Research Article Electronic Journal of Plant Breeding, Vol.17 (2):

Factor analysis-based evaluation of stalk yield and quality traits in sweet sorghum genotypes

Farrag F.B. Abu-Ellail[1*], Karim Y.M. Yousef[1], Nahid A. A. Morsi[2] and Ramy N. F. Abdelkawy[3*]

https://doi.org/10.37992/2026.1702.016

Abstract

Sweet sorghum is one of the most important global crops for energy, food, and feed, valued for its climate resilience and adaptability to diverse growing conditions. A total number of 25 sweet sorghum genotypes were evaluated to identify the best genotype (s) with high stalk yield and quality during the 2024 and 2025 growing seasons in alpha lattice design. The results showed significant variation among the tested genotypes across the eight evaluated traits. The superior genotypes exhibiting high average values for stalk and quality yield traits were MN4080, Rex, Honey, Cukarcirak, AGSC2, and Sugar drip, whereas Dall, Grcoba, Atlas, Williams, Roma, and MN1500 and Rio recorded the lowest performance across the evaluated traits. The stalk yield was significantly correlated with stalk diameter, juice extraction percentage and juice yield. To address multicollinearity and clarify complex trait relationships, factor analysis was applied, reducing the variables to three main factors that accounted for over 80 % of the total variation. The first factor, which included stalk diameter, stalk yield, juice extraction percentage, and juice yield, accounted for 40.17 % of the total variation. Both factor and cluster heat map analyses consistently grouped MN4080, Rex, Honey, Cukarcirak, AGSC2 and Sugar drip, as promising genotypes for enhancing sweet sorghum stalk yield and improving quality yield. Using factor scores in ANOVA proved more effective than relying on the original data, making these scores more reliable for selection purposes and applicable in sweet sorghum breeding programs.

Keywords: Sweet sorghum, stalk yield, juice yield, correlation, cluster heat map and factor analysis.

INTRODUCTION

Sweet sorghum (*Sorghum bicolor Moench, L.*) is a climate-resilient crop of global importance due to its ability to thrive in drought-prone and marginal environments. It serves as a key source of bioethanol, contributing to reduced dependence on fossil fuels (Salem *et al.,* 2025; Shanalin *et al.,* 2025). Beyond, sweet sorghum supports both food and feed security, providing grain for human consumption and fodder for livestock. It's adaptability and multipurpose use make it a promising and sustainable crop for addressing future global challenges. Sweet sorghum plays a significant role across several industries, including brewing, animal feed, food production, and sugar manufacturing (Bakari *et al.,* 2023; Khalifa and Eltahir, 2023; Abu-Ellail *et al.,* 2023b). It is increasingly recognized as a strategic crop for mitigating food insecurity, particularly in the context of climate change (Khalifa and Eltahir, 2023). Despite being a very promising crop with high potential, sweet sorghum cultivation in Egypt remains limited, primarily due to a lack of sufficient juicer extractors to process the stalks for syrup production. However, despite its vital role, the majority of sweet sorghum production particularly in Western countries is primarily used as animal feed, highlighting a gap in its broader utilization. Given its resilience to heat and drought, underutilization in Western diets, and its favorable nutritional and functional properties (Khoddami *et al.,* 2023), there is a compelling need to prioritize the development of improved genotypes that combine high biomass and grain yields with superior juice quality.

[1]Sugar Crops Research Institute (SCRI), Agricultural Research Center (ARC), Giza 12619, Egypt
[2]Cell Research Department, Field Crop Research Institute, Agricultural Research Center, Giza 12619, Egypt
[3]Central Laboratory for Design and Statistical Analysis Research, Agricultural Research Center, Giza 12619, Egypt

Corresponding author : Farrag F.B. Abu-Ellail, Sugar Crops Research Institute (SCRI), Agricultural Research Center (ARC), Giza 12619, Egypt. *E-Mail: farrag_abuellail@yahoo.com

How to cite this artilce: Farrag F.B. Abu-Ellail, Karim Y.M. Yousef, Nahid A. A. Morsi and Ramy N. F. Abdelkawy. 2026. Factor analysis-based evaluation of stalk yield and quality traits in sweet sorghum genotypes. *Electronic Journal of Plant Breeding*, 17(2): doi:10.37992/2026.1702.016

Received: 08.11.2025 **Accepted**: 10.04.2026 **Online**: 05.05.2026

Multivariate analyses are widely applied to summarize and interpret patterns of variation within crop genotype populations (Karthik and Hanamaratti, 2025). These statistical tools are effective for identifying key traits and reducing dataset complexity, enabling a clearer understanding of trait relationships and their variability, as seen in factor analysis (Abd El-Saber *et al.,* 2020; Shrestha, 2021; Mohiy and El-Moselhy, 2022). Additionally, allow for the statistical grouping of similar data points into clusters using methods such as hierarchical clustering.

This hierarchical clustering and heat map approach helps to identify clusters of genotypes with comparable effects on the measured characteristics.

In sweet sorghum breeding programs, the selection of superior genotypes relies on multiple traits, with particular emphasis on final stalk yield and its quality. Improving productivity, therefore, depends on effective selection strategies that can identify high-yielding genotypes well-suited to specific environmental conditions. The association between stalk yield and quality traits is also a critical consideration. In this study, factor analysis is employed to identify the key traits that significantly influence stalk yield. The cluster heatmap (CHM) illustrates the impact of diverse sweet sorghum genotypes on stalk yield and quality traits, utilizing Euclidean distance and complete linkage. The present study aimed to evaluate the performance of 25 sweet sorghum genotypes across nine traits to facilitate the selection of superior genotypes. It also sought to identify key stalk yield and quality traits that are significantly associated and contribute to the differentiation of genotypes into distinct groups. Furthermore, factor analysis was employed to determine the principal traits that significantly influence stalk and juice yield.

