 1458B, Ngwerere Road, s/n, 310142, Chongwe District, Lusaka Province, Zambia
[3] University of Illinois Urbana-Champaign, Feed the Future Innovation Lab, 1301 West Gregory Drive, s/n, 61801, Urbana, IL, United States of America

al. 2024). Beyond balancing yield and grain quality, soybean breeders must also contend with pronounced genotype-by-environment (G×E) interactions, which undermine predictability and reduce the consistency of genotypes across locations and cropping seasons (Crossa 1990).

Evidence from multi-environment trials across major production regions demonstrates that the G×E variance component for yield and quality traits is substantial, particularly in tropical and subtropical environments, complicating the identification of broadly adapted genotypes (Obua et al. 2021, Abebe et al. 2024). In the African context, climatic and management heterogeneity across countries further intensifies G×E, underscoring the need for localized, tailored selection strategies (Araújo et al. 2025a). Consequently, well-designed multi-environment trial networks and robust statistical approaches to quantify G×E are essential for identifying genetic materials that combine high mean performance, stability, and desirable grain quality.

Linear mixed models have become the standard analytical framework for multi-environment trials, as they enable modeling of random effects and the explicit decomposition of G×E interactions, while providing high-precision Best Linear Unbiased Predictions (BLUPs) of genotypic values (Smith et al. 2005). Building on this foundation, more recent methodologies integrate both mean performance and stability into a single metric, facilitating selection in scenarios characterized by strong crossover interactions. Among these approaches, the Weighted Average Absolute Scores (WAASB) model combines interaction effects with BLUPs, producing a stability measure jointly associated with performance that has proven effective across multiple crops, including soybean (Olivoto et al. 2019a).

However, because modern soybean breeding requires simultaneous gains in yield, earliness, and grain quality, selection strategies must explicitly account for the antagonistic relationships among protein, oil, and yield described above. Yield, phenological cycle, and grain-quality attributes are tightly interconnected, combining favorable and unfavorable genetic correlations that can constrain the response to univariate selection. Under these conditions, multi-trait, simultaneous selection indices become essential tools because they integrate multiple traits into a single decision criterion, allowing breeders to weight traits jointly, balance gains among them, and minimize undesirable changes in correlated traits, even in the presence of strong G×E (Hazel 1943, Pešek and Baker 1969, Ambrósio et al. 2024, Ridara et al. 2025). Among the available approaches, indices such as the Multi-Trait Stability Index (MTSI) extend this framework to multi-environment trials by combining mean performance, stability, and the desired direction of improvement into a single ranking of genotypes with an overall desirable agronomic profile (Olivoto et al. 2019b).

As soybean expands into diverse regions of Africa, the Pan-African Soybean Variety Trials (PATs) provide a unique opportunity to evaluate a large number of genotypes under contrasting climatic and management conditions, generating robust information for recommending well-adapted cultivars. Despite this potential, few studies have jointly addressed crossover G×E, performance stability, and the simultaneous selection of grain-quality traits. In this context, combining linear mixed models with performance-stability metrics such as WAASB-BLUP and MTSI offers a practical strategy for addressing the trade-off between yield and grain quality. Therefore, the objectives of this study were to: i) select soybean genotypes with high performance and stability across multi-environment trials conducted in different African countries; and ii) identify ideotypes that harmonize yield, earliness, and grain quality, particularly protein and oil content, using the MTSI applied to PATs.

MATERIAL AND METHODS

Phenotypic data and field trials

Soybean genotype yield trials were sourced from the Soybean Innovation Lab (SIL) with the objective of identifying high-performance genotypes adapted to target population environments (TPEs) in African countries of origin, aiming to strengthen farmer production. This initiative led to the establishment of the PATs (Figure 1) in collaboration with the African Agricultural Technology Foundation, the Syngenta Foundation for Sustainable Agriculture, and the International Institute of Tropical Agriculture (Araújo et al. 2025b). A total of 145 soybean genotypes were evaluated for grain yield (GY), number of days to maturity (NDM), oil content (OIL), protein content (PROT), and 100-grain weight (W100) across the 2018–2021/22 growing seasons. Experiments were conducted in seven African countries: Malawi, Mali, Mozambique, Nigeria, Rwanda, Zambia, and Zimbabwe (Figure 2), with up to two trials per country in one or both seasons, totaling 24 environments (Table 1). Trials were arranged as randomized complete block designs (RCBD) with three replications.

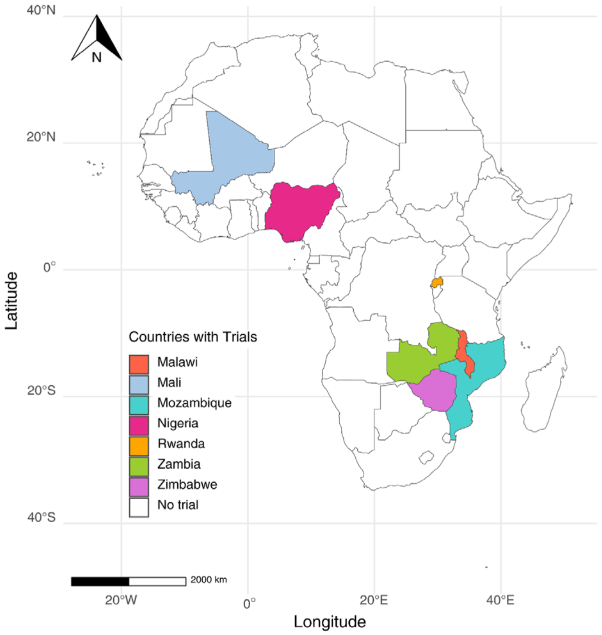

Figure 1. Countries where the trials were conducted (Malawi, Mali, Mozambique, Nigeria, Rwanda, Zambia, and Zimbabwe) between the 2018/19 and 2021/22 growing seasons, evaluating 145 soybean varieties for grain yield (GY), number of days to maturity (NDM), oil content (OIL), protein content (PROT), and 100-grain weight (W100).

