



Research article

Multi-Trait Evaluation of Soybean (*Glycine max* L.) Genotypes for Seed Yield, Oil Productivity, and Water Productivity under Semi-Arid Irrigated Conditions

Hussein Tuama Abbood

Department of Crop Science, College of Agriculture, Wasit University, Wasit 52001, Iraq

ARTICLE INFO

Keywords:

Glycine max; genotype evaluation; water use efficiency; oil productivity; multi-trait selection index; semi-arid irrigated agriculture; principal component analysis

Article history:

Received: 30 Jan 2026
Accepted: 15 Sep 2026
Online: 30 Sep 2026

ABSTRACT

Background: Soybean (*Glycine max* L.) is a globally important source of vegetable protein and oil, and genotype selection is a key lever for improving productivity and water productivity in irrigated semi-arid production systems. **Objective:** This study evaluated ten soybean genotypes for seed yield, oil yield, oil content, protein content, water use efficiency (WUE), and oil water productivity under semi-arid irrigated conditions, with irrigation water held constant across genotypes so that observed differences reflect genotypic performance rather than irrigation-treatment effects. **Methods:** The experiment followed a randomized complete block design (RCBD) with genotype as the treatment factor and replication as the blocking factor (10 genotypes $\times$ 3 replications = 30 experimental units). Analysis of variance (ANOVA) was performed for each trait, genotype means were separated using the least significant difference (LSD) test at $\alpha = 0.05$, and genotype relationships were further examined using Pearson correlation, principal component analysis (PCA), Ward hierarchical clustering, and a weighted multi-trait selection index. **Results:** Genotype effects were highly significant for seed yield ($F = 118.90$, $P = 3.568 \times 10^{-14}$, $CV = 3.85\%$), oil yield ($F = 31.62$, $P = 3.116 \times 10^{-9}$, $CV = 8.69\%$), and WUE ($F = 151.68$, $P = 4.165 \times 10^{-15}$, $CV = 3.75\%$). Genotype G10 (an imported drought-tolerant genotype) recorded the highest seed yield (4.96 t ha^{-1}), oil yield (1.033 t ha^{-1}), WUE (0.954 kg m^{-3}), and multi-trait selection index score (93.3 of 100), followed by G9 (a local genotype from Abu Ghraib) and G8. **Conclusion:** Under semi-arid irrigated conditions with constant water supply, genotype identity was the principal driver of productivity and water-productivity differences among the ten soybean genotypes evaluated, and G10, G9, and G8 emerged as the most promising genotypes for seed yield, oil productivity, and water productivity.

1. Introduction

Soybean (*Glycine max* L) is the world's leading oilseed crop and a major source of protein and vegetable oil in the global food and feed industry, closely linked to livestock and human nutrition as well as bio-based raw material in industrial applications (Guo et al., 2022). Soybean production in the world is projected to increase beyond 420 million metric tons for the 2024/25 marketing year in

✉ *Corresponding author:
hutuama@uowasit.edu.iq (Hussein Tuama Abbood)

response to the rising global demand for protein and edible oil. Soybean seed is an important source of protein (with 38–40% protein) and oil (with 19–22% oil), and both fractions are main focus in breeding programs, as protein content is often negatively correlated with oil content, a correlation, in part, attributed to antagonism between two related biosynthetic pathways for utilizing common carbon and energy (Araujo et al., 2026).

In semi-arid production areas, irrigation serves to stabilize potential yield of soybean, although the degree to which potential yield is attained is highly dependent on the genotype. Considerable genotypic differences in seed yield and WUE have been reported among soybean genotypes tested under irrigated and water stress field conditions (Fried et al., 2019), and enhanced seed and water use efficiency (WUE) due to breeding has been reported in soybean cultivars adapted to dry and semi-dry growing environments (Feng et al., 2022). Since irrigation is an expensive and ever more scarce resource input in dryland farming, the identification of genotypes that simultaneously express high seed yield, oil yield and water use efficiency would be a great benefit to breeders, agronomists and farmers who work within the constraints of diminishing water supplies (Soliman et al., 2025).

However, evaluating genotypes in a single trait basis results incomplete for overall merit since seed yield, oil yield, protein content and water productivity do not necessarily co-vary in the same direction among genotypes. Multivariate methods such as principal component analysis (PCA) and hierarchical clustering are commonly executed to compress genotype-by-trait variation, allow identification of the traits which best differ genotypes, and cluster genotypes into performance groups (Odong et al., 2013). 2.1.2. Weighted multi-trait selection indices Generally, breeders need to select superior genotypes for multiple traits of economic or agronomic importance but often some traits are more important or carrying more weight than others in the selection process, and these can be combined into a single selection index rank to simplify the selection decisions.

Although these combined statistics have their merit, for semi-arid irrigated soybean genotype evaluation, even the most recent studies do not consider using ANOVA-based mean comparison with correlation analysis, PCA, Ward clustering and, as well, formal multi-trait selection index as a single evaluation scheme especially that for p and wp performance measures followed by oil and protein quality traits. This represents a gap in research: genotype testing trials in semi-arid, irrigated environments are typically published using univariate mean comparisons only, rather than a combined multivariate ranking which breeders can immediately apply to make selection decisions.

It should be emphasized the fact that in the current study all genotypes and replications were equally subjected to irrigation; while the experiment was not conducted as a drought or watering-treatment experiment, and no contrast between water regiments was conducted. Thus, the assumptions should be interpreted strictly as differences between genotypes in productivity, oil productivity and water productivity under one constant semi-arid irrigated water regimen, and not as any implication of the ability to resist drought or respond to drought-treatments (Ruttanaprasert et al., 2025).