MATERIALS AND METHODS

Plant material and the experimentation

Field experiments were carried out over two consecutive growing seasons, May 2024 and May 2025, at the Agricultural Research Station in Giza Governorate, Egypt (30º 02' N latitude and 31º 13' E longitudes with an altitude of 22.50 meters above sea level). In each season, the sweet sorghum genotypes were evaluated using an alpha lattice design with three replications. Each replication consisted of 25 plots (25 genotypes), distributed across 5 blocks, with 5 experimental units per block. Genotypes were randomly assigned to these plots. Each plot comprised five rows, 3.5 meters in length, with a row spacing of 60 cm and plant spacing of 20 cm within the row. All agronomic practices for sweet sorghum cultivation were implemented in accordance with the standard recommendations of the Ministry of Agriculture and Land Reclamation. The names and origins of the genotypes studied are detailed in **Table 1**.

Observations and measurements

At harvest stage, a random sample of 10 plants was collected from each plot across two growing seasons to evaluate various traits, including stalk diameter (cm), stalk length (cm), number of plants /m², stripped stalk yield (tons/feddan), total soluble solids % (TSS), sucrose content % (SU), juice purity % (PU), juice extraction % (JE), and juice yield (tons/feddan). One feddan is equivalent to 4200 m². Stalk diameter (SD) was measured at the midpoint of the stalk, while stalk length (SL) was taken from the soil surface to the top of the panicle. TSS % was determined using a hand-held refractometer. Sucrose content in the clarified juice was measured with an automated saccharimeter following AOAC (2005) methods. Juice purity % (PU) was calculated using the formula: (Sucrose / Brix) × 100. Juice extraction % (JE) was computed as: (juice weight / stalk weight) × 100. Juice yield (JY), expressed in tons per feddan, and was calculated by multiplying the stripped stalk yield by the juice extraction percentage and dividing by 100.

Statistical analysis

Analysis of variance (ANOVA) was conducted separately for each growing season. Prior to performing the combined ANOVA over both seasons, Bartelett's test (1937) was applied to verify the homogeneity of error variances. For traits meeting the homogeneity assumption, a combined ANOVA was performed to assess the significance of differences among the 25 genotypes, treating replication and season as random effects, and genotypes as fixed factors. Statistical analyses were carried out following the procedures outlined by Gomez and Gomez (1984), using GenStat software version 19 (VSN International Ltd. 2017). Mean comparisons were performed using Tukey's test at the 5% significance level (Tukey, 1949). Additionally, simple correlation coefficients for stalk yield and associated traits across both seasons were calculated following Snedecor and Cochran (1981).

Correlation and cluster heat map analyses were done by R software (R Core Team, 2025), while factor analysis was conducted using SPSS software version 27 (IBM Corp. 2020). Following the confirmation of data suitability for factor analysis based on Bartlett's test of sphericity ($p < 0.05$), determinant of collinearity, and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy factor analysis was carried out following the approach described by Cattell (1965), to reduce a large set of correlated variables into a smaller number of uncorrelated components (factors). After factor extraction, the factor loading matrix was subjected to varimax orthogonal rotation, as proposed by Kaiser (1958), to enhance interpretability.

Communalities, representing the proportion of variance in each variable explained by the common factors, were estimated based on the highest correlation coefficients within each array, following the method of Seiller and Stafford (1985). Factors with eigenvalues greater than one (from a total of eight traits) were retained for subsequent ANOVA, in accordance with Shrestha (2021). The proportion of variance explained by each factor was calculated as the factor's sum of squared loadings divided by the number of variables, assuming orthogonal rotation. Cluster analysis was conducted using Euclidean distance as the dissimilarity measure, following the method described by Everitt (1993). The genetic parameters, including phenotypic coefficient of variation (PCV %), genotypic coefficient of variation (GCV %), broad-sense heritability (h^2_b %) and genetic advance (GA %), were determined following the method described by Allard (1960). Broad-sense heritability

Table 1. Key code, name and origin of twenty-five sweet sorghum genotypes

Key Code	Name	Origin
G1	MN4080	Mississippi
G2	Honey	Mississippi
G3	Williams	Kentucky/Georgia
G4	AgSc2	unknown
G5	Rex	Mississippi
G6	Sugar drip	Oklahoma
G7	Cukarcirak	unknown
G8	Brands	Mississippi
G9	AGSC3	unknown
G10	Willey	Mississippi
G11	Dall	Mississippi
G12	MN1054	Mississippi
G13	MN1500	Mississippi
G14	MN1383	Mississippi
G15	Roma	South Africa
G16	Ramada	Mississippi
G17	Rio	Texas
G18	Atlas	South Africa
G19	Rona	unknown
G20	GKAron	unknown
G21	MN4423	Mississippi
G22	SS301-1	Nigeria
G23	Grcoba	unknown
G24	MN2756	Mississippi
G25	Tracy	Texas

was computed using the equation $h^2_b = Vg / Vp$, where Vg represents the genotypic variance and Vp denotes the phenotypic variance.

RESULTS AND DISCUSSION

The results of Bartelett's test (1937) showed homogeneous error variance for most attributes in two years, allowing for a pooled analysis of variance over years.