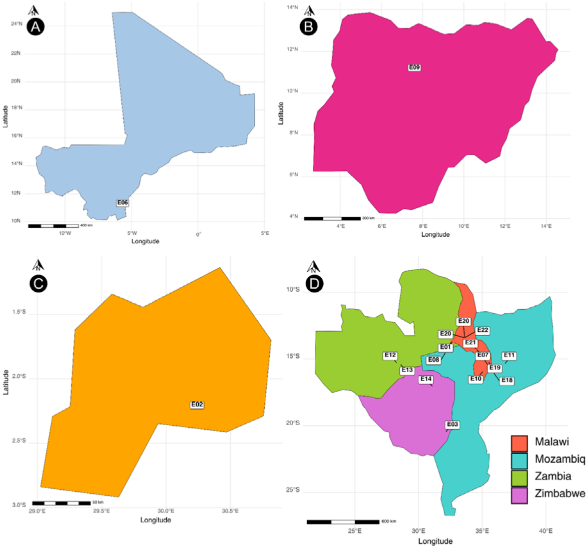

Figure 2. Geographic distribution of soybean experiments from the Pan-African Soybean Trials Network with available protein and oil data. The map displays the evaluated environments in Mali (A), Nigeria (B), Rwanda (C), Malawi, Mozambique, Zambia and Zimbabwe (D). Each environment, identified by a code from E01 to E24, represents a unique combination of trial, year, and season.

Each plot consisted of four 5-m rows sown at density of 25 plants m^{-1}, giving a gross area of 10 m^2; the harvest area comprised the two central rows. Agronomic management followed location-specific best practices for soybean production.

Statistical analyses

Variance components were estimated by restricted maximum likelihood (REML; Patterson and Thompson 1971). Genotypic values were predicted by best linear unbiased prediction (BLUP; Henderson 1949) under the experimental variance–covariance structure. All analyses were conducted using ASReml-R v4.1.2 (Butler et al. 2009) and the *metan* package (Olivoto and Lúcio 2020) within the R environment (R Core Team 2025).

For single-environment analyses, the following linear mixed model was fitted:

$$\boldsymbol{y} = \boldsymbol{\mu}\mathbf{1} + \boldsymbol{Xb} + \boldsymbol{Zg} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the vector of phenotypic observations ($n \times 1$); μ is the intercept; $\mathbf{1}$ is a vector of ones with dimension ($n \times 1$); $\boldsymbol{b}$ is the vector of fixed block effects, with incidence matrix $\boldsymbol{X}$ of dimension ($n \times p$), where p is the number of fixed-effect levels; $\boldsymbol{g}$ is the vector of random genotypic effects, assuming $\boldsymbol{g} \sim N(\mathbf{0}, \boldsymbol{I}\sigma^2_g)$, where σ^2_g is the genetic variance and I is an identity matrix; the incidence matrix $\boldsymbol{Z}$ has dimension ($n \times v$), with v being the number of genotypes. The residual errors are represented by $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{I}\sigma^2_e)$, where σ^2_e is the residual variance, with dimension $n \times 1$. This model was fitted independently for each environment and trait to estimate the genetic effects under each experimental condition.

The likelihood-ratio test (LRT; Wilks 1938) was performed to test the significance of the genotype, environment, and G×E interaction effects:

$$LRT = -2(LogL - LogL_f)$$

where L denotes the log-likelihood of the reduced model, and L_f corresponds to the log-likelihood of the whole model. The LRT value was compared against the Chi-square test (χ^2) distribution.

For the multi-environment trial analyses, the following mixed model was used:

$$\mathbf{y} = \mu\mathbf{1}_n + \mathbf{X}_1\,\mathbf{a} + \mathbf{X}_2\,\mathbf{b} + \mathbf{Z}g + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is the ($n \times 1$) vector of phenotypic observations; μ is the intercept; $\mathbf{1}_n$ is a vector of ones with dimension ($n \times 1$); a is the vector of fixed environmental effects ($a \times 1$), with incidence matrix $\mathbf{X}_1$ of dimension ($n \times a$); $\mathbf{b}$ is the vector of fixed block effects nested within environments ($pa \times 1$), with incidence matrix $\mathbf{X}_2$ of dimension ($n \times p$); $\mathbf{g}$ is the vector of random genotypic effects within environments, assuming $\mathbf{g} \sim MVN(\mathbf{0},\boldsymbol{\Sigma}_g \otimes \mathbf{I}_v)$, where $\boldsymbol{\Sigma}_g \otimes \mathbf{I}_v$ is the Kronecker product between the genetic covariance matrix among environments, with dimensions ($g \times g$), and $\mathbf{I}_v$, an identity matrix of order v, with v representing the number of genotypes. The incidence matrix $\mathbf{Z}$ has dimension ($n \times v$). ε is the vector of residual errors, assumed as $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0},\boldsymbol{\Sigma}_\varepsilon \otimes \mathbf{I}_n)$, where $\boldsymbol{\Sigma}_\varepsilon \otimes \mathbf{I}_n$ is the Kronecker product between the residual covariance matrix ($n \times n$) and $\mathbf{I}_n$, an identity matrix of dimension ($n \times n$). This model was fitted separately for each phenotypic trait evaluated (GY, NDM, OIL, PROT, and W100), allowing the estimation of genotypic effects and the G×E interaction across trial environments.