The purpose of this study was to assess ten soybean genotypes for seed, oil and protein contents, seed and oil yield, water use efficiency (WUE) and oil water productivity (OWP) under semi-arid irrigated conditions based on randomized complete block design (RCBD) and to combine ANOVA-LSD mean comparison with correlation analysis, PCA, Ward hierarchical clustering and a weighted multi-trait selection index to reveal the genotypes exhibiting the maximum potential for productivity and water productivity. It was hypothesized that genotype would have a significant effect for all productivity and water-productivity related traits as well as that multivariate ranking would identify a consistent subset of superior genotypes among the various analytical approaches.

2. Material and Method

2.1 Experimental Design and Treatments

The experiment was carried out as a randomized complete block design (RCBD), with soybean genotype as the treatment and replication as the block (Gomez & Gomez, 1984), the experimental design and analysis were in accordance with the standard field-experiment for agricultural research. Ten soybean genotypes (designated G1- G10; Table 3) were tested in three replicates, resulting in a total of 30 experimental units (10 genotypes $\times$ 3 replications). Genotypes were randomly assigned to plots within each replication block. Water for irrigation was applied evenly and consistently among all genotypes and replications during the growing season; therefore, no irrigation or drought treatment factor was considered in the experiment, and no comparison of water regimes was achieved. All the differences mentioned below are thus the effects of genotypes under a fixed irrigated semi-arid water regime.

2.2 Traits Measured

These data were collected for seed yield (t ha^{-1}), oil content (%), oil yield (t ha^{-1}), protein content (%), water use efficiency (WUE, kg m^{-3}), and oil water productivity (kg oil m^{-3}), in addition to associated agronomic and yield-component traits and seed-quality/oil-composition traits summarized in the pooled figures (Figures 2 and 3). Oil yield was calculated from seed yield and oil content, and WUE was defined as seed yield per unit of water applied (kg seed m^{-3} water); oil water productivity was defined parallel as oil yield per unit of water applied (kg oil m^{-3} water). Because oil water production and WUE are expressed per unit of the same fixed water input, variation among genotypes in these two traits represent differences among genotypes in the efficiency of seed and oil production rather than differences in the amount of water delivered.

2.3 Data Quality Control

The data (30 observations = 10 genotypes $\times$ 3 replications) were examined for internal consistency before analysis. In one case, correction was made to the seeds-per-plant value when the raw value disagreed with the product of the number of pods per plant and the number of seeds per pod. The fatty-acid compositions were said to be checked for their sum to 100%; however, 15 recorded fatty-acid profiles deviated from 100% by more than one percentage point and are treated here as major fatty acids only, until the confirmation of laboratory normalization procedures by the team collecting the original data. These quality-control procedures and their results are summarized in Table 1.

2.4 Statistical Analysis

A separate analysis of variance (ANOVA) suitable for RCBD was calculated for each quantitative character to test for the effect of genotype (Gomez & Gomez, 1984). When the effects of genotype were significant, means were separated based on the least significant difference (LSD) test at $\alpha = 0.05$. Pearson correlation coefficients were estimated between major yield, water-yield, and quality traits to investigate associations among traits. Principal component analysis (PCA) was applied to the multivariate trait data to project the data onto fewer orthogonal components that capture the primary axes of genotypic variation, and Ward's minimum-variance hierarchical clustering method (Ward,

1963) was used—following common recommendations of integrating PCA with hierarchical clustering to characterize genetic/phenotypic structure in genotype-evaluation datasets (Odong et al., 2013)—to identify groups of genotypes with similar performance. Finally, a weighted multi-trait selection index in the spirit of the classical discriminant-function selection index (Smith, 1936; Hazel, 1943) was derived from seed yield, oil yield and WUE and employed to generate an overall single composite rank (in the 0–100 index scale) of genotype merit. All statistical analyses were performed using $\alpha = 0.05$ as the conventional level of significance.

3. Results

3.1 Dataset and Quality-Control Summary

The dataset consists of 30 observations on 10 genotypes in 3 replications under RCBD. One of the seeds-per-plant values was corrected for internal inconsistency with pods per plant $\times$ seeds per pod and 15 records on fatty-acid composition showed sums deviating from 100% more than one percentage point and they are considered to major fatty acids only. The irrigation water was also verified to be the same for all experimental units; thus, the data enable comparison of genotypes under one water regime, and are not intended as testing of irrigation-treatment or drought-effect (Table 1).

Table 1. Dataset and quality-control summary.

QC Item	Value	Interpretation
Number of observations	30	30 observations = 10 genotypes $\times$ 3 replications
Number of genotypes	10	Ten soybean genotypes were evaluated
Number of replications	3	Three field blocks/replications; RCBD analysis applied
Seeds-per-plant corrected values	1	Corrected where raw values were inconsistent with pods per plant $\times$ seeds per pod
Fatty acid sums differing from 100% by >1 point	15	Interpreted as major fatty acids unless laboratory normalization is confirmed
Irrigation water variation	Constant	No irrigation-treatment or drought-effect comparison is claimed from these data

3.2 ANOVA and Genotype Effects

Genotype was significantly ($P < 0.0001$) for all six major publication traits (Table 2). Significance ($p < 0.001$) of seed yield ($F = 118.90$, $CV = 3.85\%$), oil yield ($F = 31.62$, $CV = 8.69\%$), WUE ($F = 151.68$, $CV = 3.75\%$) and oil water productivity ($F = 36.08$, $CV = 8.78\%$) and significance ($p < 0.01$) of protein content ($F = 4.65$, $CV = 2.10\%$) and ($p < 0.05$) of oil contents ($F = 3.00$, $CV = 5.24\%$) were observed. The precision of the experiment was considered high, revealed by the small values of CV for seed yield and WUE (both were less than 4%) for these two important productivity traits.

