OPEN ACCESS

Middle East Research Journal of Biological Sciences
ISSN 2789-7710 (Print & ISSN 2958-2091 (Online)
Frequency: Bi-Monthly
DOI: 10.36348/merjbs.2024.v04i03.002

KUWAIT SCHOLARS PUBLISHER
Website: http://www.kspublisher.com/
Email: office@kspublisher.com

Multivariate Analyses of Shattering and Seed Yield Related Morphological Traits Reveal High Yielding Sesame Genotypes Exhibit Low Degree of Shattering

Sintayehu Gedifew[1*]
[1]Ethiopian Institute of Agricultural Research, Assosa Agricultural Research Center, Assosa, Ethiopia

Abstract: Sesame production faces substantial challenges, particularly in terms of shattering. To address this issue, sesame breeding programs focus on developing cultivars with minimized shattering. A pivotal aspect in achieving high-yielding and shatter-resistant cultivars lies in comprehending the association between shattering and traits related to seed yield. Thus, this study aimed to examine the correlation between shattering and morphological traits associated with seed yield, as well as to characterize genotypes based on seed yield and shattering related traits. This study utilized 64 sesame genotypes, employing an 8 x 8 simple lattice design. The study revealed significant positive correlations between shattering and the duration from capsule opening to maturity, as well as the length of cracking on opened capsules. Notably, shattering exhibits a negative correlation with seed yield related morphological traits, such as plant height and branches, suggesting taller genotypes with more branches experience lower shattering. Similarly, shattering-related traits showed a significant negative correlation with yield related morphological traits. This study advocates selecting sesame genotypes with reduced shattering while maintaining high-yielding characteristics. Principal Component Analysis (PCA) of sesame genotypes reveals essential insights, with the first four components explaining 72.90% of the total variation. Seed yield and related traits contribute significantly to PC1, emphasizing their importance in explaining variability. Capsule length, shattering (%) and days from capsule opening to maturity have large scores on PC2. PCA confirmed genotypic differences, aiding breeders in selecting high-yielding, low-shattering varieties like AsARC-acc-SG-013 for future breeding programs. Cluster analysis grouped the 64 sesame genotypes into two clusters, where Cluster I and Cluster II represent 40.62% and 59.38% of the total genotypes, respectively. Cluster analysis identifies traits distinguishing Cluster I from II, including plant height, branches, capsules, capsule-bearing zone length, seed yield, and shattering-related traits. Genotypes belonging to Cluster I exhibit superiority for desirable traits.

Keywords: Capsule Opening, Correlation Coefficients, Days to First Capsule Opening, Seed Retention, Principal Component Analysis, Cluster Analysis.

RESEARCH PAPER

***Corresponding Author:**
Sintayehu Gedifew
Ethiopian Institute of Agricultural Research, Assosa Agricultural Research Center, Assosa, Ethiopia

How to cite this paper:
Sintayehu Gedifew (2024). Multivariate Analyses of Shattering and Seed Yield Related Morphological Traits Reveal High Yielding Sesame Genotypes Exhibit Low Degree of Shattering. *Middle East Res J Biological Sci*, *4*(3): 58-71.

Article History:
| Submit: 19.04.2024 |
| Accepted: 21.05.2024 |
| Published: 05.06.2024 |

Copyright © 2024 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

Sesame (*Sesamum indicum L.*, 2n=26), a valuable oilseed crop, faces numerous genetic and environmental challenges in its cultivation. Among these genetic hurdles, shattering poses a substantial adverse effect to the overall yield of sesame. Globally, two prevalent types of cultivars exist: shattering and non-shattering, experiencing seed losses of 50–90% and 30–50%, respectively (Qureshi *et al*., 2022). The splitting open of capsules at maturity leads to the release and subsequent falling of seeds to the ground, resulting in a significant reduction in overall yield. Shattering can cause a considerable yield loss of up to 50% in sesame (Ahmed *et al*., 2023), influenced by factors such as genetic control, environmental conditions, and agricultural management practices (Maity *et al*., 2021; Ahmed *et al*., 2023). The negative impact of seed yield loss due to shattering, both before and during harvesting, extends to the income of farmers and poses a threat to food security.

Mitigating yield loss in sesame involves various strategies, including developing cultivars with minimized shattering through targeted breeding programs, ensuring optimal harvest timing, and implementing improved agronomic practices. Reduction of seed yield loss can be achieved by opting for shatter-resistant genotypes (Ahmed *et al*., 2023) and adopting enhanced management practices, such as the application of paclobutrazol (Mehmood *et al*., 2021; Ahmed *et al*.,

2023). In pursuit of diverse breeding objectives for desirable characteristics, sesame breeders are actively working towards creating varieties with decreased shattering (Qureshi *et al*., 2022). Comprehending the specific morphological traits linked to seed shattering empowers breeders to precisely target these traits in their breeding programs. Through the selection or modification of these traits, there is the potential to develop crop varieties that exhibit reduced shattering. An effective breeding program aimed at mitigating seed yield loss due to shattering involves the careful selection or modification of genetic factors associated with shattering. This process requires a comprehensive understanding of the correlation between shattering and morphological traits related to seed yield. The ultimate goal is not to only develop cultivars that are characterized by low degree of shattering but also high yielding. Consequently, breeding programs should prioritize the development of cultivars possessing a combination of traits that positively influence both overall yield and the reduction of shattering.

Multivariate analysis methods are utilized to explore associations among traits and identify crucial traits for crop improvement (Visioni *et al*., 2013; Almeida *et al*., 2014). Correlation analysis has been applied to pinpoint key traits linked to frost tolerance in barley (Visioni *et al*., 2013) and reveal the relationships among yield-related traits in maize (Almeida *et al*., 2014). Morphological traits, such as the number of capsules per plant, the length of the capsule-bearing zone on the stem, and plant height, play a role in maximizing seed yield in sesame (Wang *et al*., 2023). Breeding efforts targeting these desired morphological traits may contribute to the development of high-yielding varieties with minimal seed shattering. Favorable associations between yield-related traits and shattering contribute to a higher overall sesame yield. Furthermore, Principal Component Analysis (PCA) is a statistical technique widely used across various scientific fields to reduce dataset dimensionality (Jolliffe, 2002). This involves transforming original variables into a new set of uncorrelated variables, known as principal components (Shlens, 2014). In plant breeding, PCA is extensively applied to assess genetic diversity (Singh *et al*., 2016) and select traits related to sesame yield (Gedifew, 2022). Furthermore, cluster analysis is employed to characterize and classify the germplasm of crops based on key traits, in order to select superior genotypes for breeding programs.

Therefore, the present study aimed to determine the extent of the relationship between seed shattering and seed yield related morphological traits. Additionally, the study aims to scrutinize the characteristics of genotypes concerning both seed yield and traits related to shattering.

2. MATERIALS AND METHODS

2.1 Plant Materials and Design

In this study, 64 sesame genotypes (Table 1) used as the plant materials. These materials were grown at Pawe Agricultural Research Center (PARC) in the 2019 cropping season, employing an 8 x 8 simple lattice design and adhering to recommended agronomic practices.

2.2 Data Collected

Data collected from non-destructive samples included the recording of Days to 90% maturity (DM) and days to the first capsule opening (DFCO) on a plot basis. To assess the uniformity in capsule ripening, the duration from the first capsule opening to 90% maturity (DM-DFCO) was computed by subtracting the days to the first capsule opening from the days to 90% maturity. Various parameters, such as plant height (PH), plant height to first branching (PHFB), length of the capsule-bearing zone (LCBZ), number of branches per plant (BPP), number of capsules on the main stem per plant (NCMS), total number of capsules per plant (CPP), and the number of opened capsules per plant (OCPP), were recorded on five randomly selected plants**.** The percentage of opened capsules per plant (POCPP) was calculated using the formula:

$$POCPP = \left(\frac{OCPP}{CPP}\right)100$$

Measurements of capsule length (CL) and capsule width (CW) were obtained from five randomly chosen unopened capsules. For opened capsules, the length of the opened capsule (LOC) and the length of cracking on the opened capsule (LCOC) were recorded on five instances. Subsequently, the percentage of cracking on opened capsules (PCOC) was determined using the formula:

$$PCOC = \left(\frac{LCOC}{LOC}\right)100$$

Seed yield per plant (SYPP) was recorded as the average seed yield of five random plants.

