



Multivariate characterization & Identification of superior genotypes for yield and yield components in sesame (*Sesamum indicum* L)

C Anjana¹, N Sabitha², D Bharathi³, B Ramana Murthy⁴

¹ Department of Genetics and Plant Breeding, S.V. Agricultural College, ANGRAU, Tirupati, Andhra Pradesh, India

² Associate Professor, Department of Genetics and Plant Breeding, S.V. Agricultural College, ANGRAU, Tirupati, Andhra Pradesh, India

³ Principal Scientist, Genetics and Plant Breeding, Regional Agricultural Research Station, Tirupati, Andhra Pradesh, India

⁴ Assistant Professor, Department of Statistics and Computer Applications, S.V. Agricultural College, ANGRAU, Tirupati, Andhra Pradesh, India.

Corresponding Author: N Sabitha

DOI: <https://doi.org/10.66856/ijaps.2026.8.4.8298>

Abstract

Sesame (*Sesamum indicum* L.) breeding requires simultaneous improvement of yield, phenology, plant architecture, seed attributes and quality traits, however, correlations among traits can complicate direct selection. The present study evaluated 45 sesame genotypes for 14 agronomic and quality-related traits using principal component analysis (PCA), factor analysis and the multi-trait genotype-ideotype distance index (MGIDI). Five principal components had eigenvalues greater than one and collectively accounted for 69.02% of the total variation. PC1, PC2, PC3, PC4 and PC5 explained 24.52, 18.45, 9.65, 8.70 and 7.71% of the variation, respectively. The retained factor solution grouped the traits into five interpretable dimensions. FA1 was primarily associated with number of capsules per plant, SCMR, seed yield per plant and harvest index, FA2 with plant height, first capsule height, days to maturity and specific leaf area, FA3 with primary and secondary branching, FA4 with oil content and days to 50% flowering, and FA5 with thousand-seed weight and capsule length. The MGIDI values ranged from 2.754 to 6.552, with G30, G33, G35, G43, G17, G8 and G39 occupying the first seven positions. At a 15% selection intensity, these seven entries would be retained according to the supplied MGIDI ranking. The largest reported selection differentials and expected gains occurred for number of capsules per plant and seed yield per plant, with selection differential percentages of 32.52 and 32.28%, respectively. The combined PCA-factor analysis-MGIDI framework reduced the dimensionality of the 14-trait dataset and provided an integrated criterion for multi-trait genotype prioritization. The conclusions are conditional on the supplied ideotype directions and derived statistical outputs, the original experimental design, location, season, replication and raw genotype × trait data should be verified before submission.

Keywords: *Sesamum indicum*, multivariate analysis, principal component analysis, factor analysis, MGIDI, ideotype, selection gain, sesame breeding

Introduction

Sesame (*Sesamum indicum* L.) is an important oilseed crop in which improvement of productivity is closely linked with plant architecture, phenology, seed characteristics and oil-related attributes. Because these traits are often correlated, selection for a single character may produce correlated changes in other traits and may not adequately represent the breeder's overall objective. Multivariate approaches are therefore valuable for describing the structure of variation and for identifying genotypes with balanced combinations of desirable attributes. Principal component analysis (PCA) transforms correlated variables into orthogonal components and summarizes the major dimensions of variation. Factor analysis provides a complementary interpretation by grouping correlated traits into latent dimensions, thereby facilitating biological interpretation and reducing redundancy among selection targets. The multi-trait genotype-ideotype distance index (MGIDI) integrates factor analysis with a genotype-ideotype distance framework. Traits are rescaled according to the desired direction of selection, correlated traits are represented through latent factors, and each genotype is evaluated according to its distance from the specified ideotype. Lower MGIDI values represent smaller multivariate distance from the target

combination of traits. The method was proposed to facilitate multi-trait selection while accounting for trait correlation and avoiding arbitrary weighting of individual traits (Olivoto and Nardino, 2021)^[1].

Materials and Methods

Experimental material and traits

The analysis comprised 45 sesame genotypes and 14 traits. The available analytical material contained derived PCA, factor-loading, selection-gain and MGIDI outputs rather than the original plot-level observations. Consequently, this manuscript reports the statistical outputs supplied for the study and does not infer unreported experimental details such as location, season/year, number of replications, plot size, randomization or genotype origin. These details should be inserted from the original field records before submission.

The trait abbreviations were retained as supplied: PH, plant height; NPB, number of primary branches; NSB, number of secondary branches; NCP, number of capsules per plant; CL, capsule length; SCMR, SPAD chlorophyll meter reading; TSW, thousand-seed weight; HI, harvest index; OC, oil content; SYP, seed yield per plant; DFF, days to

50% flowering; DM, days to maturity; FCH, first capsule height; and SLA, specific leaf area. The exact measurement units and operational definitions should be checked against the original field protocol.