Table 2. ANOVA summary for key publication traits.

Trait	Mean	F (genotype)	P (genotype)	Sig.	CV (%)	LSD0.05
Seed yield (t ha ⁻¹)	3.677	118.90	3.568×10^{-14}	***	3.85	0.243
Oil content (%)	20.817	3.00	0.0227	*	5.24	1.870
Oil yield (t ha ⁻¹)	0.773	31.62	3.116×10^{-9}	***	8.69	0.115
Protein content (%)	38.460	4.65	0.0027	**	2.10	1.386
Water use efficiency (kg m ⁻³)	0.684	151.68	4.165×10^{-15}	***	3.75	0.044
Oil water productivity (kg oil m ⁻³)	0.144	36.08	< 0.0001	***	8.78	0.022

3.3 Genotype Mean Performance

The LSD test ($\alpha = 0.05$) categorized the ten genotypes into some overlapping but mostly unique statistical groups for each characteristic (Table 3). The genotype G10 (Imported drought-tolerant genotype) obtained the highest seed yield of (4.96 t ha⁻¹, grouping letter 'a') and also placed at the first statistical group in oil yield, WUE and oil water productivity together with G8 and G9 for these traits. Genotype G1 (Hassan) produced the least seed yield (2.367 t ha⁻¹, grouping letter 'g') and the lowest oil yield, WUE and oil water productivity but the highest protein content (39.600%) in number.

Table 3. Major genotype means with LSD grouping letters.

Genotype	Name	Seed yield (t ha ⁻¹)	Oil yield (t ha ⁻¹)	WUE (kg m ⁻³)	Oil WP (kg oil m ⁻³)	Oil (%)	Protein (%)
G1	Hassan	2.367 g	0.458 d	0.424 h	0.082 e	19.333 b	39.600 a
G2	Lee 74	2.667 f	0.516 d	0.477 g	0.092 e	19.333 b	39.033 a
G3	QT5	2.867 f	0.573 c	0.519 g	0.104 d	20.000 b	39.100 a
G4	TN12	3.180 e	0.647 c	0.580 f	0.118 d	20.333 b	39.300 a
G5	Hawija 101	3.700 d	0.771 b	0.679 e	0.141 c	20.833 a	39.333 a
G6	Soybean Parents	3.700 d	0.784 b	0.685 e	0.145 c	21.167 a	38.867 a
G7	Energy 2	4.233 c	0.926 a	0.796 d	0.174 b	21.833 a	37.733 b
G8	QTLI	4.433 b	0.998 a	0.842 c	0.189 a	22.500 a	36.833 b
G9	Local genotype (Abu Ghraib)	4.667 b	1.028 a	0.886 b	0.195 a	22.000 a	37.467 b
G10	Imported drought-tolerant	4.960 a	1.033 a	0.954 a	0.199 a	20.833 a	37.333 b

3.4 Productivity and Water-Productivity Traits

Seed yield, oil yield, WUE, and oil water productivity followed a general pattern of ranking between genotypes, with an upward trend from G1 to G10 (Figure 1). Numerically, G10 was the highest for all four traits, followed by G9 and G8, while G1, G2, and G3 occupied the bottom statistical classes throughout. This uniformity in four productivity-is-linked traits suggests that the ranking among genotypes would be reasonably robust and would not simply be a by-product caused by any one definition of trait.

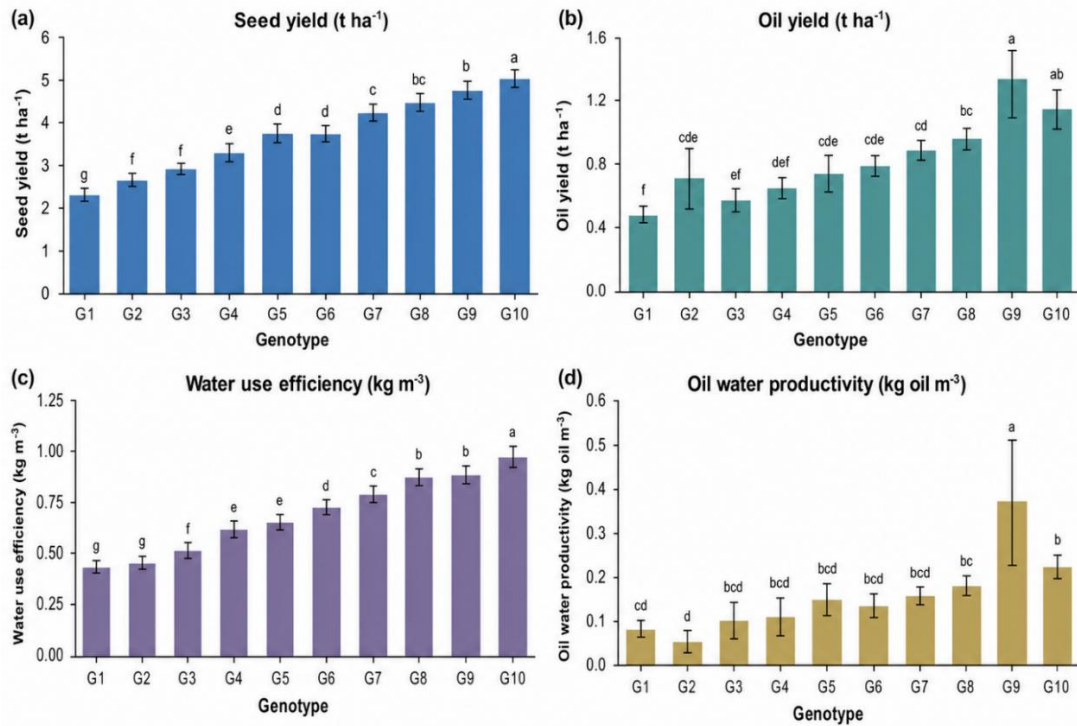


Figure 1. Combined productivity and water-productivity traits of soybean genotypes, (mean $\pm$ SE); letters = LSD at $P \leq 0.05$.