Table 1: List of the plant materials employed in the experiment

GenNo	Genotype	GenNo	Genotype	GenNo	Genotype	GenNo	Genotype
1	EBI17697	17	EBI28320	33	AsARC-acc-SA-017	49	KG-012 (2)
2	EBI17702	18	EBI202514	34	AsARC-acc-SA-019	50	MG-012 (1)
3	EBI17703	19	EBI207957	35	AsARC-acc-SA-020	51	MG-012 (2)
4	EBI17704	20	Abasena	36	AsARC-acc-SA-022	52	MT-023 (1)

GenNo	Genotype	GenNo	Genotype	GenNo	Genotype	GenNo	Genotype
5	EBI17708	21	AsARC-acc-S-001	37	AsARC-acc-SG-005	53	MT-075 (1)
6	EBI23548	22	AsARC-acc-S-003	38	AsARC-acc-SG-013	54	Setit-1
7	EBI23565	23	AsARC-acc-S-004	39	AsARC-acc-SG-018	55	Setit-2
8	EBI28301	24	AsARC-acc-S-006	40	GK-012 (1)	56	TM-023 (2)
9	EBI28302	25	AsARC-acc-S-010	41	GK-012 (2)	57	TZ-013 (1)
10	EBI28303	26	AsARC-acc-S-022	42	GM-012 (1)	58	TZ-013 (2)
11	EBI28304	27	AsARC-acc-SA-002	43	GM-012 (2)	59	TZ-054 (1)
12	EBI28306	28	AsARC-acc-SA-007	44	Gondar-1	60	TZ-054 (2)
13	EBI28308	29	AsARC-acc-SA-008	45	HM-012 (1)	61	ZT-013 (1)
14	EBI28309	30	AsARC-acc-SA-009	46	HM-012 (2)	62	ZT-013 (2)
15	EBI28316	31	AsARC-acc-SA-011	47	Humera-1	63	ZT-054 (1)
16	EBI28318	32	AsARC-acc-SA-016	48	KG-012 (1)	64	ZT-054 (2)

Information obtained from the destructive sample included recording the number of seeds dropped per opened capsule (SDPOC), which was determined by examining empty holes in five opened capsules. Additionally, the number of seeds dropped per opened capsule when the capsule is inverted downward (SDPOCI) and the number of seeds retained per opened capsule (SRPOC) were tallied for five opened capsules. Subsequently, shattering (Sh) in percent, per opened capsule was calculated using the formula:

$$Sh = \left(\frac{SDPOC + SDPOCI}{SDPOC + SDPOCI + SRPOC}\right)100$$

2.3 Statistical Analysis

The genotype means for the considered traits underwent Pearson correlation analysis, utilizing the *metan* R package (Oliveto and Lucio, 2020) within R software version 4.2.2 (R Core Team, 2022). To visually represent the relationships of traits, scatter plots were generated using the *ggpubr* R package (Kassambara, 2023). Additionally, a correlation heatmap was constructed to illustrate the associations between seed yield-related morphological traits, shattering, and shattering-related traits. This heatmap was created using the '*plot*' function for the '*corr_coef*' object.

Principal component analysis was conducted through the '*prcomp*' function in R software. The f*actoextra* package (Kassambara and Mundt, 2016) was employed to produce the scree plot of eigenvalues, biplot of individuals and variables, and assess the contribution of individuals and variables to each principal component. Kaiser's rule (Kaiser, 1960), suggesting the retention of principal components with eigenvalues exceeding 1, was utilized as a criterion for determining the optimal number of components.

For hierarchical clustering of genotypes, the squared Euclidean distance generated by the '*dist*' function in R software was employed. The complete linkage method clustered genotypes into genetically distinct groups using the '*hclust*' and plot functions. The ideal number of clusters was identified by evaluating the elbow point on the scree plot's curve, generated by the '*fviz_nbclust*' function within the *factoextra* package (Kassambara and Mundt, 2016). The means of clusters were computed to characterize genotypes within each cluster based on the considered traits.

3. RESULTS AND DISCUSSION

3.1 Correlation Analysis

Shattering represents the most detrimental trait causing substantial yield losses in sesame both before and during harvesting. Figure 1 illustrates the relationships of shattering and shattering-related traits. The findings revealed a significant positive correlation between shattering (Sh) and the number of days from the first capsule opening to maturity (DM_DFCO) (r = 0.30*) and the length of cracking on opened capsules (%) (PCOC) (r = 0.28*). Additionally, a non-significant positive correlation was observed between shattering and the percentage of opened capsules per plant (POCPP) (r= 0.15). Furthermore, the number of days from the first capsule opening to maturity displayed a non-significant positive correlation with the length of cracking on opened capsules (%) (r = 0.05) and the percentage of opened capsules per plant (r = 0.22), while the length of cracking on opened capsules (%) and the percentage of opened capsules per plant showed a non-significant positive correlation (r = 0.20). These results imply that a longer duration between the first capsule opening and maturity, along with an increased length of cracking on opened capsules, significantly contributes to shattering in sesame. A prolonged gap between the first capsule opening and maturity leads to seed drop from the bottom capsules while waiting for the capsules at the top of the plant to mature. To minimize shattering, harvesting is recommended when the lower capsules become dry and rupture, even if yield sacrifice is expected from the upper part of the plant (Qureshi *et al*., 2022). Although this study indicates a negligible positive impact of the percentage of opened capsules per plant on shattering, it does not guarantee the absence of significant yield loss due to shattering in sesame. This is because shattering in this experiment was measured by the number of seeds dropped per opened capsules, rather than the number of

seeds dropped per plant. In summary, developing sesame genotypes with a shorter duration between the first capsule opening and maturity, a reduced length of cracking on opened capsules, and fewer opened capsules per plant at maturity is crucial for minimizing seed yield loss caused by shattering.

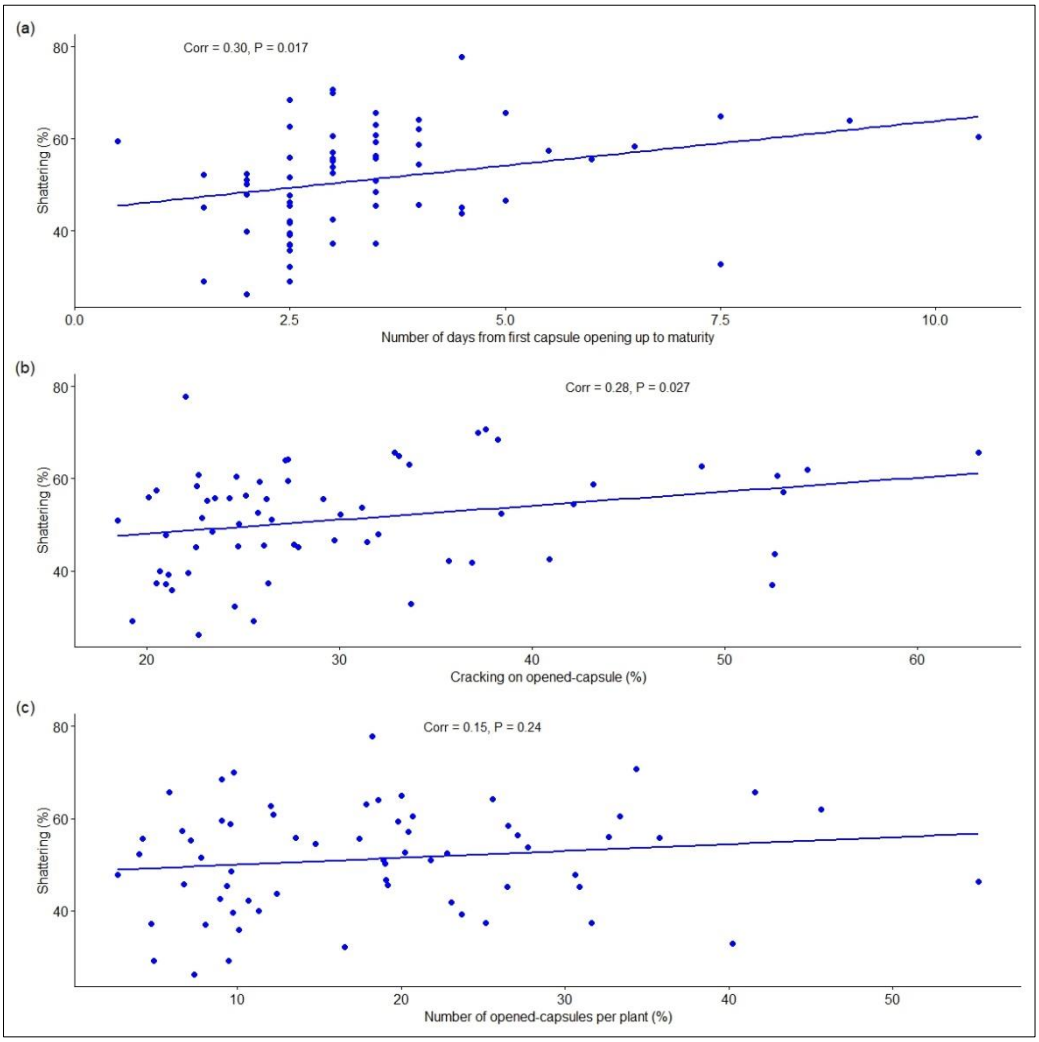

Figure 1: Scatter plots that depict the associations between shattering and its related traits

Figure 2 presents a scatter plot illustrating the relationships between seed yield and morphological traits. Seed yield (SYPP) exhibited a highly significant positive correlation with plant height (PH) (r = 0.60***), length of the capsule-bearing zone (LCBZ) (r = 0.79***), number of branches per plant (BPP) (r = 0.69***), number of capsules on the main stem per plant (NCMS) (r = 0.82***), and the total number of capsules per plant (CPP) (r = 0.91***). Conversely, morphological traits such as plant height to the first branch (PHFB) (r = 0.084), capsule length (CL) (r = 0.21), and capsule width (CW) (r = 0.081) showed a non-significant positive correlation with seed yield. Previous research by Khairnar and Monpara (2013), Abate *et al*., (2015), and Abhijatha *et al*., (2017) indicated a noteworthy correlation between the number of capsules per plant and sesame seed yield. Similarly, Teklu *et al*., (2017) documented a highly significant positive phenotypic correlation between seed yield and traits such as the length of the capsule-bearing zone, the number of capsules per plant, and the number of primary branches per plant. In Bulgarian sesame breeding efforts targeting mechanized harvesting and increased seed yield, Georgiev *et al*., (2008) identified the total number of capsules per plant and the number of capsules on the main stem per plant as the most crucial traits for optimizing seed yield. Therefore, enhancing seed yield in sesame can be achieved through indirect selection for morphological traits such as plant height, the length of the capsule-bearing zone, number of branches per plant, number of capsules on the main stem per plant, and the total number of capsules per plant. However, when pursuing indirect selection to maximize seed yield, it is essential to consider the association of these seed yield-related traits with undesirable traits such as shattering and its related traits.