Mean performance

Mean performances for the studied stalk traits are presented in **Table 2**. For stalk diameter, Rex genotype exhibited the largest significant diameter (2.82 cm), followed by MN4423 (2.74 cm), suggesting their superiority in structural robustness, which is often associated with higher biomass accumulation. In contrast, Grcoba had the smallest diameter (1.44 cm), with Dall close behind at 1.69 cm. Concerning stalk length, the genotypes *viz*., Tracy, Honey, Cukarcirak and MN1383 had a high length (254.20, 232.50, 225.60 and 220.60 cm, respectively), while Atlas, Williams, Dall and Rio were recorded the shortest stalk length (136.71, 142.20, 152.90 and 165.30 cm, respectively) (**Table 2**). However, greater plant height does not always translate into higher yield due to potential lodging or inefficient assimilate partitioning.

Regarding the number of plants per square meter, Brands genotype showed a highly significant number (49.83 plants/m^2), followed by Tracy (40.50 plants/m^2). Conversely, AGSC3 had the lowest plant density (19.50 plants/m^2), with Ramada genotype close behind at 19.67/m^2. In terms of stalk yield, MN4423, Rex, MN4080, Willey, GKAron and Sugar drip exhibited a highly significant yield of 15.08, 15.07, 13.93, 13.85, 13.74 and 13.67 ton/fed, respectively, whereas the lowest stalk yield was 8.54 and 8.93 for MN1500 and Roma, respectively (**Table 2**). The eleven genotypes (MN4080, Honey, Rex, Sugar drip, Cukarcirak, MN1054, Ramada, Rio, GKAron, MN4423 and SS301-1) showed a stalk yield above the general mean (11.76 ton/fed). Plant density showed considerable variation, with Brands and Tracy recording the highest values. Despite this, Brands produced relatively low stalk yield (8.71 ton/fed), suggesting that excessive plant density may lead to increased competition for resources, ultimately limiting yield performance. In contrast, genotypes such as MN4423 and Rex combined moderate plant density with superior stalk diameter and yield, highlighting the importance of balanced trait expression. The overall means for stalk diameter (2.13 cm), stalk length (194.78 cm), plant density (29.98 plants/m²), and stalk yield (11.76 ton/fed), along with

Table 2. Mean performance of sweet sorghum genotypes for stalk yield traits combined over 2024 and 2025 growing seasons

Code	Traits/ Genotype	Stalk diameter (cm)	Stalk length (cm)	Number of plants /m^2	Stalk yield (ton/fed)
G1	MN4080	2.28abcde	177.80ghij	31.50bcd	13.93ab
G2	Honey	2.29abcde	232.50abc	37.50abc	12.97abc
G3	Williams	1.81ef	142.20kl	30.17bcd	9.41cde
G4	AGSC2	2.19abcde	190.40fghi	29.67bcd	11.7abcde
G5	Rex	2.82a	180.00ghij	38.00abc	15.07a
G6	Sugar drip	2.25abcde	204.10cdefg	29.00bcd	13.67ab
G7	Cukarcirak	2.24abcde	225.60abcd	32.83bcd	12.56abcd
G8	Brands	2.04cdef	219.20bcdef	49.83a	8.71de
G9	AGSC3	2.24abcde	241.90ab	19.50d	11.34abcde
G10	Willey	2.58abc	185.60ghi	26.17bcd	13.85ab
G11	Dall	1.69ef	152.90jkl	32.50bcd	10.02bcde
G12	MN1054	1.88def	183.60ghi	25.67bcd	12.32abcde
G13	MN1500	2.02cdef	183.90ghi	28.33bcd	8.54e
G14	MN1383	1.88def	220.60bcde	22.67cd	11.08bcde
G15	Roma	2.02cdef	184.40ghi	28.00bcd	8.93de
G16	Ramada	2.09bcdef	174.10hij	19.67d	12.89abc
G17	Rio	2.13bcde	165.30ijkl	28.67bcd	11.93abcde
G18	Atlas	2.04cdef	136.7l	27.33bcd	9.67cde
G19	Rona	2.01cdef	201.70defgh	25.83bcd	10.26bcde
G20	GKAron	2.51abcd	195.20efgh	24.50bcd	13.74ab
G21	MN4423	2.74ab	217.20bcdef	30.17bcd	15.08a
G22	SS301-1	2.27abcde	216.00bcdef	24.83bcd	13.18abc
G23	Grcoba	1.44f	166.00ijk	26.83bcd	10.77bcde
G24	MN2756	1.72ef	218.30bcdef	39.83ab	10.72bcde
G25	Tracy	1.97cdef	254.20a	40.50ab	11.61abcde
Mean		**2.13**	**194.78**	**29.98**	**11.76**
Standard error		**0.13**	**5.47**	**3.14**	**0.73**

moderate to high coefficients of variation especially for plant density (25.60%) and stalk yield (15.30%) (**Table 2**) further confirm the presence of significant variability. This level of variation is advantageous for breeding programs, as it enhances the selection of superior genotypes.

Developing a comprehensive framework to assess the stalk yield characteristics of sweet sorghum genotypes in Egypt is crucial. This approach will provide a theoretical foundation for the production, utilization, and cultivation of the crop in the region. The consistent superiority of these genotypes across key stalk yield traits demonstrates their strong potential for both bioethanol production and fresh consumption. Their enhanced performance is likely due to favorable combinations of yield-related attributes particularly larger stalk diameter and higher stalk yield which play central roles in determining overall productivity and processing efficiency (Zaghloul *et al.,* 2022; Abu-Ellail *et al.,* 2023a, b; Abu-Ellail *et al.,* 2025).

In contrast, genotypes such as Dall, Grcoba, Atlas, Williams, Roma, and MN1500 showed comparatively lower performance across the measured traits, indicating reduced agronomic suitability under the evaluated conditions. This variation highlights the importance of multi-trait selection approaches in identifying elite genotypes capable of meeting both yield and quality requirements for commercial cultivation and breeding programs. The present results align closely with previous reports, which similarly emphasized the value of integrated trait selection for improving sugarcane productivity and processing potential (Zaghloul *et al.,* 2022; Abu-Ellail *et al.,* 2023a, b; Abu-Ellail *et al.,* 2025).