3.5 Agronomic and Yield-Component Traits

A related group of general agronomic and yield component (plant height, pods per plant, seeds per pod and their corresponding growth measure) traits averaged by group is presented in Figure 2. High seed-yielding genotypes (G8, G9, G10) exhibited a good overall performance for most of the yield component traits, thus the superior performance of these genotypes for high seed yield could be attributed to the combined effect of several yield components and not to a single predominant character.

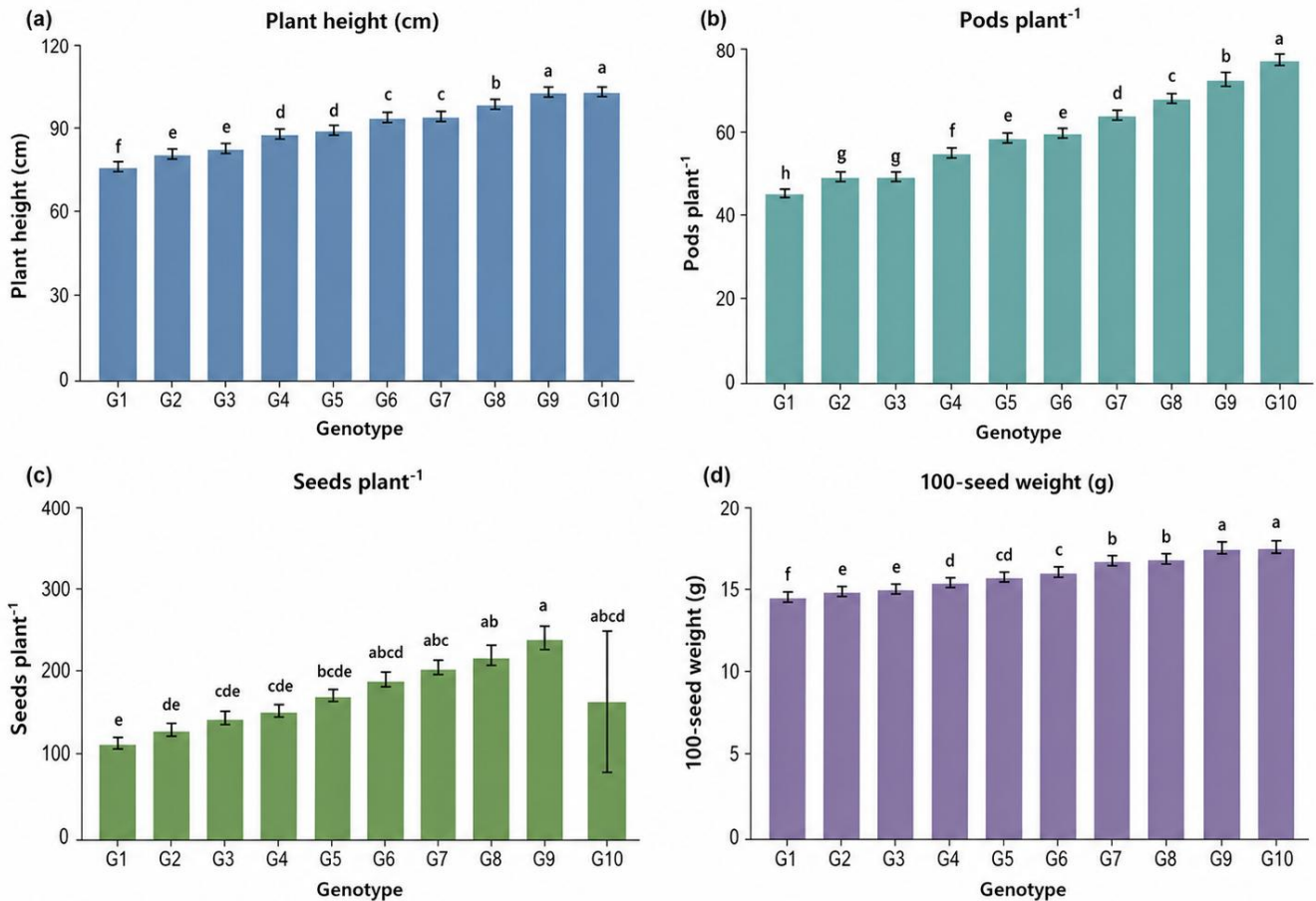


Figure 2. Combined agronomic growth and yield-component traits of soybean genotypes, (mean $\pm$ SE); letters = LSD at $P \leq 0.05$.

3.6 Seed-Quality and Oil-Composition Traits

Oil content and protein content varied from 19.333% (G1, G2) to 22.500% (G8) and 36.833% (G8) to 39.600% (G1), respectively (Table 3; Figure 3). Genotypes (G7, G8, G9, G10) which had maximum oil content and oil yield were also observed to have a minimum protein content as compared to genotypes with minimum oil content and oil yield (G1–G4). This trend aligns with the well-known negative correlation between oil and protein accumulation in the seed of soybean.