The broad-sense heritability was estimated as:

$$H_g^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_i^2 + \hat{\sigma}_e^2} \;,$$

where σ_g^2 is the genetic variance; σ_i^2 is the G× E interaction variance; and σ_e^2 is the residual variance. The environmental coefficient of variation (CV_e) was calculated using the following formula:

$$CV_e = \left(\frac{\sqrt{\hat{\sigma}_e^2}}{\mu}\right) \times 100,$$

where μ is the overall mean for the respective trait.

We estimated the genetic correlation ($\hat{r}_g$) between each pair of traits as follows:

$$\hat{r}_g = \frac{\hat{\sigma}_{gXY}}{\sqrt{\hat{\sigma}_{gX}^2 \hat{\sigma}_{gY}^2}} \;,$$

where $\hat{\sigma}_{gXY}$ is the genetic covariance between traits X and Y; $\hat{\sigma}_{gX}^2$ is the genetic variance of trait X; and $\hat{\sigma}_{gY}^2$ is the genetic variance of trait Y.

The difference of $\hat{r}_g$ from zero was verified using the Student's t-test (Student 1908), at a 5% probability level. The significance of the estimated genetic correlation was evaluated using Student's *t*-test (Student 1908) at the 5% probability level. Superior genotypes were selected using the Multi-trait Selection Index (MTSI; Olivoto et al. 2019a, b), applying a 10% selection intensity to increase grain yield (GY), oil content (OIL), protein content (PROT), and 100-grain weight (W100), while reducing days to maturity (NDM). The selection index is defined as a linear combination of standardized trait values:

$$I = b_1X_1 + b_2X_2 + \cdots + b_nX_n$$

where I is the selection index value; (X_1, X_2, … ,X_n) are the rescaled trait values; and (b_1, b_2, … ,b_n) are the weights assigned to each trait according to the MTSI procedure. Trait interrelationships were accounted for by factor analysis, and genotype ranking was based on the Euclidean distance to the ideotype in the factor space. Genotypes with smaller distances to the ideotype were considered superior.

RESULTS AND DISCUSSION

Malawi and Zimbabwe each accounted for 37.5% of the trials, followed by Zambia with 8.33%. The remaining 16.67% corresponded to one environment per country in the other four countries (Table 1). Overall, 41.67% of the experiments were conducted during the first cropping season (April to August), whereas 58.33% were conducted during the second season (September to January). The first season showed a higher mean GY than the second, with a difference of 537.63 kg ha^{-1}. For NDM, OIL, PROT, and W100, the mean values between seasons were similar.

The LRT indicated highly significant effects ($p < 0.001$) of genotype (G), environment (E), and the G × E interaction for all traits evaluated (Table 2). These results demonstrate substantial genetic variability among genotypes, strong contrasts among environments, and differential genotypic responses across sites and seasons. The effect of blocks within environments was also significant ($p < 0.05$), suggesting residual within-site heterogeneity and justifying its inclusion in the model. The significance of the G × E component confirms the re-ranking of genotypic performance across environments, a pattern already reported in multi-environment soybean networks (Araújo et al. 2022, Abebe et al. 2024, Stella et al. 2025). It supports the use of mixed models and stability metrics in subsequent selection steps.

The overall means for GY, NDM, OIL, PROT, and W100 were 2600.0 kg ha^{-1}, 97.2 days, 20.8%, 41.3%, and 17.7 g, respectively (Table 1). Across environments (E), GY ranged from 1404.96 kg ha^{-1} in E03 to 3391.23 kg ha^{-1} in E15, while the maturity period varied from 81 days (E11) to 127 days (E23). The observed variation is directly associated with

Table 1. Summary of the mean values for grain yield (GY, kg·ha^{-1}), oil content (OIL, %), protein content (PROT, %), 100-grain weight (W100, g) and number of days to maturity (NDM, days) for 145 soybean varieties evaluated across 24 trials distributed throughout the African continent