Mean performances for the quality of studied stalk yield traits are given in **Table 3**. Concerning total soluble solids percentage (TSS %), Honey, MN4080, and SS301-1 showed the highest mean value, ranged from 18.71 to 19.43 %. Meanwhile, the lowest mean value

Table 3. Mean performance of sweet sorghum genotypes for quality yield traits combined over 2024 and 2025 growing seasons

Code	Traits Genotype	Total soluble solids (%)	Sucrose (%)	Purity (%)	Juice extraction (%)	Juice yield (ton/fed)
G1	MN4080	18.90ab	8.68bc	45.86g	34.28abc	4.87ab
G2	Honey	19.43a	9.58abc	49.79efg	39.82a	5.24a
G3	Williams	17.07abcde	9.33abc	55.88bcdefg	26.93bcdefg	2.57efghij
G4	AGSC2	18.27abc	10.01abc	54.93cdefg	34.23abc	3.96abcdefg
G5	Rex	17.17abcde	8.35c	49.33efg	30.8abcde	4.64abc
G6	Sugar drip	16.73abcde	10.14abc	60.54abcdefg	30.29abcdef	4.13abcde
G7	Cukarcirak	16.23abcde	10.57abc	65.52abcde	36.90ab	4.69abc
G8	Brands	14.63de	10.82ab	74.48a	22.21defg	1.97hij
G9	AGSC3	15.90bcde	10.01abc	63.16abcdef	28.96abcdefg	3.22bcdefghij
G10	Willey	14.93cde	10.50abc	71.52ab	26.87bcdefg	3.71abcdefgh
G11	Dall	15.96bcde	9.64abc	61.59abcdefg	18.31g	1.84ij
G12	MN1054	14.33e	9.77abc	67.94abcd	27.01bcdefg	3.29bcdefghij
G13	MN1500	18.16abc	8.73bc	48.19fg	19.68fg	1.67j
G14	MN1383	15.79bcde	10.70abc	68.5abc	20.92efg	2.34fghij
G15	Roma	15.89bcde	9.53abc	59.96abcdefg	25.24cdefg	2.19hij
G16	Ramada	16.07abcde	9.10abc	58.73abcdefg	31.42abcde	4.04abcdef
G17	Rio	16.40abcde	8.38c	51.62defg	21.90defg	2.66efghij
G18	Atlas	18.21abc	10.65abc	59.14abcdefg	24.30cdefg	2.27ghij
G19	Rona	17.90abcd	11.26a	63.12abcdef	29.08abcdefg	2.88defghij
G20	GKAron	16.62abcde	11.24a	67.8abcd	32.31abcd	4.44abcd
G21	MN4423	17.84abcd	10.03abc	56.42bcdefg	22.00defg	3.32bcdefghij
G22	SS301-1	18.71ab	11.03ab	59.19abcdefg	22.56defg	2.98cdefghij
G23	Grcoba	16.81abcde	10.07abc	59.93abcdefg	21.33efg	2.29ghij
G24	MN2756	17.03abcde	9.81abc	57.76bcdefg	23.11defg	2.44efghij
G25	Tracy	16.26abcde	10.79ab	66.37abcd	29.38abcdef	3.45bcdefghi
Mean		**16.85**	**9.95**	**59.89**	**27.19**	**3.24**
Standard error		**0.65**	**0.44**	**3.10**	**2.05**	**0.33**

was obtained in the genotypes *viz*., MN1054, Brands and Willey, ranged from 14.33 to 14.93 %. These genotypes exhibited a value lower than the general mean (16.85 %) (**Table 3**). The high TSS % suggests their strong capability for concentrating soluble solids, which is a key indicator of sugar-rich stalks, whereas the low values imply limited sugar accumulation efficiency. Regarding sucrose percentage, the Rona and GKAron exhibited the highest values (11.26 and 11.24 %, respectively), demonstrating their strong potential for sugar recovery, while the Rex (8.35%), Rio (8.38 %), and MN4080 (8.68 %) showed the lowest sucrose percentages, which may indicate higher proportions of non-sucrose components. For purity percentage, Brands exhibited the highest value (74.48 %), followed by Willey (71.52 %), reflecting cleaner juice with fewer impurities. The MN4080, Rex, and Honey genotypes showed low purity percentages (45.86, 49.33, and 49.79 %, respectively), suggest higher amounts of non-sucrose materials affecting sugar processing efficiency (**Table 3**). Concerning juice extraction percentage, Honey, Cukarcirak, AGSC2 and MN4080 genotypes showed the high of juice extraction percentage (39.82, 36.90, 34.23 and 34.28 %, respectively), indicate superior stalk juiciness and ease of mechanical extraction. Meanwhile Dall, MN1500, and MN1383 showed the lowest juice extraction percentage. For juice yield, four genotypes exhibited the highest juice yield: Honey, MN4080, Cukarcirak and Rex ranged from 5.24 to 4.64 ton/fed, reflect their combined advantages in juice volume and stalk biomass. Meanwhile, MN1500, Dall, Brands and Roma showed the lowest juice yield. The twelve genotypes (MN4080, Honey, AGSC2, Rex, Sugar drip, Cukarcirak, Willey, MN1054, Ramada, GKAron, MN4423 and Tracy) showed a juice yield above the general mean (3.24 ton/fed).