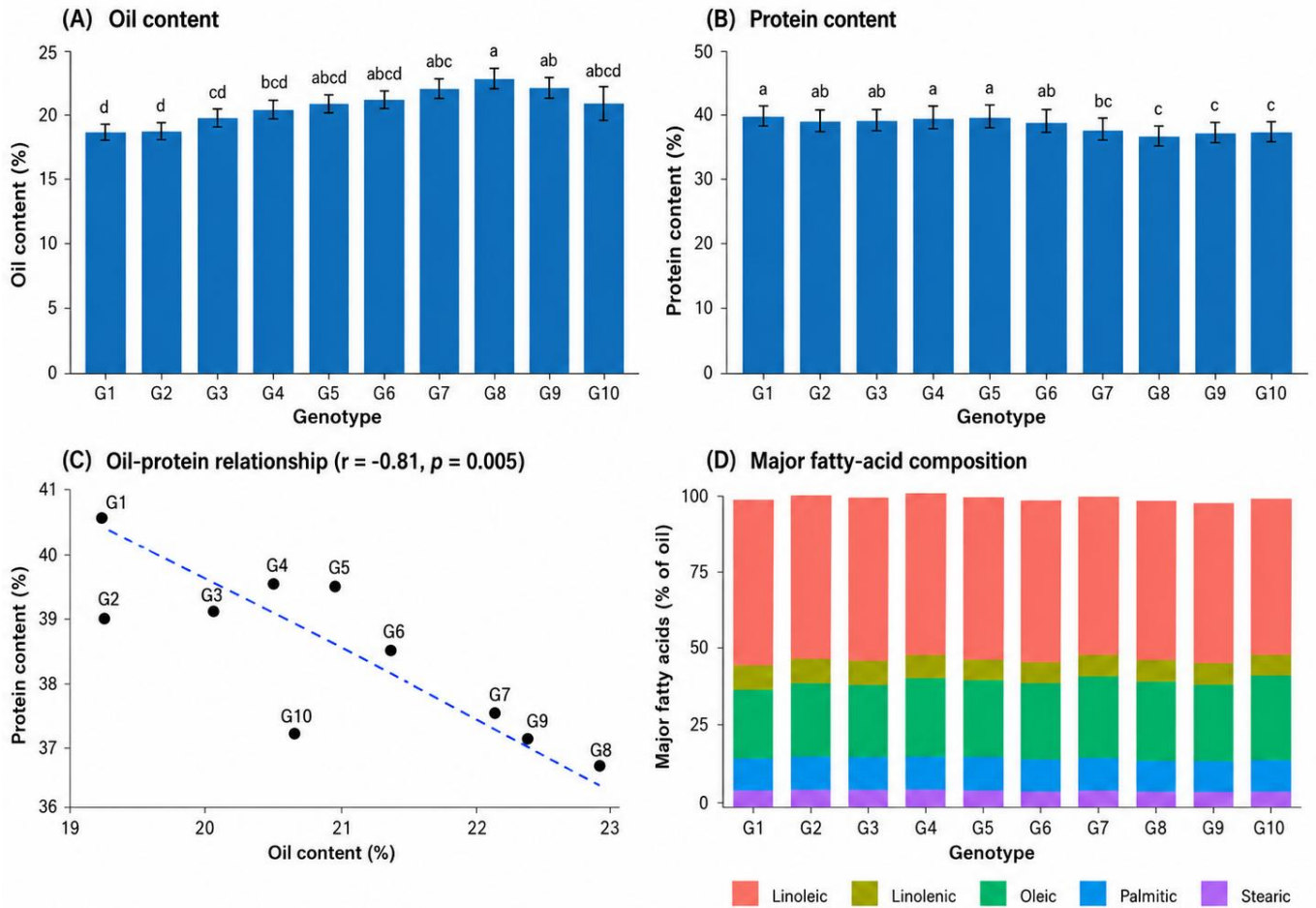


Figure 3. Combined seed-quality and oil-composition traits of soybean genotypes.

3.7 Correlation Analysis

The bivariate relationships of the key traits (Figure 4) revealed a strong positive correlation of seed yield with oil yield, WUE and oil water productivity follow the same pattern of parallel genotype ranking observed in Table 3 for these traits. Oil and protein content were negatively related in the genotypes, following the trend in Section 3.6 and in accordance with negative oil–protein associations described in soybean seed (Guo et al., 2022).