Countries	Environments	Years	Seasons	Traits				
				GY	OIL	PROT	W100	NDM
Malawi	E01	2018	1	2922.34	22.05	41.03	16.54	115.43
	E07	2019	1	3264.01	22.48	40.60	18.03	95.59
	E08	2019	1	2869.76	20.64	40.96	18.03	126.77
	E10	2019	2	1429.17	19.87	40.47	16.37	111.52
	E18	2020	1	2914.89	21.44	43.25	18.93	109.18
	E19	2020	1	2701.37	21.45	41.42	18.59	98.63
	E20	2020	1	3296.96	21.45	41.27	18.33	127.16
	E21	2021	1	2834.19	21.88	42.25	19.35	117.16
	E22	2021	1	2965.99	21.16	40.95	18.22	129.25
Mali	E06	2019	1	2866.67	21.03	43.92	15.18	97.18
Mozambique	E11	2019	2	1791.35	19.70	41.55	16.67	81.03
Nigeria	E09	2019	1	1756.79	21.27	40.71	15.75	110.36
Rwanda	E02	2018	2	2058.92	20.15	42.20	13.22	131.67
Zambia	E12	2019	2	2196.59	19.90	41.06	15.44	102.56
	E13	2019	2	2843.52	20.19	40.11	21.17	106.91
Zimbabwe	E03	2018	2	1404.96	22.53	38.23	16.77	133.86
	E04	2018	2	1451.08	22.27	40.72	17.37	120.45
	E05	2018	2	2955.90	20.92	41.59	16.04	106.17
	E14	2019	2	1556.37	20.47	39.50	18.44	106.88
	E15	2019	2	3391.23	19.67	41.08	17.75	123.94
	E16	2019	2	3218.11	18.61	41.71	19.15	110.94
	E17	2019	2	2634.28	18.66	41.64	17.11	129.04
	E23	2022	2	3364.14	20.23	39.89	20.89	136.72
	E24	2022	2	3314.95	18.23	43.03	15.64	131.96
Overall mean				2600.00	20.80	41.30	17.70	114.00

Table 2. Summary of the likelihood ratio test for grain yield (GY, kg·ha^{-1}), number of days to maturity (NDM, days), oil content (OIL, %), protein content (PROT, %), and 100-grain weight (W100, g) in 145 soybean varieties evaluated across 24 trials distributed throughout the African continent

Traits	LogL			LRT			χ^2		
	G	E	G×E	G	E	G×E	G	E	G×E
GY	-10629.44	-10810.63	-10724.75	22.79	385.16	213.40	$1.81e^{-06}$	$1.99e^{-67}$	$2.48e^{-48}$
OIL	-592.15	-504.80	-558.31	325.77	151.04	258.07	$8.03e^{-73}$	$8.11e^{-21}$	$4.53e^{-58}$
PROT	-1342.42	-1293.46	-1361.77	193.67	95.73	232.34	$5.03e^{-44}$	$7.60e^{-11}$	$1.84e^{-52}$
W100	-2141.15	-1991.47	-2008.83	421.14	121.78	156.50	$1.37e^{-93}$	$1.98e^{-15}$	$6.58e^{-36}$
NDM	-3561.22	-3608.09	-3652.61	202.64	296.39	385.44	$5.53e^{-46}$	$2.46e^{-49}$	$8.14e^{-86}$

Likelihood Ratio Test (LRT); Log-Likelihood test (LogL); p-values from the Chi-square test (χ^2); genotype effect (G), environment effect (E), and genotype × environment interaction (G×E).

differences in altitude among trial sites: higher-altitude environments exhibit lower mean temperatures, which prolong development and increase the NDM, whereas lower altitudes shorten the cycle (Oliveira et al. 2006, Alliprandini et al. 2009). In soybeans grown in tropical regions, this effect of altitude tends to be more pronounced than that of latitude (Lubis et al. 2021, Stella et al. 2025), which is consistent with the pattern observed in this study, where higher-altitude environments included some of the most productive sites (e.g., E15, 1341 m). In contrast, low-altitude environments yielded less (e.g., E03, 421 m).

Mean OIL ranged from 18.23% (E18) to 22.54% (E03), while PROT varied from 38.23% (E03) to 43.92% (E06) (Table 1). Notably, environment E18 combined the lowest OIL with one of the highest PROT values (43.35%), indicating that specific environmental conditions can shift the balance between these two traits. These contrasts are consistent with the strong environmental influence on the genes regulating oil and protein synthesis during grain filling and maturation (Duan et al. 2023) and anticipate the trade-offs discussed in the genetic correlations. For W100, the values ranged from 13.22 g (E02) to 21.18 g (E13), reflecting differences in assimilate availability during grain filling and contributing to the heterogeneity observed among environments.

The pattern observed between the coefficient of variation (CV) and broad-sense heritability (H_g^2) was negative: more precise environments exhibited higher H_g^2 (Figure 3). GY displayed the highest and most dispersed CVs, ranging from 9.36% to 28.45%, reflecting its intense sensitivity to environmental variation. In contrast, W100 exhibited lower CV values, ranging from 4.98% to 13.59%. The CVs for NDM and OIL were low, ranging from 0.23% to 6.22% and from 1.34% to 6.30%, respectively. PROT recorded the lowest CVs, ranging from 1.20% to 3.24%, indicating greater experimental precision. H_g^2 showed that a substantial proportion of the phenotypic variance was attributable to genetic variation (Table 3), a key condition for achieving efficient selection and cumulative genetic gains in breeding programs conducted in heterogeneous tropical environments (Pour-Aboughadareh et al. 2022).