Primarily, the goal of this investigation was to examine the connection between shattering and its associated traits and morphological traits related to seed yield (refer to Figure 3). Shattering (%) displayed a notable negative correlation with plant height (r = -0.30*) and the number of branches per plant (r = -0.31*). Conversely, there was a non-significant negative correlation observed between shattering (%) and plant height to first branching (r = -0.11), the length of the capsule-bearing zone (r = -0.12), the number of capsules on the main stem per plant (r = -0.06), and the number of capsules per plant (r = -0.15). Additionally, shattering (%) showed a non-significant positive correlation with capsule length (r = 0.16) and capsule width (r = 0.11). Traits associated with shattering, such as the length of cracking on opened capsules (%) and the number of opened capsules per plant (%), displayed a significant negative correlation with morphological traits that exhibited a notable positive correlation with seed yield per plant. Specifically, the number of opened capsules per plant (%) demonstrated a significant negative correlation with plant height (r = -0.36**), the length of the capsule-bearing zone (r = -0.50***), the number of branches per plant (r = -0.39**), the number of capsules on the main stem per plant (r = -0.44***), and the number of capsules per plant (r = -0.52***). The length of cracking on opened capsules (%) exhibited a significant negative correlation with plant height (r = -27*) and the number of branches per plant (r = -0.28*), while showing a non-significant negative correlation with plant height to first branching (r = -0.24), the length of the capsule-bearing zone (r = -0.09), the number of capsules on the main stem per plant (r = -0.23), and the number of capsules per plant (r = -0.20).

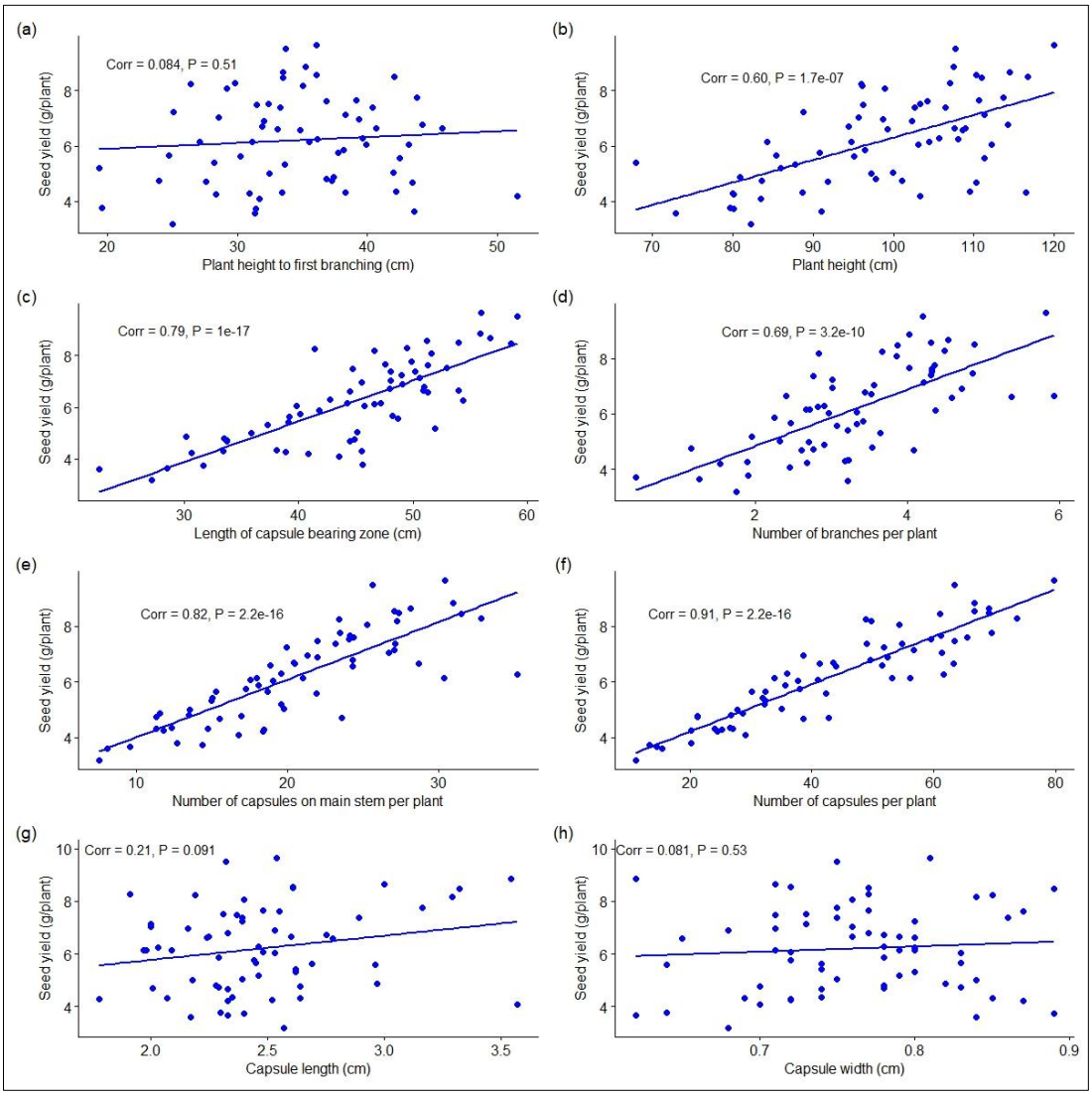

Figure 2: Scatter plots visualizing the relationships between seed yield and seed yield related morphological traits.

In this study, a confirmed significant negative correlation was identified between shattering and both plant height and the number of branches per plant. This implies that taller genotypes and those with a higher number of branches tend to exhibit lower shattering. Additionally, a non-significant negative correlation was found between shattering and yield-related morphological traits, such as the length of the capsule-bearing zone on the main stem, the number of capsules on the main stem per plant, and the number of capsules per plant. This suggests that genotypes with an extended capsule-bearing zone on the main stem and a high capsule-bearing capacity tend to have a lower shattering percentage. The study also revealed that genotypes with a lower proportion of opened capsules per plant were tall, had an extended capsule-bearing zone on the main stem, and displayed a high number of branches and capsules per plant. Furthermore, genotypes with a minimal length of cracking on opened capsules were tall and exhibited a high capsule-bearing and branching capacity. Overall, the consistently observed inverse correlation between yield-related morphological traits and shattering implies that genotypes with high seed-yielding characteristics tend to exhibit lower levels of shattering. Consequently, this study suggests that selecting for a relatively low degree of shattering within partially shattering populations of sesame genotypes does not adversely impact seed yield. In contrast, completely non-shattering sesame lines (indehiscent types) have lower yields and undesirable traits, making threshing more challenging (Qureshi *et al*., 2022). The authors argue that shattering types, especially from indehiscent types, surpass non-shattering types in terms of yield, quality, and suitability for manual harvesting. The gene governing shattering in indehiscent cultivars may be associated with detrimental effects that reduce the productivity of sesame.