In general, the best genotypes with high mean values overall studied traits were MN4080, Honey, AGSC2, Rex, Sugar drip and Cukarcirak, while Williams, Dall, MN1500, Roma, Rio, Atlas and Grcoba were the lowest genotypes (**Table 3**). The significant variation among genotypes in TSS %, sucrose percentage, purity, juice extraction percentage, and juice yield reflects substantial genetic diversity in stalk quality traits. These differences indicate genotype-specific variation in physiological efficiency, sugar synthesis, and assimilate storage capacity. Higher TSS %, sucrose percentage, and purity values suggest superior sugar accumulation potential, consistent with observations by Zaghloul *et al.* (2022) in sorghum. Likewise, differences in juice extraction and juice yield point to variability in stalk structure and fiber content, as highlighted by Abu-Ellail *et al.* (2023; 2025). Overall, the diversity observed provides valuable opportunities for selecting genotypes with enhanced sugar content and improved processing quality.

Correlation analysis

The data in **Fig. 1** denote the relationship between stalk and quality yield traits of twenty-five sweet sorghum genotypes. According to Pearson's correlation coefficients, several traits exhibited significant positive associations. Stalk yield showed strong positive correlations with stalk diameter, juice extraction percentage, and juice yield. In addition, stalk length was positively correlated with sucrose percentage, while sucrose percentage also exhibited a significant positive relationship with purity percentage. Notably, a very strong positive correlation was observed between juice extraction percentage and juice yield ($r = 0.90^{**}$) (**Fig.1**), followed by the correlations of stalk yield with stalk diameter and juice yield ($r = 0.82^{**}$ and 0.80^{**}, respectively). This suggests that these parameters were positively associated, as they tended to increase or decrease together. This implies a positive interrelationship among these parameters, as changes in one were accompanied by similar changes in the others. Moreover, the significant positive correlation between sucrose content and purity % indicates that selecting genotypes with higher juice output may also lead to improved sucrose accumulation, which is desirable for both quality and processing efficiency.

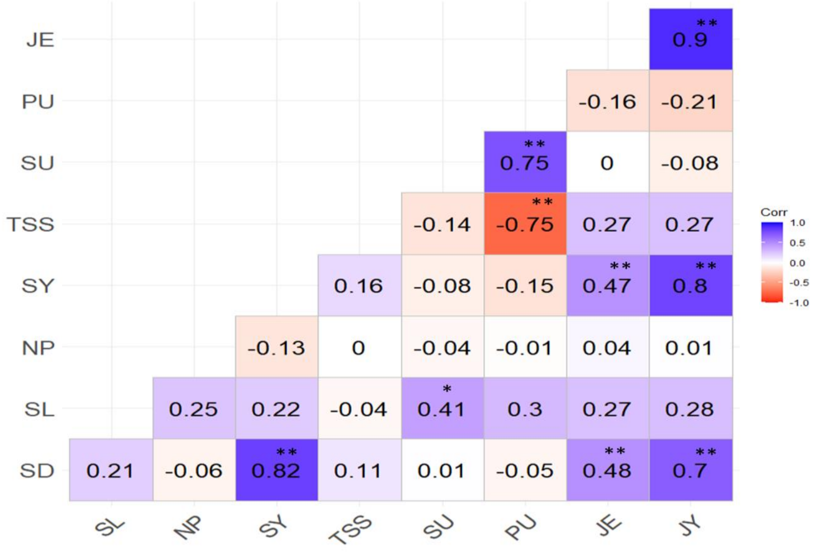
Although the correlation between stalk length (SL) and sucrose % (SU) was relatively low ($r = 0.41^{*}$) (**Fig. 1**), it remained statistically significant, implying that longer stalks may slightly favor sugar accumulation. In contrast, some negative correlations were observed, such as between total soluble solids % (TSS) and purity % (PU) ($r = -0.75$), this finding implies that high TSS values in sweet sorghum may reflect the accumulation of non-sugar constituents rather than sucrose, which reduces juice purity %. Consequently, relying solely on TSS as a selection criterion could be misleading, and integrating both TSS and purity % provides a more accurate assessment of sugar yield potential. The weak negative relationship between juice extraction (JE) and purity % (PU) may indicate that high extraction efficiency does not always correspond to higher purity, possibly due to differences in stem structure or moisture content, which can influence the proportion of non-sugar constituents in the extracted juice. Similar observations have been reported in sweet sorghum by Shukla *et al*. (2017); and Kanbar *et al*.(2020).

The results emphasize the importance of considering correlated traits in selection strategies to achieve balanced genetic improvement without compromising key yield and quality attributes. The observed associations indicate the potential for indirect selection, whereby improvement in one trait may lead to favorable changes in related traits, thereby enhancing the efficiency of genotype evaluation and breeding for dual-purpose sweet sorghum (Shukla *et al*., 2017; Kanbar *et al*., 2020). However, reliance solely on correlation coefficients may be insufficient, as they only describe pairwise relationships and fail to capture the complex, multivariate interactions among traits (Abd El-Saber *et al*., 2020). This limitation is particularly critical for complex traits such as yield, which are governed by multiple interdependent components. In this context, factor analysis provides a more comprehensive approach by identifying the underlying structure of trait interrelationships and highlighting the most influential variables contributing to yield performance. Consequently, integrating correlation and factor analysis offers a more reliable framework for selection decisions and genetic improvement in sweet sorghum breeding programs.

Factor analysis

Fig. 2 presents the eigenvalues of each component from the initial solution, with the scree plot indicating the optimal number of components based on the last notable decline in eigenvalues, where the slope becomes less than one, typically after the third component. This suggests that the first three components are most appropriate. As shown in **Table 4**, factor analysis results for the sweet sorghum traits (including eigenvalues, variance percentages, and cumulative variance) before and after rotation revealed that only three out of eight variables had eigenvalues exceeding one and were thus retained (Pattanashetti *et al*., 2025).