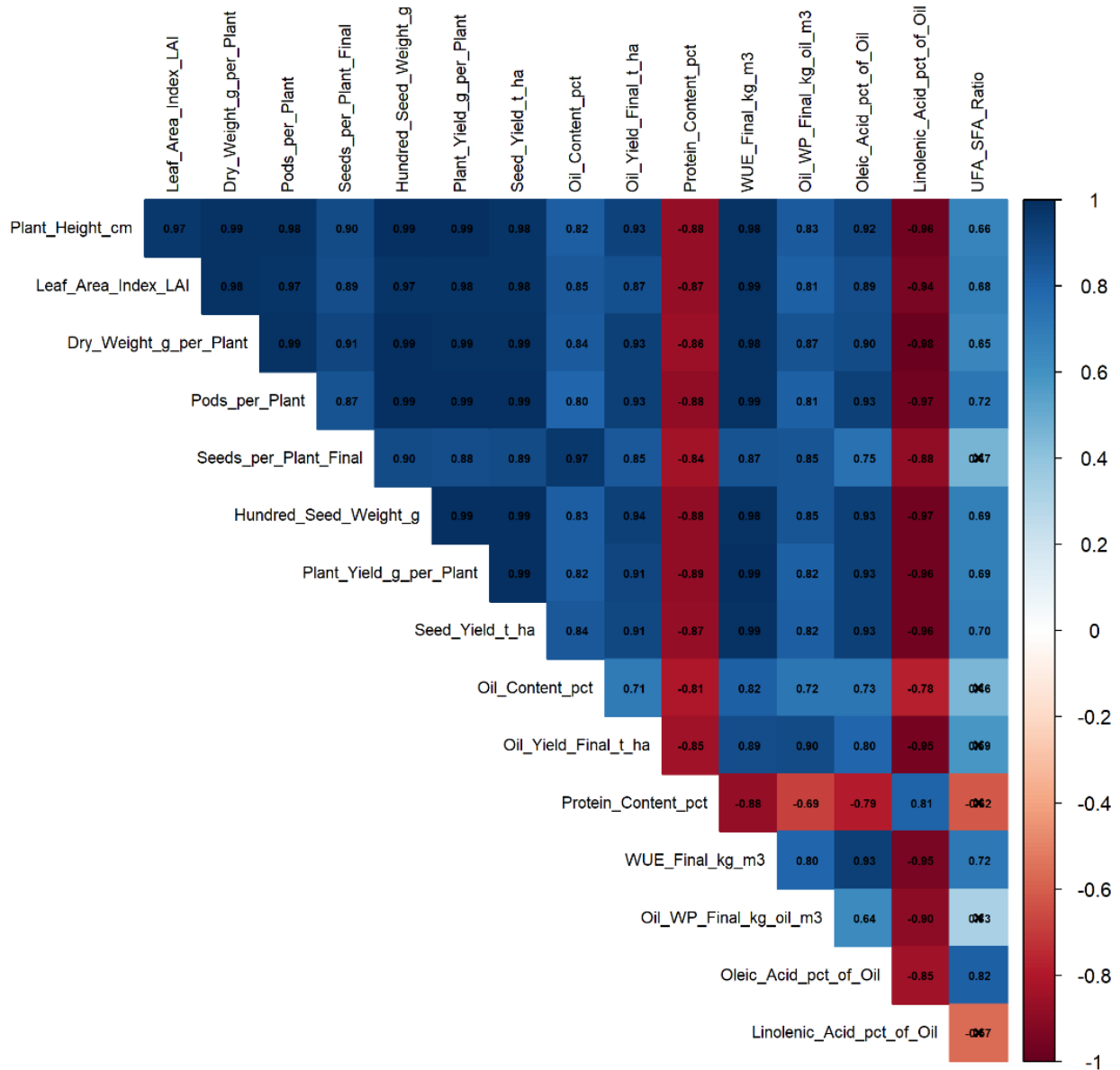


Figure 4. Correlation analysis

3.8 PCA and Cluster Analysis

Principal component analysis (PCA) and Ward hierarchical clustering of the pooled traits (Figure 5) separated the genotypes into clusters which were very similar to the productivity ordering defined above: a group of high-productivity genotypes G8, G9, G10; a group of moderate genotypes G5, G6, G7; and a group of low-productivity genotypes G1, G2, G3, G4. Such clustering pattern promoted the univariate ANOVA-LSD results the correlational pattern, suggested that genotype separation in this study is largely accounted for by a concerted productivity/water productivity axis more so than by independently varying traits.