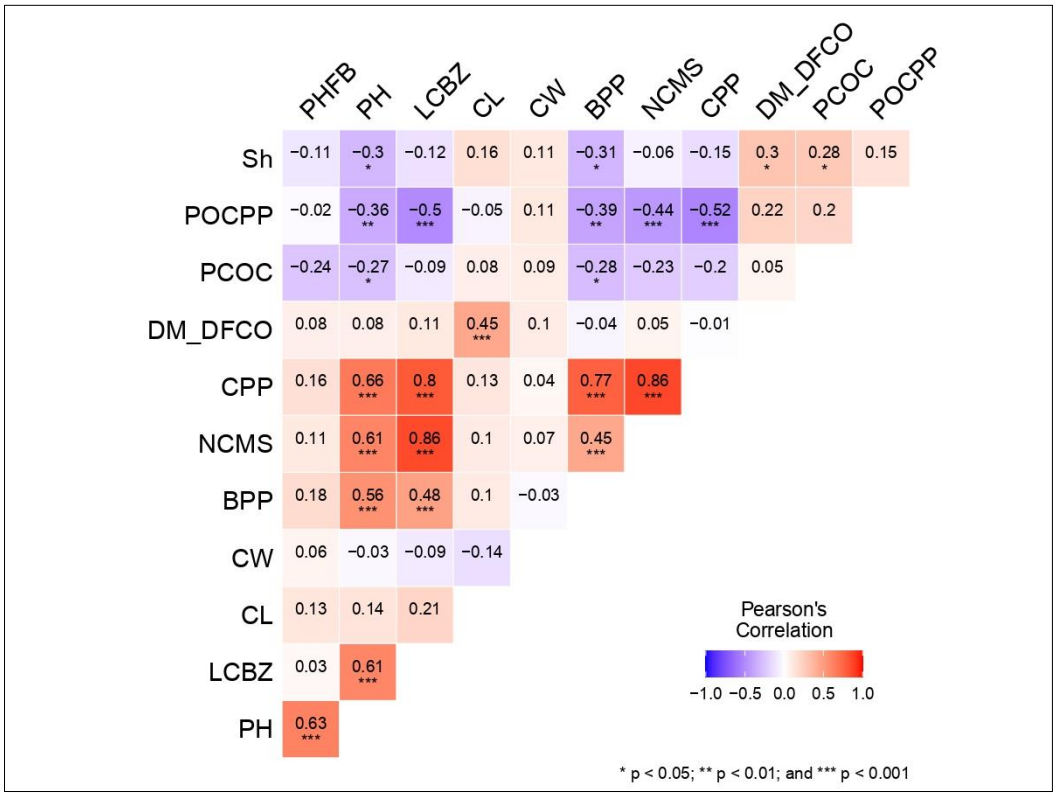

Figure 3: A correlation heatmap illustrating the relationships among seed yield-related morphological traits, shattering, and shattering-related traits

3.2 Principal Component Analysis

The findings from the study using Principal Component Analysis (PCA) on sesame genotypes reveal important insights into the variability and relationships among various traits. PCA successfully reduced the dimensionality of the dataset within the first four Principal Components (PCs) with an eigenvalue greater than 1 (Figure 4A). As suggested by Kaiser (1960), principal components with eigen values greater than 1 should be kept as potential contributor of the variation. The first principal component (PC1) accounted for 39% of the variability, with subsequent contributions from PC2 (13.70%), PC3 (11.30%), and PC4 (8.90%) as depicted in Figure 4B. The first four principal components explained approximately 72.90% of the total variation (Table 2) among the 64 sesame genotypes. This suggests that a substantial amount of information can be captured with a reduced set of variables.

Principal component analysis (PCA), as demonstrated by Arriel *et al*., (2007), allows for the simultaneous comprehension of discerning traits that account for explained genetic variability and the interrelationships among these traits. The present study identified the relative importance of traits in contributing to variability. Seed yield and its related traits, such as plant height, capsule-bearing zone length, number of branches, and number of capsules, exhibited significant contributions to the first principal component (PC1), accounting for 39% of the variation. Characters which had high scores to the PC implies that traits would be the highest contributor of the variation explained (Akbar *et al.*, 2011; Singh *et al.*, 2017). Seed yield per plant (-0.40) and seed yield related traits, such as plant height (-0.35), the length of capsule-bearing zone (-0.38), number of branches per plant (-0.34), number of capsules on the main stem per plant (-0.39), and number of capsules per plant (-0.42) showed high negative score loading to PC1 (Table 2). Similarly, the first principal component exhibited strong associations with traits such as plant height, number of capsules per plant (Ercan *et al*., 2002; Furat and Uzun, 2010; Gedifew, 2022; Mukhtambica *et al*., 2023), as well as the number of branches per plant and seed yield (Furat and Uzun, 2010; Gedifew, 2022; Mukhtambica *et al*., 2023). On the other hand, the number of days from first capsule opening up to maturity (0.57), capsule length (0.49), percent of cracking on opened capsule (0.31), and percent of shattering (0.49) had large score loading to PC2. PC3 and PC4 also showed distinct trait contributions to overall variability. The contribution of individual traits to specific principal components was identified. For example, the number of capsules per plant had the highest contribution to PC1 (Figure 5A), emphasizing its significance in explaining the variability among genotypes. Similarly, the number of days from capsule opening up to maturity was the highest contributor to PC2 (Figure 5B). The correlation analysis, in conjunction with principal component analysis (PCA), provides valuable insights into the relationships among seed yield-related morphological traits and shattering traits in sesame genotypes. In the present study, the correlation analysis indicated a negative correlation between seed yield-related morphological traits and shattering, a relationship that is further supported by the principal component analysis.

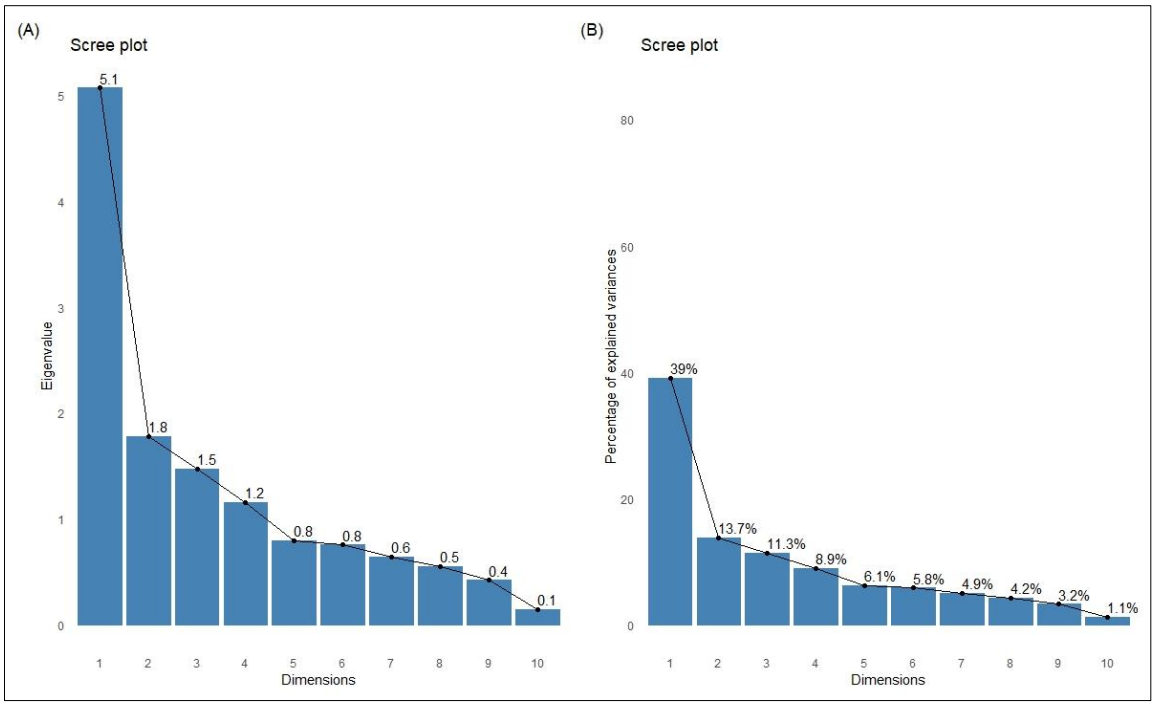

Figure 4: Eigenvalues (A) and percentage of explained variances (B) of principal components

Table 2: Eigenvalues and score load of original variables to the principal components

	Principal components			
	PC1	**PC2**	**PC3**	**PC4**
Eigenvalues	5.10	1.80	1.50	1.20
Variance Proportion (%)	39	13.70	11.30	8.90
Cumulative Variance (%)	39	52.70	64	72.90
Plant height	-0.35	-0.04	0.36	-0.07
Plant height to first branch	-0.12	-0.04	0.68	-0.18

Length of capsule bearing zone	-0.38	0.15	-0.18	0.09
Capsule length	-0.08	0.49	0.21	0.41
Capsule width	0.01	0.12	-0.04	-0.84
Number of branches per plant	-0.34	-0.11	0.05	0.01
Number of capsules on the main stem per plant	-0.39	0.11	-0.14	-0.12
Number of total capsules per plant	-0.42	0.04	-0.12	-0.08
Seed yield per plant	-0.40	0.10	-0.19	-0.07
Number of days from first capsule opening to maturity	0.00	0.57	0.27	0.03
Cracking on opened capsule (%)	0.13	0.31	-0.33	-0.08
Number of opened capsules per plant (%)	0.27	0.15	0.22	-0.19
Shattering (%)	0.12	0.49	-0.16	-0.12