The first three extracted factors together explained 80.12 % of the total variation among genotypes (6.41 eigenvalue out of 8 variables), indicating a strong underlying structure within the dataset (**Table 4**). The first factor, with an eigenvalue of 3.21, accounted for 40.17 % of the variance and was primarily defined by stalk diameter, stalk yield, juice extraction percentage, and juice yield. The second factor, associated mainly with stalk length, sucrose percentage, and purity, explained 24.86 % of the variance (eigenvalue = 1.99), indicating a distinct quality-related dimension. The third factor (eigenvalue = 1.21) contributed 15.09 % of the variance and was mainly related to plant density (plants/m²).



	SL	NP	SY	TSS	SU	PU	JE	JY
JE								0.9**
PU							-0.16	-0.21
SU						0.75**	0	-0.08
TSS					-0.14	-0.75**	0.27	0.27
SY				0.16	-0.08	-0.15	0.47**	0.8**
NP			-0.13	0	-0.04	-0.01	0.04	0.01
SL		0.25	0.22	-0.04	0.41*	0.3	0.27	0.28
SD	0.21	-0.06	0.82**	0.11	0.01	-0.05	0.48**	0.7**

Fig. 1. Pearson's correlation coefficients between yield and the studied traits

* and ** indicates significant at the 0.05 and 0.01 level of probability, respectively

Dark blue shades indicate strong positive correlations and red shades indicate negative correlations.

Where, stalk diameter (SD), stalk length (SL), number of plants /m^2 (NP), stalk yield (SY), total soluble solids % (TSS), sucrose% (SU), purity% (PU), juice extraction% (JE), and juice yield (JY).

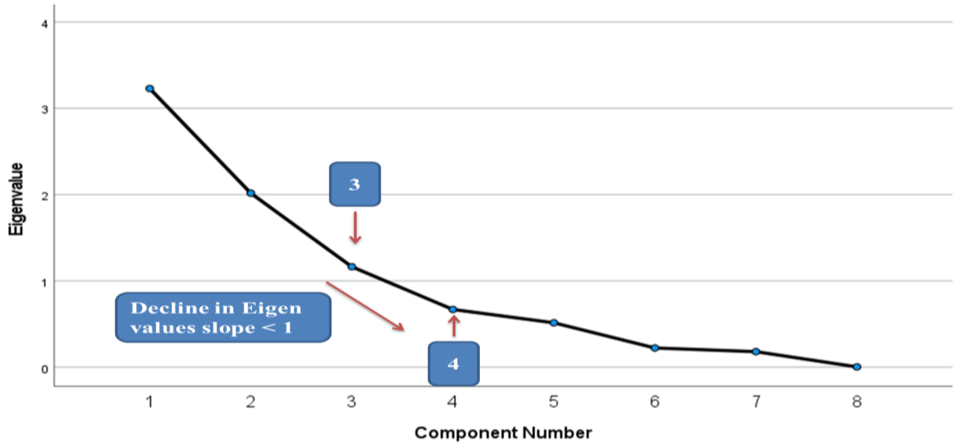


Fig. 2. The scree plot illustrates the eigenvalues corresponding to the component number response for the evaluated traits

These traits showed the highest loadings and also closely related to biomass production and productivity, suggesting that they form a primary component of yield performance in sweet sorghum. High communality values across variables indicate that the extracted factors effectively capture the majority of the variation in the dataset. For example, the communality of juice yield was 0.94, meaning that 94 % of its variance is explained by the first three factors, highlighting the strong representativeness of these factors for key traits.

The communality value represents the proportion of a variable's variance that is explained by the retained factors. In this case, the communality of juice yield is 0.94, indicating that the first three extracted factors account for 94 % of the total variability in juice yield, while the remaining 6 % is attributed to unique or unexplained variance (including measurement error or other factors). This high communality suggests that the selected factors adequately capture the variation in juice yield. Such high communalities demonstrate that the factor analysis model reliably summarizes complex trait interrelationships, supporting its use in identifying the most influential traits and guiding selection decisions in sweet sorghum breeding programs. High factor loading values indicate strong correlations between individual variables and their

Table 4. Varimax rotated factor analysis results for eight studied sweet sorghum traits

Traits	Rotated component matrix			Communality	Loading
	Factor1	Factor2	Factor3		
Stalk diameter	0.85	0.05	-0.12	0.74	0.85
Stalk length	0.34	0.55	0.52	0.68	0.55
Number of plants/m^2	-0.08	-0.06	0.92	0.85	0.92
Stalk yield	0.89	-0.03	-0.16	0.81	0.89
Sucrose percent	-0.02	0.93	-0.01	0.86	0.93
Purity percent	-0.16	0.90	-0.05	0.83	0.90
Juice extraction percent	0.79	-0.05	0.23	0.68	0.79
Juice yield	0.96	-0.09	0.11	0.94	0.96
Eigen value	3.21	1.99	1.21	6.41	
Variance (%)	40.17	24.86	15.09	80.12	
Cumulative variance (%)	40.17	65.03	80.12		

corresponding components, meaning that variables with higher loadings serve as clearer and more direct indicators of the underlying factors. Such variables contribute most to defining the structure and interpretation of each component. The rotated component matrix therefore plays a crucial role in clarifying these relationships, enhancing interpretability, and ensuring that each factor represents distinct trait groupings. These insights align with the observations of Abd El-Saber *et al.* (2020) and Mohiy and El-Moselhy, (2022), who emphasized the importance of rotation in revealing meaningful component patterns.