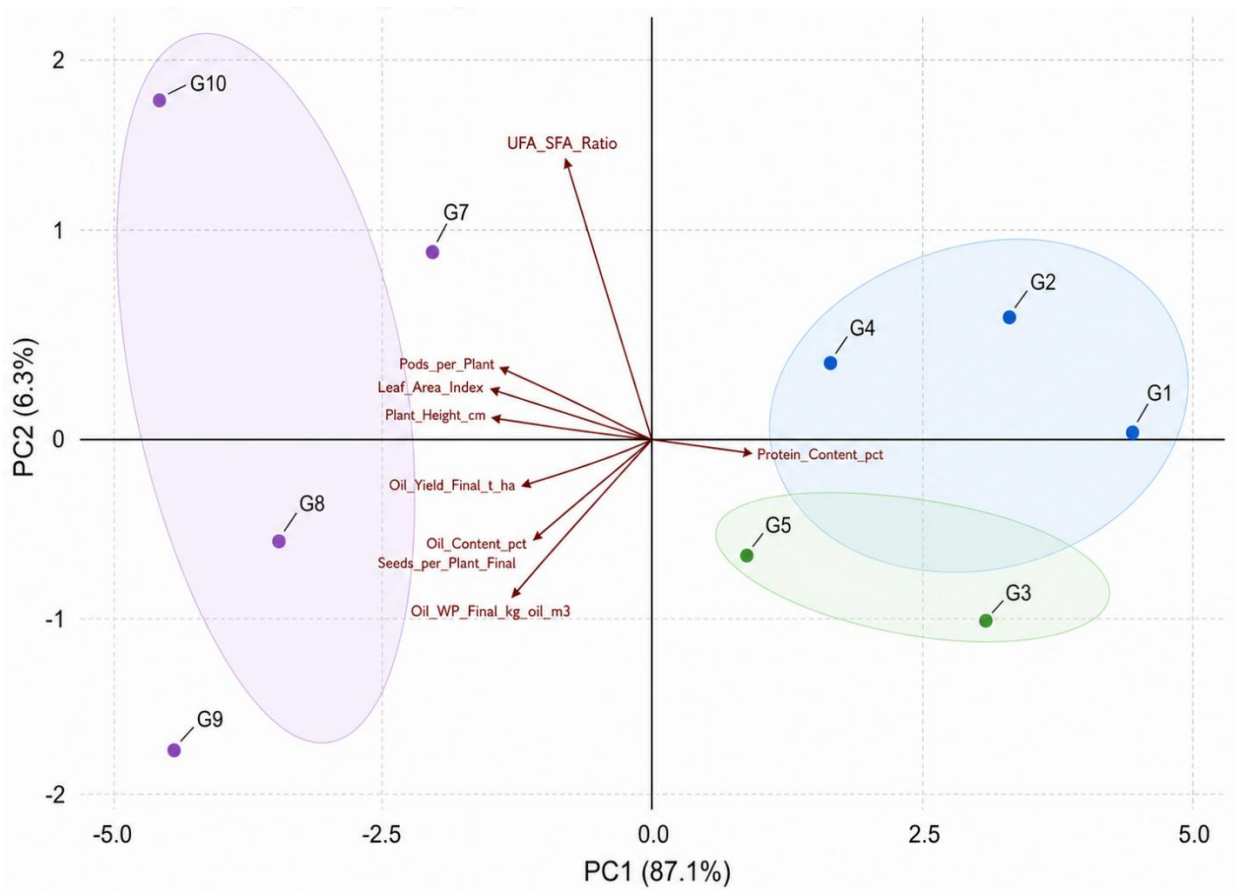


Figure 5. Biplot of PCA derived from the agronomic, yield, seed quality, and WUE traits of soybean genotypes.

3.9 Selection Index with Multiple Traits

Considering seed yield, oil yield and WUE, the weighted multi-trait selection index gave best rank to G10 (index = 93.3 out of 100; seed yield 4.96 t ha^{-1} , oil yield 1.033 t ha^{-1} , WUE 0.954 kg m^{-3}) followed by G9 (86.10) and G8 (80.03) (Table 4). The lowest genotype ranked was G1 (6.18). This ranking in the selection index agreed with the ANOVA-LSD results (Table 3), and with the results of the PCA/clustering (Section 3.8), where G10, G9 and G8 were the superior genotypes for seed yield, oil yield and water yield/pod under the semi-arid irrigated conditions of the present study.

Table 4. Multi-trait selection index ranking.

Rank	Genotype	Name	Index	Seed yield (t ha ⁻¹)	Oil yield (t ha ⁻¹)	WUE (kg m ⁻³)
1	G10	Imported drought-tolerant	93.27	4.96	1.033	0.954
2	G9	Local genotype (Abu Ghraib)	86.10	4.67	1.028	0.886
3	G8	QTLI	80.03	4.43	0.998	0.842
4	G7	Energy 2	74.04	4.23	0.926	0.796
5	G6	Soybean Parents	54.12	3.70	0.784	0.685
6	G5	Hawija 101	52.28	3.70	0.771	0.679

7	G4	TN12	35.22	3.18	0.647	0.580
8	G3	QT5	21.44	2.87	0.573	0.519
9	G2	Lee 74	15.10	2.67	0.516	0.477
10	G1	Hassan	6.18	2.37	0.458	0.424

4. Discussion

The pronounced and highly significant effects of genotype on seed yield, oil yield, WUE, and oil water productivity (Table 2) suggest that, in the constant-irrigation semi-arid conditions of this study, genotype was the largest source of variability in productivity and water productivity result in these tested plots. This is in line with earlier assessments of soybean germplasm that revealed significant genotypic differences for yield and WUE across irrigated and moisture stressed environments (Fried et al., 2019) and with reports that breeding has quantitatively modified water-use patterns and yield among soybean cultivars bred for dry and semi-dry areas of (Feng et al., 2022).

G10 was followed by G9 and G8, and they performed better than other genotypes in terms of seed yield, oil yield, WUE, oil water productivity, and the overall selection index. Since seed yield, oil yield, and WUE are related in a mathematical and biological way-oil yield is seed yield $\times$ oil content, and WUE and oil water productivity are similarly calculated based on the same constant water input data the fact that G10, G9, and G8 were superior in a coordinated manner for those traits indicated that it is truly higher seed and oil production per plant that underlies their superiority rather than a difference in water supplied, given that irrigation water was the same for all genotypes. This pattern agrees with reports that genotypes exhibiting a combination of high yield and high WUE can be directly identified in the field and selected for water-limited or high-value irrigated production systems (Soliman et al., 2025).

The atomized oil and protein contents were negatively related among genotypes (Section 3.6; Figure 4), consistent with the well-known inverse relationship between these two seed components in soybean, which is thought to be due to the competing nature of oil and protein biosynthesis for the same photosynthate and energy supply in seed filling (Guo et al., 2022). This trade-off applies for genotype selection: genotypes such as G8, G9 and G10 exhibiting the best combination of high oil yield with relatively low protein content and are more appropriate for production systems that focus on oil, while genotypes such as G1 combining a low yield with high protein content may continue to exist in where protein content is the single most important quality attribute.

The agreement of ANOVA-LSD grouping, correlation pattern, PCA/Ward clustering and weighted multi-trait selection index on the largest number of the same top ranked genotypes (G10, G9, G8) makes this ranking more robust than that of any other individual analysis. Multivariate analyses such as PCA-cum-hierarchical clustering are commonly applied in genotype evaluation studies largely because they discern whether genotypic differentiation is shaped by an integrated suite of traits or is influenced by multiple independent traits (Odong et al., 2013), the close grouping of G8, G9, and G10 and G1–G4 on the other hand denote a joint productivity/water-productivity axis of variation instead of trait-specific tricks of the mind.