Principal component analysis

PCA was summarized using eigenvalues, percentage variance explained and cumulative variance. Components with eigenvalues >1 were retained according to the Kaiser criterion. The first five components were retained and subsequently used as the dimensional basis for the factor interpretation. The analysis was intended to summarize the major dimensions represented by the 14 traits.

Factor analysis and trait grouping

Exploratory factor analysis was used to organize correlated traits into latent dimensions. Five factors (FA1–FA5) were supplied in the analytical output. Absolute loadings of approximately 0.50 or greater were treated as salient for interpretation. Because factor orientation is arbitrary, biological interpretation emphasized the magnitude and grouping of trait loadings rather than the algebraic sign alone.

MGIDI and ideotype definition

MGIDI was calculated using the multi-step framework described by Olivoto and Nardino (2021) ^[1]: (i) rescaling traits to a common 0–100 range while preserving the desired direction of selection; (ii) factor analysis to account for trait correlation; (iii) specification of the desired ideotype; and (iv) calculation of Euclidean distance between each genotype and the ideotype in factor space. For genotype *i*, the index can be represented as:

$$MGIDI_i = [\sum_{j=1}^f (\gamma_{ij} - \gamma_j)^2]^{1/2}$$

Where γ_{ij} is the score of genotype *i* on factor *j*, γ_j is the corresponding ideotype score, and *f* is the number of retained factors. Smaller values indicate smaller distance from the specified ideotype.

Selection differential and expected gain

The supplied output included selection differential (SD), selection differential percentage (SD%), heritability (h^2), selection gain (SG) and selection gain percentage (SG%) for each trait. These values were used directly to describe the expected response under the multi-trait selection strategy. No recalculation was performed because the raw genotype × trait matrix and experimental replication structure were not included in the supplied material.

Results

Principal component structure

Five principal components had eigenvalues greater than one and together accounted for 69.02% of the total variation among the 45 genotypes for the 14 traits (Table 1). PC1 explained the largest proportion of variation (24.52%), followed by PC2 (18.45%), PC3 (9.65%), PC4 (8.70%) and PC5 (7.71%). The cumulative variation increased to 42.96% after PC2, 61.31% after PC4 and 69.02% after PC5. The remaining nine components each had eigenvalues below one.

Table 1: Principal component variance explained by the retained components.

Component	Variance explained (%)	Cumulative variance (%)
PC1	24.52	24.52
PC2	18.45	42.96
PC3	9.65	52.61
PC4	8.70	61.31
PC5	7.71	69.02
PC6–PC14	30.98 (combined)	100.00

Factor loading pattern and communalities

The five-factor solution provided a biologically interpretable organization of the 14 traits. High communalities were observed for NPB (0.922), NSB (0.903), SYP (0.864), NCP (0.791), FCH (0.717), TSW (0.706), DFF (0.697), SCMR (0.696) and HI (0.683). In contrast, DM (0.381) and SLA (0.461) had comparatively lower communalities, indicating that a larger proportion of their variation was not represented by the retained common-factor structure.

Table 2: Factor loadings, communalities and uniquenesses for the 14 sesame traits.

Trait	FA1	FA2	FA3	FA4	FA5	Communality	Uniqueness
PH	-0.094	0.854	-0.048	-0.027	0.043	0.742	0.258
NPB	-0.173	-0.051	-0.913	0.010	0.237	0.922	0.078
NSB	-0.192	-0.027	-0.923	0.046	0.104	0.903	0.097
NCP	-0.746	0.096	-0.185	-0.108	0.424	0.791	0.209
CL	-0.092	0.103	-0.121	0.172	0.634	0.465	0.535
SCMR	-0.759	-0.073	-0.160	-0.040	-0.296	0.696	0.304
TSW	-0.063	-0.250	-0.158	-0.146	0.770	0.706	0.294
HI	-0.495	-0.473	-0.043	0.452	0.086	0.683	0.317
OC	0.078	-0.000	-0.054	0.790	0.023	0.634	0.366
SYP	-0.788	0.047	-0.110	-0.011	0.478	0.864	0.136
DFF	0.041	0.577	0.109	0.592	0.004	0.697	0.303
DM	0.094	-0.447	-0.326	-0.258	0.004	0.381	0.619
FCH	0.037	-0.828	0.038	-0.149	0.079	0.717	0.283
SLA	-0.174	-0.641	-0.087	0.109	0.027	0.461	0.539

FA1 was characterized mainly by NCP (−0.746), SCMR (−0.759), HI (−0.495) and SYP (−0.788), representing a

productivity/physiological dimension. FA2 was dominated by PH (0.854), FCH (0.577) and SLA (−0.641), with DM

also contributing (−0.447). FA3 was strongly associated with NPB (−0.913) and NSB (−0.923), representing branching architecture. FA4 was dominated by OC (0.790) and DFF (0.592), whereas FA5 was characterized by TSW (0.770) and CL (0.634).