Future analyses may primarily focus on stalk diameter, juice extraction percentage, and juice yield, although additional insights can still be obtained by saving component. Overall, the results indicate that selecting genotypes with higher positive factor score (FS) values is an effective approach for identifying superior sweet sorghum candidates for both stalk yield and quality traits. The factor score values for the first selected factor—representing the most yield-influential traits—were used to evaluate the genotypes. Using the first factor scores (FS) enabled effective screening of genotypes based on key yield traits. As shown in **Fig. 3**, genotypes were classified based on FS values: those with high stalk and quality yields appeared on the positive side of the histogram, while genotypes with lower yields showed negative FS values, indicating poorer performance. As shown in the histogram (**Fig. 3**), thirteen genotypes (Rex, Honey, MN4080, GKAron, Cukarcirak, MN4423, Sugar drip, Willey, Ramada, AGSC2, SS301-1 and Tracy) displayed positive factor score values for stalk and quality yield traits, placing them in the positive region. The negative FS values observed in the remaining genotypes indicate comparatively lower yield potential, highlighting their limited suitability for selection. In contrast, genotypes such as Rex, Honey, MN4080, and GKAron exhibited the highest FS values, marking them as the most promising candidates for breeding and cultivation. These results are consistent with Abd El-Saber *et al.* (2020), reinforcing the utility of factor scores in effectively discriminating high-performing genotypes based on combined trait performance.

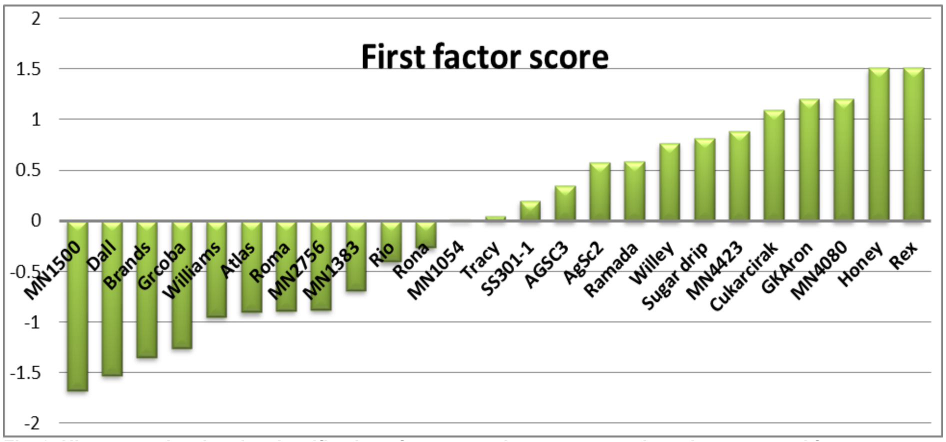


Fig. 3. Histogram showing the classification of sweet sorghum genotypes based on extracted factor scores of first factor concluding the best yield traits

Cluster heatmap

Cluster heatmap analysis is a valuable tool for uncovering the underlying relationships among genotypes and organizing them into a hierarchical structure. In this study, hierarchical clustering using the complete linkage method was applied to differentiate the 25 evaluated genotypes. Based on stalk yield and quality-related traits, the genotypes were grouped into distinct clusters, as depicted in the cluster heat map (**Fig. 4**). The traits were clustered into three main groups. The first group included stalk yield, stalk diameter, juice yield, and juice extraction %. The second group included stalk length, sucrose percentage, and purity %. The third group only included the number of plants per m². These results are the same as those obtained from factor analysis.

The cluster heatmap showed significant variation among the 25 sorghum genotypes based on eight yield and quality traits. The genotypes were grouped into distinct clusters, with some genotypes sharing similar trait profiles. Genotypes *viz*., MN4080, Rex, Honey, Cukarcirak, AGSC2 and Sugar drip showed high values (red shades) in stalk yield, stalk diameter, juice yield, and juice extraction, indicating superior performance in productivity traits and suggests their potential for selection in breeding programs aimed at improving bioethanol production or high yield for animal feed. Conversely, genotypes such as Dall, Grcoba, Atlas, Williams, Roma, MN1500 and Rio were characterized by lower values (blue shades) across most traits. Brands exhibited a particularly high value for purity % and the number of plants per m², while Tracy exhibited the highest value for stalk length. Meanwhile AGSC3, SS301-1, MN1383, Rona, MN4423, Willey, GKAron, MN1054 and Ramada showed relatively balanced performance across most traits, these genotypes may require improvement or could be used as contrasting parents in hybridization schemes. Traits such as sucrose content and purity showed significant variation among genotypes, helping in distinguishing those with higher quality.

The clustering patterns revealed that several genotypes share similar physiological and agronomic characteristics, particularly regarding juice-related traits and key yield components. This demonstrates the effectiveness of hierarchical clustering in grouping genotypes with comparable performance profiles. When combined with heatmaps, this approach simplifies the interpretation of complex multivariate relationships, allowing clearer visualization of trait associations and genotype similarities. A multivariate approach that integrates both productivity and quality traits enhances selection efficiency by allowing simultaneous evaluation of multiple important attributes. This framework facilitates the identification of superior sweet sorghum genotypes with balanced performance across key traits. The findings align with recent