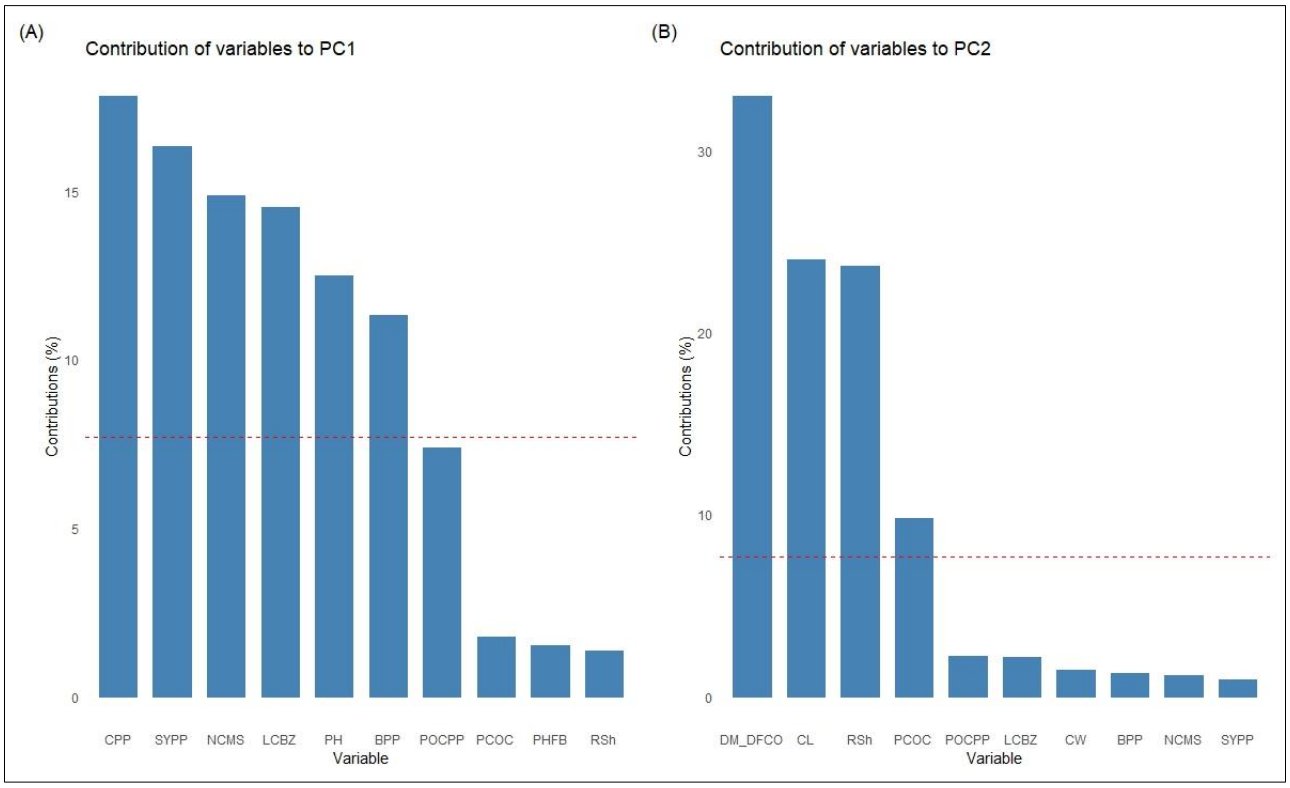

Figure 5: Graph anticipating the contributions of variables to PC1 (A) and PC2 (B)

The first principal component (PC1) demonstrates a negative loading scores for yield-related traits but a positive loading scores for shattering-related traits, reinforcing the inverse relationship between these two sets of characteristics. For instance, traits related to seed yield were negatively loaded on PC1, while shattering and its related traits such as percent of cracking on opened capsule and number of opened capsules per plant (%) were positively loaded on PC1. In general, the PCA analysis provided a comprehensive understanding of trait variability, trait associations, and the grouping of sesame genotypes. This information can be valuable for breeders and researchers in selecting genotypes with desirable traits and understanding the underlying genetic factors influencing sesame characteristics. Selection of individuals underlying the traits with high score loading to the first PCs is suggestable to exploit variability (Akbar *et al.*, 2011; Mukhthambica *et al*., 2023). The loading scores on PC1, particularly for seed yield and related traits, indicated a clear separation of genotypes into two distinct groups. This suggests that genotypes with similar characteristics cluster together, providing valuable information for grouping and classification. Genotypes, such as KG-012(1), TZ-054(2), ZT-054(2), KG012(2), Setit-2, and ZT-054(1), exhibited high positive loading scores on PC1 (Figure 6), indicating a high degree of shattering and low seed-yielding characteristics. In contrast, genotype AsARC-acc-SG-013, AsARC-acc-SA-008, and Gondar-1, which had high negative loading scores to PC1 were found to be high seed yielding genotypes with low degree of shattering.

The relative contribution of genotypes to PC1 and PC2 is highlighted in Figure 7. Genotype AsARC-acc-SG-013 had the largest contribution to PC1, followed by KG-012 (1) and TZ-054 (2) (Figure 7A). On the other hand, genotype EBI28306, EBI28304, and EBI28320 were the major contributor to PC2 (Figure 7B). This information is crucial for breeders and researchers in identifying genotypes that may require targeted improvement in terms of reducing shattering and enhancing seed yield. The study affirms the existence of variations among the studied genotypes for the considered traits. This variation offers opportunities for genetic improvement through targeted selection and breeding strategies. Genotypes with high seed yield and low shattering characteristics, such as AsARC-acc-SG-013, present promising avenues for future breeding programs. In general, the findings underscore the importance of understanding the relationships between seed yield-related traits and shattering in sesame genotypes. This knowledge can guide breeding efforts to enhance desirable traits, leading to improved seed and overall crop performance.

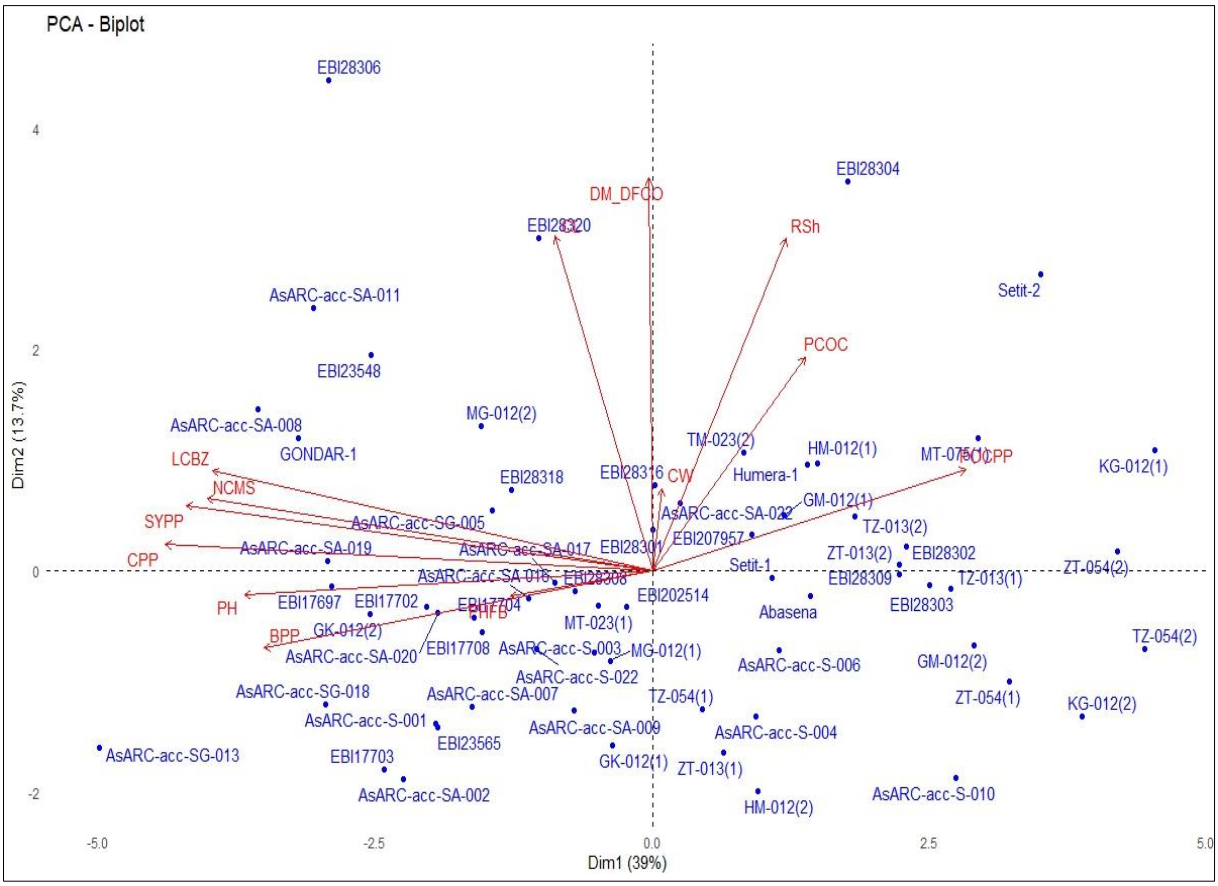

Figure 6: Principal component biplot of variables and sesame genotypes (PC1 vs PC2)