The genetic correlation analysis revealed a moderate positive and different from zero by the Student's t-test association between GY and NDM ($\hat{r}_g = 0.39$), indicating that higher-yielding genotypes tend to exhibit longer cycles, consistent with previous results in soybean (Diers et al. 2018). This pattern shows that selection based exclusively on yield tends to shift

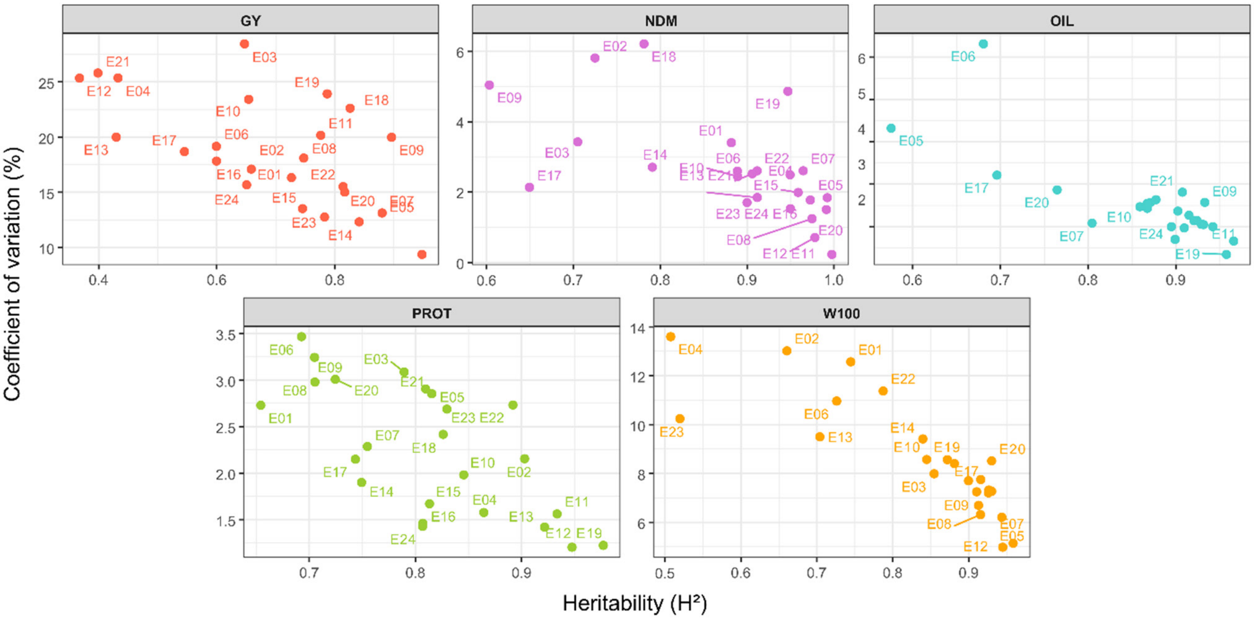

Figure 3. Biplot of the coefficient of variation and heritability for 24 environments for grain yield (GY, kg·ha⁻¹), number of days to maturity (NDM, days), oil content (OIL, %), protein content (PROT, %), and 100-grain weight (W100, g) in 145 soybean varieties evaluated across 24 trials distributed throughout the African continent.

Table 3. Response to selection based on the Multi-trait Selection Index (MTSI) for 145 soybean genotypes evaluated across 24 environments (2018–2021/22), including broad-sense heritability (H^2_g), environmental coefficient of variation (CV_e), original population mean (Xo), mean of the selected lines (Xs), selection differential (SD), and predicted selection gain (SG) for grain yield (GY), number of days to maturity (NDM), oil content (OIL), protein content (PROT), and 100-grain weight (W100) in trials conducted in seven African countries

Trait	H^2_g	CV (%)	X_o	X_s	SD	SD (%)	SG	SG (%)
GY	0.83	19.00	2600.00	3074.00	489.50	18.94	406.80	15.74
NDM	0.97	3.19	114.00	108.8	-7.17	-6.18	-6.95	-5.99
OIL	0.97	2.51	20.80	21.22	0.59	2.87	0.58	2.80
PROT	0.96	2.29	41.30	41.92	0.28	0.68	0.27	0.65
W100	0.97	8.60	17.70	19.36	2.27	13.30	2.22	13.01

the breeding program toward later-maturing materials, which may be undesirable in several production systems. The correlation between GY and OIL was positive ($\hat{r}_g$ = 0.24) and significant, suggesting that higher-yielding genotypes tend to have higher oil content. Similarly, OIL showed a significant and negative correlation with PROT ($\hat{r}_g$ = -0.36), a pattern widely reported in oilseed crops (Finoto et al. 2021, Jin et al. 2023, Abdelghany et al. 2025) and consistent with carbon competition between lipid and protein synthesis pathways during grain filling (Li et al. 2024, Mo et al. 2024).

In this work, PROT showed a negative and different from zero correlation with W100, indicating that heavier seeds tend to have lower protein content, corroborating observations by Wang et al. (2020) and Duan et al. (2023). In contrast, W100 was positively and significantly correlated with OIL ($\hat{r}_g$ = 0.32) and with NDM ($\hat{r}_g$ = 0.26), suggesting that larger seeds are associated with higher oil content and longer cycles (Duan et al. 2023). Finally, no linear correlations were detected between GY and PROT, or amid NDM and OIL, or between NDM and PROT (Figure 4), indicating relative genetic independence between these pairs of traits and creating opportunities for more targeted gains via multi-trait selection.

The observed correlation patterns among GY, NDM, OIL, PROT, and W100, together with strong G×E and environmental heterogeneity, indicate that univariate selection would tend to increase losses in at least one of the target traits. In this context, the use of a simultaneous index became decisive for this dataset. The MTSI was therefore applied to combine mean performance, stability, and the desired direction of response into a single metric, generating a ranking of the 145 genotypes according to their proximity to the ideotype (Figure 5) and serving as the basis for selecting the genotypes (Olivoto et al. 2019b).