The multi-trait selection index itself is based on the well-established principle of bundling up several traits into one trait that is weighted by economic or agronomic importance in genotype selection (Smith, 1936; Hazel, 1943), and the univariate and multivariate results concur well with that of this study and indicate that it can be used practically to rank genotypes for breeders who conduct their experiments in comparable irrigated semi-arid environments.

The caution for comparison extends to the general literature on soybean yield and WUE under water limited environments, much of which involves an explicit manipulation of levels of irrigation or drought stress (Fried et al., 2019; Feng et al., 2022; Soliman et al., 2025). Since irrigation water was non limiting in this experiment, the genotype rankings presented here represent performances under a single well-watered semi-arid moisture regime and do not allow prediction of genotype performance under water-limited or drought stressed conditions, and the extension of the word “drought-tolerant” beyond the genotype name itself (as was for G10 in the original germplasm designation) would be at odds with the present results. From a practical perspective, the equal performance of G10 and G9, and also closely followed by G8 in productivity traits, water-productivity traits under stable, sufficient irrigation implies potential candidates for more extensive multi-locational and multi-season testing, in semi-arid irrigated soybean production systems, especially where high seed and oil yield per unit of irrigation water application is an economic concern.

5. Conclusion

With irrigation water kept the same across genotypes, the identity of the genotype was the most significant factor that separated the seed yield, oil yield, water use efficiency, and oil water productivity of the ten genotypes grown in this randomized complete block design experiment in an irrigated, semi-arid condition with constant irrigation water. Genotype G10 (an imported drought-tolerant genotype, tested here only under full, constant irrigation) was the best candidate overall, attaining the highest seed yield (4.96 t ha^{-1}), oil yield (1.033 t ha^{-1}), WUE (0.954 kg m^{-3}), and multi-trait selection index value (93.3 out of 100), followed by G9 (a local genotype from Abu Ghraib) and G8 (QTLI). These results highlight the potential of genotype selection, rather than irrigation-regime manipulation, which was not evaluated here, as a point of intervention to enhance productivity, oil yield, and water productivity within the context of semi-arid irrigated soybean production.

References

- Araujo, G. R., da Silva, F. L., Adegas, F. S., Velini, E. D., & Alcántara-de la Cruz, R. (2026). Brazil as a global leader in soybean production: Historical trends, current status, and future perspectives. *International Journal of Plant Production*, 20(2), Article 56. <https://doi.org/10.1007/s42106-026-00434-x>
- Araujo, G. R., da Silva, F. L., Adegas, F. S., Velini, E. D., & la Cruz, R. A. D. (2026). Brazil as a Global Leader in Soybean Production: Historical Trends, Current Status, and Future Perspectives. *International Journal of Plant Production*, 20(2), 56. <https://doi.org/10.1007/s42106-026-00434-x>

- Feng, Y. Y., Richards, R. A., Jin, Y., et al. (2022). Yield and water-use related traits in landrace and new soybean cultivars in arid and semi-arid areas of China. *Field Crops Research*, 283, 108559. <https://doi.org/10.1016/j.fcr.2022.108559>
- Fried, H. G., Narayanan, S., & Fallen, B. (2019). Evaluation of soybean [*Glycine max* (L.) Merr.] genotypes for yield, water use efficiency, and root traits. *PLOS ONE*, 14(2), e0212700. <https://doi.org/10.1371/journal.pone.0212700>
- Gomez, K. A., & Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research* (2nd ed.). John Wiley & Sons.
- Guo, B., Sun, L., Jiang, S., Ren, H., Sun, R., Wei, Z., Hong, H., Luan, X., Wang, J., Wang, X., Xu, D., Li, W., Guo, C., & Qiu, L.-J. (2022). Soybean genetic resources contributing to sustainable protein production. *Theoretical and Applied Genetics*, 135(11), 4095–4121. <https://doi.org/10.1007/s00122-022-04222-9>
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, 28(6), 476–490. <https://doi.org/10.1093/genetics/28.6.476>
- Odong, T. L., van Heerwaarden, J., van Hintum, T. J. L., van Eeuwijk, F. A., & Jansen, J. (2013). Improving hierarchical clustering of genotypic data via principal component analysis. *Crop Science*, 53(4), 1546–1554. <https://doi.org/10.2135/cropsci2012.04.0215>
- Smith, H. F. (1936). A discriminant function for plant selection. *Annals of Eugenics*, 7(3), 240–250. <https://doi.org/10.1111/j.1469-1809.1936.tb02143.x>
- Soliman, A. A., Mousa, M. I., Ibrahim, M. A., Baiomy, K. A., Shaaban, S. A., Shabana, M. M. A., Mohamed, E. N. M., Rehan, M., Yu, H., & He, Y. (2025). Agronomic and anatomic performance of some soybean genotypes under optimal and water-deficit conditions. *Frontiers in Plant Science*, 16, 1575180. <https://doi.org/10.3389/fpls.2025.1575180>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Ruttanaprasert, R., Jogloy, S., Kaewkhieo-Ngam, T., Janket, A., Puttha, R., & Aninbon, C. (2025). Impact of Mid-Season drought on tuber yield, biomass, harvest index, and Water-Use efficiency of Jerusalem artichoke in tropical regions. *Agronomy*, 15(2), 395. <https://doi.org/10.3390/agronomy15020395>