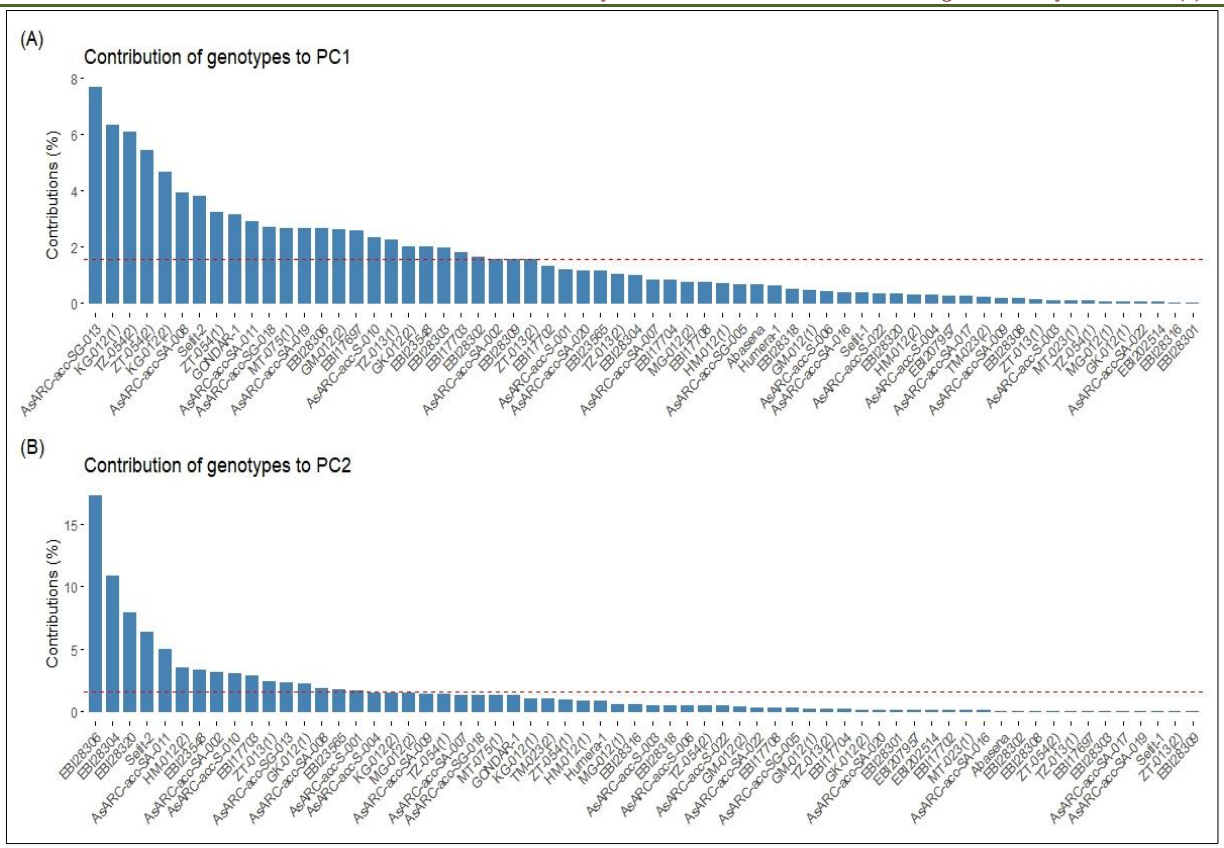

Figure 7: Graph anticipating the contributions of sesame genotypes to PC1 (A) and PC2 (B)

3.3 Cluster Analysis

The scree plot of within sum of squares against the number of clusters suggests that two clusters are appropriate for classifying the 64 sesame genotypes. This implies that there are distinct patterns or characteristics that allow for a meaningful division of the genotypes into two groups. The cluster grouping is given in Table 3 and the circular dendrogram is depicted on Figure 8. The distribution of genotypes among the clusters provides insights into the diversity within the sesame genotypes. Cluster I, with 40.62% of the genotypes, and Cluster II, with 59.38%, indicate that there is a relatively balanced representation in the two identified clusters (Table 3). Understanding the structure of the dendrogram can provide information about the similarities and dissimilarities between genotypes, helping researchers interpret the genetic or phenotypic relationships. The characteristics or traits that define these clusters would be of interest for further investigation. Understanding the traits that contribute to this clustering can be valuable for breeding programs. Investigating the specific characteristics that define each cluster and understanding the genetic or environmental factors contributing to the clustering can deepen the understanding of sesame genotypes. It opens avenues for more targeted and informed research directions. The possibility of enhancing productivity through selection and combining desirable genes can be evaluated using cluster analysis (Bandila *et al*., 2011).

Table 3: Pattern of clustering of 64 sesame genotypes based on 13 quantitative traits grown at Pawe, 2019

Cluster	List of genotypes	Number of genotypes
I	EBI17697, EBI17702, EBI17703, EBI17704, EBI17708, EBI23548, EBI23565, EBI28306, EBI28318, AsARC-acc-S-001, AsARC-acc-S-003, AsARC-acc-S-022, AsARC-acc-SA-002, AsARC-acc-SA-007, AsARC-acc-SA-008, AsARC-acc-SA-009, AsARC-acc-SA-011, AsARC-acc-SA-017, AsARC-acc-SA-019, AsARC-acc-SA-020, AsARC-acc-SG-013, AsARC-acc-SG-018, GK-012(1), GK-012(2), Gondar-1, MG-012(2)	26
II	EBI202514, EBI207957, EBI28301, EBI28302, EBI28303, EBI28304, EBI28308, EBI28309, EBI28316, EBI28320, Abasena, AsARC-acc-S-004, AsARC-acc-S-006, AsARC-acc-S-010, AsARC-acc-SA-016, AsARC-acc-SA-022, AsARC-acc-SG-005, GM-012(1), GM-012(2), HM-012(1), HM-012(2), Humera-1, KG-012(1),	38

	KG-012(2), MG-012(1), MT-023(1), MT-075(1), Setit-1, Setit-2, TM-023(2), TZ-013(1), TZ-013(2), TZ-054(1), TZ-054(2), ZT-013(1), ZT-013(2), ZT-054(1), ZT-054(2)	

The research identifies distinct traits that differentiate Cluster I from Cluster II. Cluster means for the 13 quantitative traits are presented in Table 4. Plant height, number of branches, number of capsules, the length of capsule bearing zone, seed yield and shattering related traits discriminated the two clusters. Similarly, number of branches, number of capsules, and seed yield were the most discriminating traits of sesame genotypes grouped in different clusters (Abate *et al*., 2015; Gedifew, 2022). Genotypes belonging to Cluster I exhibit taller stature, more branches, higher capsule production, and greater seed yield compared to Cluster II. This differentiation in quantitative traits indicates a significant variation in the genetic characteristics of the sesame genotypes between the two clusters. Cluster I, being characterized by taller plants with more branches and capsules, along with high seed yield, suggests that these genotypes possess desirable traits. The lower percentages of opened capsules, cracking on opened capsules, and shattering in Cluster I further highlight the potential for better seed retention characteristics, making these genotypes more attractive for cultivation.

Cluster II, in contrast, exhibits inferior traits related to seed yield. Dwarf stature, fewer branches, lower capsule production, and reduced seed yield characterize Cluster II. The higher percentages of shattering, opened capsules, and cracking on opened capsules in this cluster suggest challenges in harvest efficiency and seed quality. The cluster analysis provides valuable insights for breeding programs. Traits associated with higher seed yield and desirable traits in Cluster I can be prioritized for further breeding efforts. This information is crucial for developing improved sesame varieties that are more productive. Farmers can benefit from the information regarding the characteristics of each cluster. For example, if seed yield is a primary concern, selecting genotypes from Cluster I may be advantageous. On the other hand, farmers may need to consider specific management practices for genotypes in Cluster II to overcome challenges related to lower seed yield and potential issues during harvesting.

Table 4: Cluster means for seed yield-related morphological traits, shattering, and shattering-related traits.