The application of the MTSI with a selection intensity of 10% resulted in the selection of 14 genotypes (G114, G077, G052, G110, G080, G142, G124, G039, G005, G058, G036, G143, G050, and G049) (Figure 5). These genotypes showed the lowest MTSI values, exhibiting greater proximity to the ideotype and consistently above-average performance across multiple environments. The expected response to selection indicated a 15.74% gain in GY (≈ 406.8 kg ha^{-1}) and a 6% reduction in NDM, corresponding to about seven fewer days in the cycle. For quality traits, gains were more modest but still positive: OIL showed a relative increase of 2.80% (0.58 percentage point), PROT increased by 0.27%

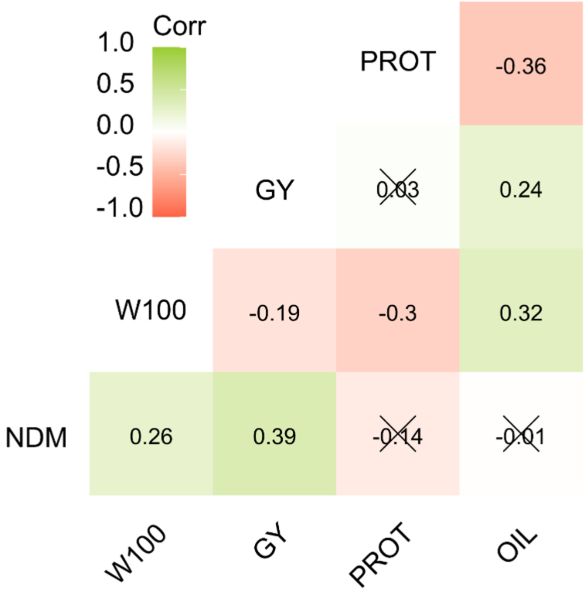

Figure 4. Genetic correlations (Corr) among grain yield (GY, kg·ha^{-1}), number of days to maturity (NDM, days), oil content (OIL, %), protein content (PROT, %), and 100-grain weight (W100, g) in 145 soybean varieties evaluated across 24 trials conducted in the African continent. The color gradient represents the strength and direction of the correlations: red indicates a strong negative correlation (−1), white indicates no correlation (0), and green indicates a strong positive correlation (+1). × indicates a non-significant correlation at 95% confidence according to the Student's t-test.

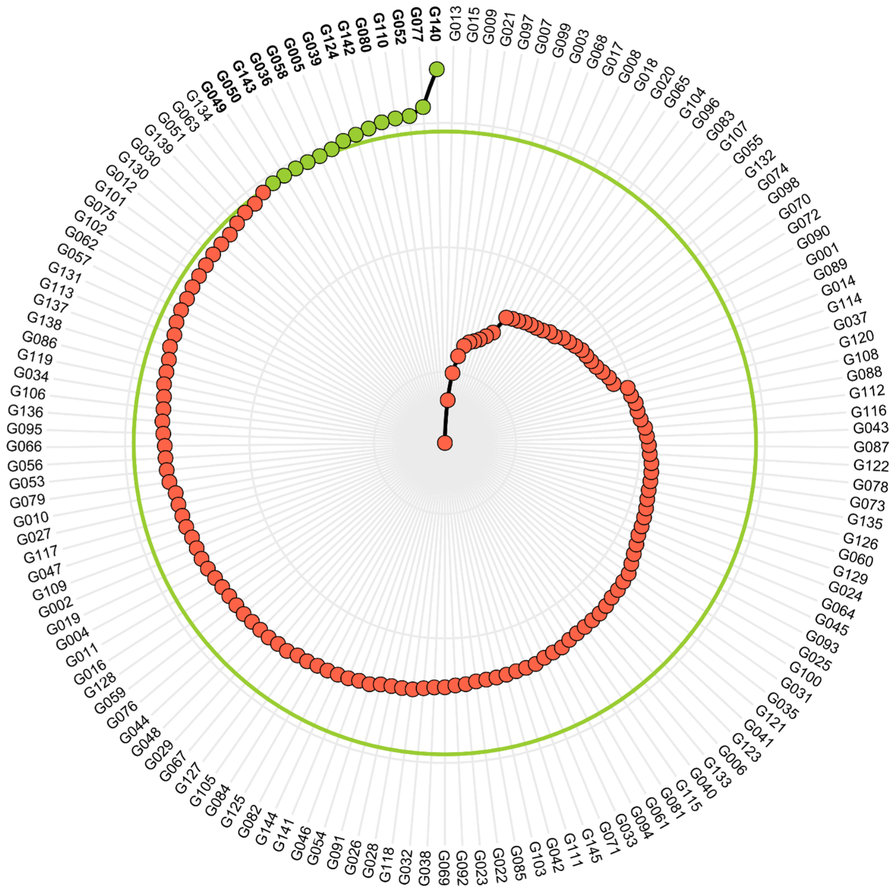

Figure 5. Classification of 145 soybean genotypes evaluated in 24 trials across the African continent according to the Multi-Trait Stability Index (MTSI). Genotypes highlighted in green and positioned above the green threshold line represent the 10% selected under the applied selection pressure.