Selection differential, heritability and expected gain

The largest positive selection differential percentages were reported for NCP (32.522%), SYP (32.282%), NPB (10.368%), NSB (9.955%) and HI (9.431%). Corresponding expected selection gains were 31.425% for NCP, 30.674% for SYP, 8.318% for NPB, 9.253% for NSB and 8.329% for HI. OC, DFF, CL and TSW also showed positive expected gains. The supplied output reported negative responses for DM, FCH and SLA, which were specified as decreasing traits. PH was listed as an increasing trait but had a negative selection differential and gain; this apparent inconsistency should be checked against the original ideotype specification.

Table 3: Selection differential, heritability and expected gain from the supplied MGIDI analysis.

Trait	Factor	SD (%)	h ²	SG (%)	Direction
NCP	FA1	32.522	0.966	31.425	increase
SCMR	FA1	3.297	0.810	2.671	increase
HI	FA1	9.431	0.883	8.329	increase
SYP	FA1	32.282	0.950	30.674	increase
PH	FA2	-0.996	0.859	-0.855	increase
DM	FA2	-0.840	0.959	-0.805	decrease
FCH	FA2	-3.728	0.938	-3.498	decrease
SLA	FA2	-1.507	0.821	-1.238	decrease
NPB	FA3	10.368	0.802	8.318	increase
NSB	FA3	9.955	0.929	9.253	increase
OC	FA4	2.266	0.958	2.171	increase
DFF	FA4	4.035	0.975	3.934	increase
CL	FA5	3.419	0.511	1.747	increase
TSW	FA5	4.167	0.935	3.895	increase

MGIDI ranking and selection

MGIDI values ranged from 2.754 to 6.552. G30 had the smallest MGIDI value (2.754), followed by G33 (2.949), G35 (3.331), G43 (3.369), G17 (3.373), G8 (3.501) and G39 (3.566). At a 15% selection intensity, approximately seven genotypes would be retained from 45 entries. Thus, the first seven entries in the supplied ranking constitute the numerical selection set under that convention.

Table 4: Top 15 sesame genotypes ranked according to MGIDI.

Rank	Genotype	MGIDI
1	G30	2.754
2	G33	2.949
3	G35	3.331
4	G43	3.369
5	G17	3.373
6	G8	3.501
7	G39	3.566
8	G16	3.583
9	G40	3.752
10	G13	3.958
11	G28	4.032
12	G42	4.061
13	G22	4.230
14	G29	4.310
15	G37	4.326

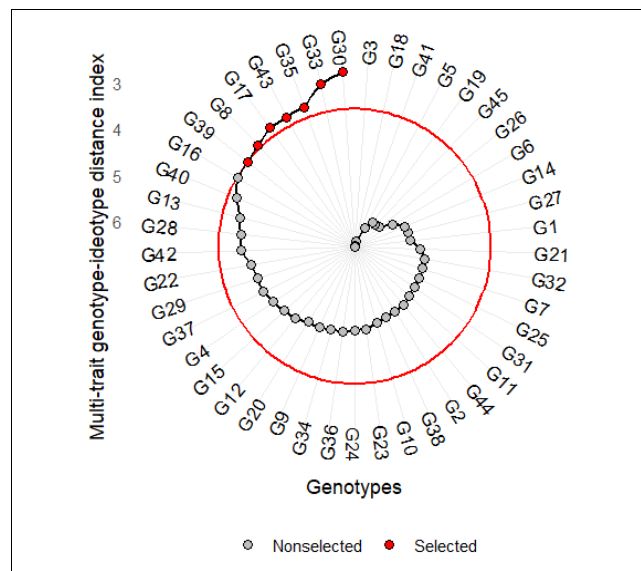


Fig 1: MGIDI-based genotype-ideotype distance profile of the 45 sesame genotypes.

Lower radial distance indicates a smaller distance from the specified ideotype; selected entries are highlighted in red in the supplied analytical figure.