Trait	Cluster	
	I	II
Plant height (cm)	106.22	93.54
Plant height to first branching (cm)	35.34	34.18
Length of capsule bearing zone (cm)	50.67	40.63
Capsule length (cm)	2.50	2.45
Capsule width (cm)	0.77	0.76
Branches per plant	4.13	2.79
Number of capsules on the main stem	26.08	16.78
Capsule per plant	60.23	31.60
Seed yield per plant (g)	7.68	5.19
Days from first capsule opening to days to 90% maturity	3.19	3.62
Percentage of cracking on opened capsule (%)	27.86	32.10
Percentage of opened capsules per plant (%)	11.01	23.50
Shattering (%)	47.99	53.43

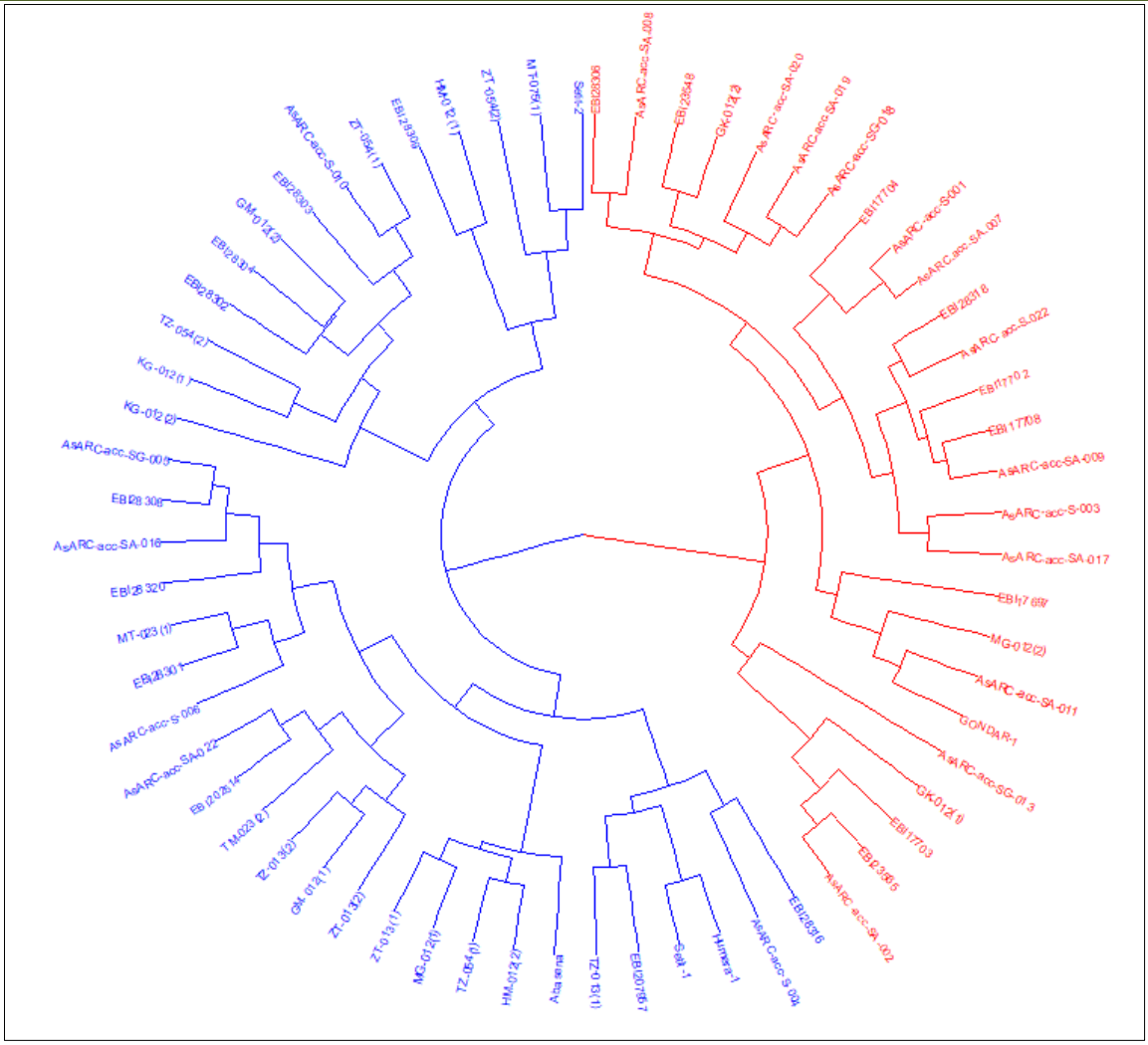

Figure 8: Agglomerative hierarchical clustering of 64 sesame genotypes based on 13 quantitative traits of sesame: Red-colored (Cluster I) & blue-colored (Cluster II)

CONCLUSION

Shattering is a significant factor causing considerable sesame yield losses. Correlation analysis revealed that genotypes with reduced shattering are characterized by tall plant stature, numerous branches, and capsules per plant, reaffirming a negative correlation between shattering and high seed-yielding traits.

Principal Component Analysis unveils key insights into trait variability and relationships. The first principal component, represented by seed yield and related traits, negatively correlated with shattering-related traits. This indicates the presence of distinct individuals. For instance, KG-012(1) exhibits high shattering and low seed yield, contrasting with AsARC-acc-SG-013.

Cluster analysis confirmed that genotypes with increased height, branching, capsule bearing, and higher seed yield show better seed retention.

Declaration of Competing Interest: The author has declared no competing interests.

Data Availability Statement: The data will be made available up on request.

Supplementary file: R codes used in correlation, principal component and cluster analyses

Load a data set from an opened R directory

data<-**read.csv**("Data_shattering.csv") # a data set has ‘ Genotype’ as a grouping variable and 13 variables (PH, PHFB,LCBZ,CL,CW,BPP,NCMS,CPP,SYPP,DM_DF CO,PCOC,POCP, Sh)

1. *Correlation analysis*

Correlation coefficients with significance test (Figure 3)

library(metan)

coeff_vars<-**corr_coef**(data[,2**:**14])

```
plot(coeff_vars, digits.cor =2,type = "upper",reorder = FALSE,size.text.lab = 14,
size.text.plot = 5)
## Scatter plot to visualize the relationship b/n shattering and its related traits (Figure 1)
library(ggplot2)
library(ggpubr)
plota<-ggscatter(data[,2:14],x="DM_DFCO", y="Sh",add="reg.line",color="blue")+
annotate("text",x=c(2),y=c(80),label=c("Corr = 0.30, P = 0.017"))+
labs(x="Number of days from first capsule opening up to maturity",y="Shattering (%)")
plotb<-ggscatter(data[,2:14],x="PCOC",y="Sh",add="reg.line",color="blue")+
annotate("text",x=c(45),y=c(80),label=c("Corr = 0.28, P = 0.027"))+
labs(x="Cracking on opened-capsule (%)",y="Shattering (%)")
plotc<-ggscatter(data[,2:14],x="POCPP", y="Sh",add="reg.line",color="blue")+
annotate("text",x=c(25),y=c(80),label=c("Corr = 0.15, P = 0.24"))+
labs(x="Number of opened-capsules per plant (%)",y="Shattering (%)")
library(metan)
arrange_ggplot(plota,plotb,plotc,ncol = 1,tag_levels = list(c("(a)", "(b)","(c)")))
## Scatter plot to visualize the relationship b/n seed yield and morphological traits (Figure 2)
plotA<-ggscatter(data[,2:14],x="PHFB", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(25),y=c(9),label=c("Corr = 0.084, P = 0.51"))+
labs(x="Plant height to first branching (cm)",y="Seed yield (g/plant)")
plotB<-ggscatter(data[,2:14],x="PH", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(85),y=c(8.5),label=c("Corr = 0.60, P = 1.7e-07"))+
labs(x="Plant height (cm)",y="Seed yield (g/plant)")
plotC<-ggscatter(data[,2:14],x="LCBZ", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(30),y=c(9),label=c("Corr = 0.79, P = 1e-17"))+
labs(x="Length of capsule bearing zone (cm)",y="Seed yield (g/plant)")
plotD<-ggscatter(data[,2:14],x="BPP", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(2.5),y=c(9),label=c("Corr = 0.69, P = 3.2e-10"))+
labs(x="Number of branches per plant",y="Seed yield (g/plant)")
plotE<-ggscatter(data[,2:14],x="NCMS", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(15),y=c(8.5),label=c("Corr = 0.82, P = 2.2e-16"))+
labs(x="Number of capsules on main stem per plant",y="Seed yield (g/plant)")
plotF<-ggscatter(data[,2:14],x="CPP", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(25),y=c(8.5), label=c("Corr = 0.91, P = 2.2e-16"))+
labs(x="Number of capsules per plant",y="Seed yield (g/plant)")
plotG<-ggscatter(data[,2:14],x="CL",  y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(2),y=c(10),label=c("Corr = 0.21, P = 0.091"))+
labs(x="Capsule length (cm)",y="Seed yield (g/plant)")
plotH<-ggscatter(data[,2:14],x="CW", y="SYPP",add="reg.line",color="blue")+
annotate("text",x=c(0.65),y=c(10), label=c("Corr = 0.081, P = 0.53"))+
labs(x="Capsule width (cm)",y="Seed yield (g/plant)")
library(metan)
arrange_ggplot(plotA,plotB,plotC,plotD,plotE,plotF, plotG, plotH,ncol = 2,tag_levels = list(c("(a)", "(b)","(c)","(d)","(e)","(f)","(g)","(h)")))
```