(0.65 percentage point), and W100 increased by 13.01% (2.22 g). Taken together, these results show that it is possible to increase yield potential without losses in oil and protein, highlighting the essential role of multi-trait selection indices, such as MTSI, in reconciling grain yield and quality in soybean multi-environment trial networks. This ability to balance responses is one of the main advantages of multi-trait indices in large trial networks (Ambrósio et al. 2024, Ridara et al. 2025).

The trials conducted in seven African countries revealed substantial firm-level environmental heterogeneity and strong G×E interactions for yield, phenological cycle, and grain quality, underscoring the challenge of identifying stable genotypes in tropical and subtropical environments. The MTSI allowed the integration of multiple traits and the selection of a group of genotypes with consistent performance, resulting in gains greater than 15% in yield, a reduction of seven days in maturity, and moderate improvements in grain technological quality, highlighting G114, G077, and G052 as promising materials. Nevertheless, limitations such as incomplete environmental coverage, moderate genetic variability for protein, the negative correlation between oil and protein, and the absence of standardized environmental covariates restrict broader gains. As future directions, incorporating genomic prediction, envirotyping, and validation in new seasons is expected to increase selection accuracy, refine the definition of recommendation zones, and accelerate the delivery of more productive cultivars adapted to African agricultural conditions.

ACKNOWLEDGEMENTS

We thank the Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo (ESALQ/USP), for its support. We also acknowledge the Coordination for the Improvement of Higher Education Personnel (CAPES) and the São Paulo Research Foundation (FAPESP, grant 2024/01868-9) for providing scholarships.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current research are available from the corresponding author upon reasonable request.

REFERENCES

Abdelghany AM, Zhang S, Li J, Li B and Sun J (2025) Seed phenotype and maturity groups as determinants of protein, oil, and fatty acid composition patterns in diverse soybean germplasm. **BMC Plant Biology 25**: 1189.

Abebe AT, Adewumi AS, Adebayo MA, Shaahu A, Mushoriwa H, Alabi T, Derera J, Agbona A and Chigeza G (2024) Genotype × environment interaction and yield stability of soybean (*Glycine max* L.) genotypes in multi-environment trials (METs) in Nigeria. **Heliyon 10**: e38097.

Alliprandini LF, Abatti C, Bertagnolli PF, Cavassim JE, Gabe HL, Kurek A, Matsumoto MN, Oliveira MAR, Pitol C, Prado LC and Steckling C (2009) Understanding soybean maturity groups in Brazil: environment, cultivar classification, and stability. **Crop Science 49**: 801-808.

Ambrósio M, Daher RF, Santos RM, Santana JGS, Vidal AKF, Nascimento MR, Leite CL, Souza AG, Freitas RS, Stida WF, Farias JEC, Souza Filho BF, Melo LC and Santos PR (2024) Multi-trait index: selection and recommendation of superior black bean genotypes as new improved varieties. **BMC Plant Biology 24**: 525.

Araújo M, Aragão WFL, Santos SP, Freitas TKT, Saraiva VC, Damasceno-Silva KJ, Dias LAS and Rocha MM (2022) Evaluation of adaptability and stability for iron, zinc and protein content in cowpea genotypes using GGE biplot approach. **Heliyon 8**: e11832.

Araújo MS, Chaves S, Ferreira GNC, Chigeza G, Leles EP, Santos MF, Diers BW, Goldsmith P and Pinheiro JB (2025b) High-resolution soybean trial data supporting the expansion of agriculture in Africa. **Scientific Data 12**: 1908.

Araújo MS, Pavan JPS, Stella AA, Fregonezi BF, Lima NF, Leles EP, Santos MF, Goldsmith P, Chigeza G, Diers BW and Pinheiro JB (2025a) Optimizing soybean variety selection for the Pan-African trial network using factor analytic models and envirotyping. **Frontiers in Plant Science 16**: 1594736.

Arce MSZ, Tagliapietra ELL, Winck JEM, Alves AF, Dalla Porta FS, Facco TB, Streck NA, Soares MF, Ferrao GDE, Debona D, Costa CHM, Bega RM, Fochesatto E, Krabbe EL and Zanon AJ (2025) Assessing genetics, biophysical, and management factors related to soybean seed protein variation in Brazil. **European Journal of Agronomy 165**: 127541.

Assefa Y, Bajjalieh N, Archontoulis S, Casteel S, Davidson D, Kovács P, Naeve S and Ciampitti IA (2018) Spatial characterization of soybean yield and quality (amino acids, oil, and protein) for United States. **Scientific Reports 8**: 14653.

Butler DG, Cullis BR, Gilmour AR and Gogel BJ (2009) ASReml-R reference manual. Available at <https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/ASReml-R-Reference-Manual-4.2.pdf>. Accessed on December 18, 2025.

Crossa J (1990) Statistical analysis of multi-location trials. **Advances in Agronomy 44**: 55-85.

Diers BW, Specht J, Rainey KM, Cregan P, Song Q, Ramasubramanian V, Graef G, Nelson R, Schapaugh W, Wang D, Shannon G, McHale L, Kantartzi SK, Xavier A, Mian R, Stupar RM, Michno JM, An YQC, Goettel W, Ward R, Fox C, Lipka AE, Hyten D, Cary T and Beavis WD (2018) Genetic architecture of soybean yield and agronomic traits. **G3: Genes, Genomes, Genetics 8**: 3367-3375.

Duan Z, Li Q, Wang H, He X and Zhang M (2023) Genetic regulatory networks of soybean seed size, oil and protein contents. **Frontiers in Plant Science 14**: 1160418.