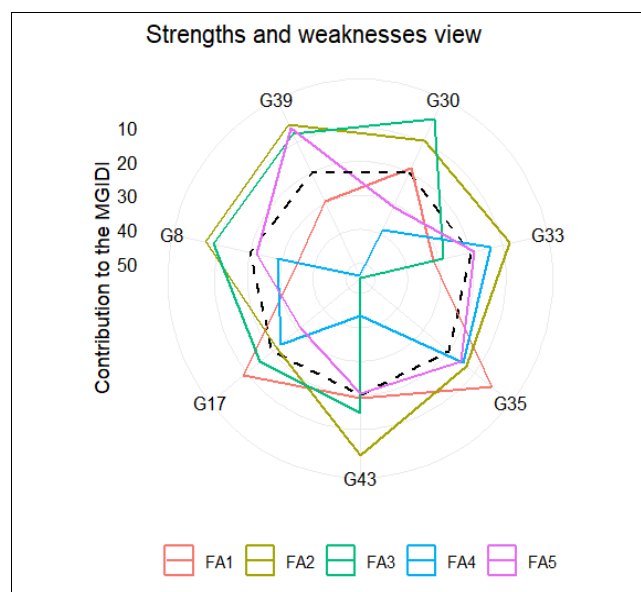


Fig 2: Strengths and weaknesses view for the selected sesame genotypes.

The contribution of FA1–FA5 to the MGIDI is shown for G39, G30, G33, G35, G43, G17 and G8. Smaller factor contributions indicate smaller distance from the ideotype for the traits represented by that factor.

Discussion

The PCA results demonstrated substantial multidimensional variation among the 45 sesame genotypes. The first five components captured 69.02% of the total standardized variation, indicating that a relatively small number of dimensions represented most of the information contained in the 14 traits. PC1 and PC2 together explained 42.96%, while the addition of PC3–PC5 increased the cumulative explanation to 69.02%. This pattern supports the use of multivariate dimension reduction for a dataset in which

yield, architecture, phenology, physiology and quality traits are considered simultaneously.

The factor-loading structure provided a more direct basis for biological interpretation. The concentration of NCP and SYP in FA1 is consistent with a productivity-related dimension, while SCMR and HI add physiological and efficiency-related information to the same latent axis. The strong loadings of NPB and NSB on FA3 indicate that branching architecture represented a distinct dimension. Similarly, OC and DFF were grouped within FA4, and TSW and CL within FA5. Such grouping can help breeders organize multi-trait selection around correlated trait complexes rather than treating all 14 variables as independent targets.

The MGIDI analysis translated the multivariate structure into a single distance measure. The low MGIDI values of G30, G33, G35, G43, G17, G8 and G39 indicate that these entries were closer to the specified ideotype in the retained factor space. The strength–weakness view further shows that the overall MGIDI of a selected genotype is composed of contributions from several factors. This is useful for breeding interpretation because two genotypes with similar overall MGIDI can differ in the factor responsible for most of their distance from the ideotype.

The reported selection gains highlight the potential for substantial response in productivity-associated traits. NCP and SYP had the largest selection differential percentages and expected gains, while branching traits and harvest index also showed positive responses. At the same time, negative responses for traits designated for reduction illustrate the trade-offs inherent in simultaneous selection. Such trade-offs are central to ideotype-based breeding, where the objective is not simply to maximize one trait but to move the population toward a specified combination of trait values.

Conclusion

The combined PCA, factor-analysis and MGIDI approach provided an integrated framework for evaluating 45 sesame genotypes across 14 traits. Five principal components with eigenvalues greater than one explained 69.02% of the total variation. The factor solution separated productivity/physiological traits, plant stature and phenology, branching architecture, oil content and flowering, and seed/capsule attributes into interpretable latent dimensions. Based on the supplied MGIDI output, G30, G33, G35, G43, G17, G8 and G39 formed the seven-entry selection set at 15% selection intensity. The largest expected responses were observed for number of capsules per plant and seed yield per plant. The results demonstrate the value of combining dimension reduction with ideotype-based multi-trait selection in sesame breeding.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Olivoto T, Nardino M. MGIDI: a novel multi-trait index for genotype selection in plant breeding. *Bioinformatics*,2021;37:1383–1389. <https://doi.org/10.1093/bioinformatics/btaa981>.
2. Olivoto T, Lúcio ADC. metan: an R package for multi-environment trial analysis. *Methods in Ecology and Evolution*,2020;11:783–789.

3. Olivoto T, Nardino M. MGIDI: a novel multi-trait index for genotype selection in plant breeding. *bioRxiv*, 2020, 2020.07.23.217778.
4. Olivoto T, et al. MGIDI: a powerful tool to analyze plant multivariate data. *Plant Methods*,2022;18:49. <https://doi.org/10.1186/s13007-022-00952-5>.
5. Jolliffe IT. *Principal Component Analysis*. 2nd ed. Springer, New York, 2002.
6. Jaya Sai Prasad G, Haritha T, Sirisha ABM, Subbaiah PV, Naveen YKV. Multi-trait Genotype-Ideotype Distance Index (MGIDI) approach for identifying superior genotypes in sesame (*Sesamum indicum* L.). *Electronic Journal of Plant Breeding*,2026;17(2):137–143. <https://doi.org/10.37992/2026.1702.017>.