2. Principal component analysis

```
rownames(data)<-c(data$Genotype)
library(factoextra)
library(ggplot2)
pca <- prcomp(data[,2:14], center= TRUE, scale = TRUE)
## Scree plot showing eigenvalues and explained variances (Figure 4)
A<-fviz_eig(pca, choice = "eigenvalue", addlabels=TRUE)
A1<-A+theme_minimal()+theme(panel.grid.major=element_blank(),panel.grid.minor=element_blank())
B<-fviz_eig(pca, addlabels = TRUE, ylim = c(0, 85))
B1<-B+theme_minimal()+theme(panel.grid.major=element_blank(),panel.grid.minor=element_blank())
library(metan)
arrange_ggplot(A1,B1,ncol=2,tag_levels = list(c("(A)","(B)",size=1)))
## Score loads of original variables to PCs (Table 2)
PC.scores.vars<-pca$"rotation"
PC.scores.vars[,1:4]
## Contributions of variables to PC1 & PC2 (Figure 5)
C<-fviz_contrib(pca, choice="var", axes = 1, top = 10 )+
theme_minimal()+theme(panel.grid.major=element_blank(),
 panel.grid.minor=element_blank())
C1<-C+labs(title="Contribution of variables to PC1",x="Variable")
D<-fviz_contrib(pca, choice="var", axes = 2, top = 10)+
theme_minimal()+theme(panel.grid.major=element_blank(),
panel.grid.minor=element_blank())
```

```
D1<-D+labs(title="Contribution of variables to PC2",x
="Variable")
library(metan)
arrange_ggplot(C1,D1,ncol=2,tag_levels = list(c("(A)"
,"(B)",size=1)))
## Biplot of variables and individuals (Figure 6)
fviz_pca_biplot(pca,col.var = "red", col.ind = "blue", re
pel=TRUE)+
theme_minimal()+theme(panel.grid.major=element_b
lank(),
panel.grid.minor=element_blank())
## Contributions of individuals to PC1 & PC2 (Figure
7)
E<-fviz_contrib(pca, choice="ind", axes = 1)+
theme(panel.grid.major=element_blank(),panel.grid.m
inor=element_blank())+ labs(title="Contribution of ge
notypes to PC1",x="Genotype")
G<-fviz_contrib(pca, choice="ind", axes = 2)+theme(p
anel.grid.major=element_blank(),
panel.grid.minor=element_blank())+
labs(title="Contribution of genotypes to PC2",x="Genot
ype")
library(metan)
arrange_ggplot(E,G,ncol=1,tag_levels = list(c("(A)","(
B)",size=1)))
```

3. Cluster analysis

```
library(NbClust)
library(factoextra)
library(dplyr)
## Optimal number of clusters
nb <- NbClust(data[,2:14], method = "kmeans")
## Hierarchical clustering with complete linkage
distance<-dist(data[,2:14])
hc.c<-hclust(distance,method="complete")
## Cluster members
cluster<-cutree(hc.c,2)
cluster
cluster<-as.data.frame(cluster)
cluster
attach(data)
cluster.members<-cbind(cluster,Genotype)
groups<-arrange(cluster.members, cluster)
groups
## Dendrogram
fviz_dend(hc.c,cex=0.44,lwd=0.65,k=2, rect=TRUE,k_
colors=c("red","blue"),
rect_border="jco",rect_fill=FALSE,horiz=TRUE, type=
"circular")+
theme_void()
## Cluster means
c.mean<-aggregate(data[,2:14],list(cluster=cluster$clus
ter),mean)
c.mean
```

REFERENCES

- Abate, M., Mekbib, F., Ayana, A., & Nigussie, M. (2015). Genetic variability and association of traits in mid-altitude sesame (Sesamum indicum L.) germplasm of Ethiopia. *American Journal of Experimental Agriculture*, *9*(3), 1-14.
- Abhijatha, A., Arya, K., Madhukar, K., Gogineni, S. (2017). Evaluation of sesame (Sesamum indicum L.) genotypes to the shaded uplands of southern region. *International Journal of Current Microbiology and Applied Science*, 6, 332-339.
- Ahmed, J., Qadir, G., Ansar, M., Wattoo, F. M., Javed, T., Ali, B., ... & Rahimi, M. (2023). Shattering and yield expression of sesame (Sesamum indicum L) genotypes influenced by paclobutrazol concentration under rainfed conditions of Pothwar. *BMC Plant Biology*, *23*(1), 1-13. DOI: https://doi.org/10.1186/s12870-023-04145-7.
- Akbar, F., Rabbani, M. A., Shinwari, Z. K., & Khan, S. J. (2011). Genetic divergence in sesame (Sesamum indicum L.) landraces based on qualitative and quantitative traits. *Pak. J. Bot*, *43*(6), 2737-2744.
- Almeida, G. D., Nair, S., Borém, A., Cairns, J., Trachsel, S., Ribaut, J. M., ... & Babu, R. (2014). Molecular mapping across three populations reveals a QTL hotspot region on chromosome 3 for secondary traits associated with drought tolerance in tropical maize. *Molecular breeding*, *34*, 701-715.
- Arriel, N. H. C., Di Mauro, A. O., Arriel, E. F., Unêda-Trevisoli, S. H., Costa, M. M., Bárbaro, I. M., & Muniz, F. R. S. (2007). Genetic divergence in sesame based on morphological and agronomic traits. *Crop Breeding and Applied Biotechnology*, 253-261.
- Bandila, S., Ghanta, A., Natarajan, S., & SUBRAMONIAM, S. (2011). Determination of genetic variation in Indian sesame (Sesamum indicum) genotypes for agro-morphological traits. *Journal of Research in Agricultural Science*, 7 (2), 88-99.
- Ercan, A. G., Taskin, K. M., Turgut, K., Bilgen, M., & Firat, M. Z. (2002). Characterization of Turkish sesame (Sesamum indicum L.) landraces using agronomic and morphologic descriptors. *Ziraat Fakultesi Dergisi Akdeniz universitesi*, *15*(2), 45-52.
- Furat, S., & Uzun, B. (2010). The Use of Agro-morphological Characters for the Assessment of Genetic Diversity in Sesame ('Sesamum indicum'L). *Plant Omics*, *3*(3), 85-91.
- Gedifew, S. (2022). Characterization and evaluation of sesame (Sesamum indicum L.) accessions. International *Journal of Agricultural and Natural Sciences, 15*(3), 226-239.
- Georgiev, S., Stamatov, S., & Deshev, M. (2008). Requirements to sesame (Sesamum indicum L.) cultivars breeding for mechanized harvesting.

Bulgarian Journal of Agricultural Sciences, 14, 616-620.

- Jolliffe, I. T. (2002). *Principal component analysis for special types of data* (pp. 338-372). Springer New York.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, *20*(1), 141-151.
- Kassambara, A. (2023). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0. URL: https://CRAN.R-project.org/package=ggpubr.
- Kassambara, A., & Mundt, F. (2016). Factoextra: Extract and Visualize the Results of Multivariate Data Analyses, R package version 1 (2016). https://rpkgs.datanovia.com/factoextra/index.html.
- Khairnar, S. S., & Monpara, B. A. (2013). Identification of potential traits and selection criteria for yield improvement in sesame (Sesamum indicum L.) genotypes under rainfed conditions. *Iranian Journal of Genetics and Plant Breeding*, *2*(2), 1-8.
- Maity, A., Lamichaney, A., Joshi, D. C., Bajwa, A., Subramanian, N., Walsh, M., & Bagavathiannan, M. (2021). Seed shattering: a trait of evolutionary importance in plants. *Frontiers in Plant Science*, *12*, 657773. DOI: 10.3389/fpls.2021.657773.
- Mehmood, M. Z., Qadir, G., Afzal, O., Din, A. M. U., Raza, M. A., Khan, I., ... & Ahmed, M. (2021). Paclobutrazol improves sesame yield by increasing dry matter accumulation and reducing seed shattering under rainfed conditions. *International journal of plant production*. DOI: https://doi.org/10.1007/s42106-021-00132-w.
- Mukhthambica, K., Bisen, R., & Ramya, K. T. (2023). Principal Component Analysis for Yield and Yield Related Traits in Sesame (Sesamum indicum L.). *Biological Forum – An International Journal*, *15*(3), 227-232.
- 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. DOI: https://doi.org/10.1111/2041-210X.13384.
- Qureshi, M., Langham, D. R., Lucas, S. J., Uzun, B., & Yol, E. (2022). Breeding history for shattering trait in sesame: classic to genomic approach. *Molecular Biology Reports*, *49*(7), 7185-7194. DOI: https://doi.org/10.1007/s11033-022-07636-2.
- R Core Team (2022) R software version 4.2.2: A language and environment for statistical computing, R foundation for statistical computing, Vienna, Austria. URL: http:// www.R-project.org.
- Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science*, *21*(2), 110-124.
- Singh, B., Goswami, A. and Kumar, A. (2017). Biotechnology approaches for quality improvement of vegetables. *Ann. Hort*, *10*(1), 7-14.
- Teklu, D. H., Kebede, S. A., Gebremichael, D. E. (2017). Assessment of genetic variability, genetic advance, correlation and path analysis for morphological traits in sesame genotypes. *International Journal of Novel Research in Life Sciences*, 4, 34-44.
- Visioni, A., Tondelli, A., Francia, E., Pswarayi, A., Malosetti, M., Russell, J., ... & Comadran, J. (2013). Genome-wide association mapping of frost tolerance in barley (Hordeum vulgare L.). *BMC genomics*, *14*, 1-13.
- Wang, C., Niu, J., Miao, H., Li, C., Duan, Y., Ju, M., ... & Zhang, H. (2024). Genetic variation, correlation, and association mapping of seed yield and its component traits in sesame. *Genetic Resources and Crop Evolution*, *71*(2), 603-619.