Finoto EL, Soares MBB, Albuquerque JAA, Monteiro Neto JLL, Maia SS, Dona S, Chagas EA, Soares-da-Silva E and Abanto-Rodriguez C (2021) Selection of soybean genotypes for yield, size, and oil and protein contents. **Australian Journal of Crop Science 15**: 48-50.

Hazel LN (1943) The genetic basis for constructing selection indexes. **Genetics 28**: 476-490.

Henderson CR (1949) Estimation of changes in herd environment**. Journal of Dairy Science 32**: 706-706.

Jin H, Yang X, Zhao H, Song X, Tsvetkov YD, Wu Y, Gao Q, Zhang R and Zhang J (2023) Genetic analysis of protein content and oil content in soybean by genome-wide association study. **Frontiers in Plant Science 14**: 1182771.

Li H, Sun J, Zhang Y, Wang N, Li T, Dong H, Yang M, Xu C, Hu L, Liu, Chen Q, Foyer C and Qi ZC (2024) Soybean oil and protein: Biosynthesis, regulation and strategies for genetic improvement. **Plant, Cell & Environment 1**: 1-17.

Lubis I, Atmaja ISF, Koesoemaningtyas T, Ghulamahdi M, Purnamawati H and Shiraiwa T (2021) Growth performance and productivity of several soybean genotypes (*Glycine max* L. Merr.) cultivated at high altitude areas in Indonesia. **Journal of Tropical Crop Science 8**: 1-7.

Messina M (2022) Perspective: Soybeans can help address the caloric and protein needs of a growing global population. **Frontiers in Nutrition 9**: 909464.

Mo W, Wang P, Shi Q, Zhao X, Zheng X, Ji L, Zhang L, Geng M, Wang Y, Wang R, Bian M, Meng X and Zuo Z (2024) Uncovering key genes associated with protein and oil in soybeans based on transcriptomics and proteomics. **Industrial Crops and Products 222**: 119981.

Obua T, Sserumaga JP, Awio B, Nganga F, Odong TL, Tukamuhabwa P, Tusiime G and Nabasirye M (2021) Multi-environmental evaluation of protein content and yield stability among tropical soybean genotypes using GGE biplot analysis. **Agronomy 11**: 1265.

Oliveira AB, Duarte JB, Chaves LJ and Couto MA (2006) Environmental and genotypic factors associated with genotype by environment interactions in soybean. **Crop Breeding and Applied Biotechnology 6**: 79-86.

Olivoto T and Lúcio AD (2020) metan: An R package for multi-environment trial analysis. **Methods in Ecology and Evolution 11:** 783-789.

Olivoto T, Lúcio ADC, Silva JAG, Marchioro VS, Souza VQ and Jost E (2019a) Mean performance and stability in multi-environment trials I: combining features of AMMI and BLUP techniques. **Agronomy Journal 111**: 2949-2960.

Olivoto T, Lúcio ADC, Silva JAG, Sari BG and Diel MI (2019b) Mean performance and stability in multi-environment trials II: Selection based on multiple traits. **Agronomy Journal 111**: 2961-2969.

Patil G, Mian R, Vuong T, Pantalone V, Song Q, Chen P, Shannon GJ, Carter TC and Nguyen HT (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. **Theoretical and Applied Genetics 130**: 1975-1991.

Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika 58**: 545-554.

Pešek J and Baker RJ (1969) Desired improvement in relation to selection indices. **Canadian Journal of Plant Science 49**: 803-804.

Pour-Aboughadareh A, Khalili M, Poczai P and Olivoto T (2022) Stability indices to deciphering the genotype-by-environment interaction (GEI) effect: An applicable review for use in plant breeding programs. **Plants 11**: 414.

Qin P, Wang T and Luo Y (2022) A review on plant-based proteins from soybean: health benefits and soy product development. **Journal of Agriculture and Food Research 7**: 100265.

R Core Team (2025) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at <https://www.r-project.org/>. Accessed on December 2, 2025.

Ridara F, Ustari D, Wicaksono AA, Algina A, Amien S, Susanto GWA, Koerniati S, Hastilestari BR, Concibido V and Karuniawan A (2025) Integrating multi-trait selection and environmental stability for soybean (*Glycine max* L. Merr.) genotype improvement in Indonesia. **Plant, Cell & Environment 1**: 1-13.

Smith AB, Cullis BR and Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science 143**: 449-462.

Stella AA, Pavan JPS, Araújo MS, Fregonezi BF, Unzimai IV, Leles EP, Santos MF, Goldsmith P, Chigeza G, Diers BW, Gathungu T, Njoroge J and Pinheiro JB (2025) Soybean selection in Kenya enhanced by multi-trait and genotype-by-environment interaction modeling. **Scientific Reports 15**: 27575.

Student (1908) The probable error of a mean. **Biometrika 6**: 1-25.

Taliercio E, Gillenwater J, Woodruff L and Fallen B (2024) *Glycine soja*, PI424025, is a valuable genetic resource to improve soybean seed-protein content and composition. **PLoS ONE 19**: e0310544.

Wang S, Liu S, Wang J, Yokosho K, Zhou B, Yu Y-C, Liu Z, Frommer WB, Ma JF, Chen LQ, Guan Y, Shou H and Tian Z (2020) Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. **National Science Review 7**: 1776-1786.

Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. **The Annals of Mathematical Statistics 9**: 60-62.