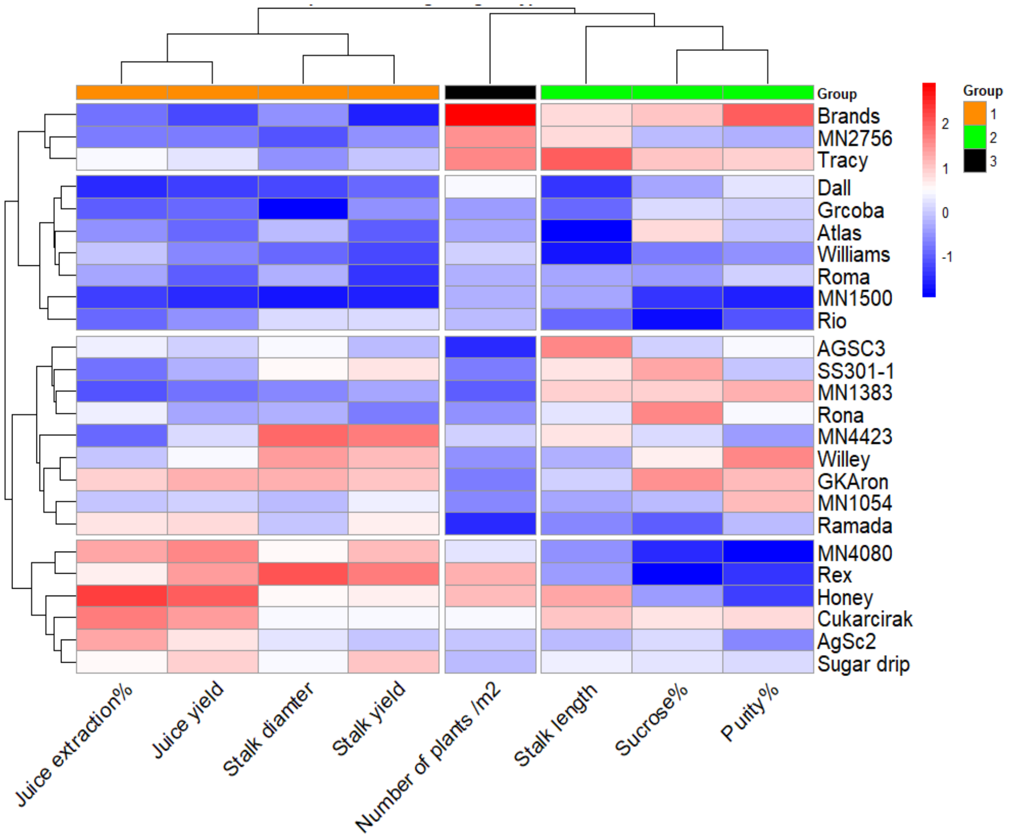


Fig. 4. Cluster heatmap of sweet sorghum genotypes under studied traits

Table 5. Variability coefficient, heritability, and predicted genetic advance for studied traits

Traits	PCV (%)	GCV (%)	h^2_b (%)	GA (%)
Stalk diameter	21.22	19.45	84.01	36.73
Stalk length	21.99	21.62	96.74	43.81
Number of plants/m^2	32.55	28.99	79.32	53.18
Stalk yield	23.27	21.53	85.58	41.03
Sucrose percent	12.15	10.38	73.11	18.29
Purity percent	17.53	15.94	82.60	29.84
Juice extraction	29.48	27.47	86.86	52.74
Juice yield	45.27	42.95	89.99	83.93

studies by Boukrouh *et al*. (2023), Somu *et al*. (2024), Alzahrani *et al*. (2025), and Bidhari *et al*. (2025), which similarly highlight the value of multivariate analysis for effective genotype discrimination and selection in sorghum and sweet sorghum.

Genetic variability

The phenotypic coefficient of variation (PCV) values for the studied sweet sorghum traits ranged from 12.15 % for sucrose content to 45.27 % for juice yield (**Table 5**), while the genotypic coefficient of variation (GCV) values followed a similar trend, ranged from 10.38 % to 42.95 %. The highest PCV and GCV were observed for juice yield (45.27 and 42.95 %, respectively) and stalk diameter (40.35 and 39.32 %, respectively), indicating high variability among genotypes. Broad-sense heritability (h^2_b) estimates were generally high, ranged from 73.11 % for sucrose content to 96.74 % for stalk length. The predicted genetic advance as a percentage of the mean (GA %) was highest for juice yield (83.93 %) and stalk diameter (36.73 %), followed by stalk length (43.81 %) (**Table 5**), indicating the potential for effective selection in these traits. In contrast, the sucrose and purity percentage demonstrated the lowest genetic advance value 18.29 % and 29.84 %, respectively. The phenotypic coefficient of variation (PCV) slightly exceeded the genotypic coefficient of variation (GCV), consistent with earlier findings by Adedugba *et al*. (2023). This narrow difference indicates that environmental influences on the expression of these traits are minimal, suggesting strong genetic control (**Table 5**).

The high heritability values (h^2_b) observed for most traits particularly stalk length (96.74 %), and juice yield (89.99 %), in combination with high genetic advance percentages, suggest that these traits are largely governed by additive gene action and can be effectively improved through direct selection. The close correspondence between PCV and GCV values for most traits indicates that environmental influence on trait expression is minimal, thus enhancing the reliability of selection. Traits such as sucrose content, which showed relatively lower heritability (73.11 %) and genetic advance (18.29 %), may require more cycles of selection or the adoption of indirect selection strategies. The high genetic advance, heritability, phenotypic and genotypic coefficients of variation observed for juice yield and number of plants/m^2 reflect a strong potential for genetic improvement, aligning with previous findings that emphasize these traits as key contributors to overall productivity and quality in sweet sorghum.

Heritability and genetic gain are critical indicators in breeding programs, as they reflect the proportion of observed variation that is genetically controlled and the expected improvement from selection, respectively. High heritability suggests that trait expression is largely governed by genetic factors rather than environmental influences, making selection more efficient and reliable. When coupled with substantial genetic gain, these traits can be effectively improved in successive breeding cycles, ultimately enhancing sorghum yield. The importance of these parameters in guiding selection decisions has been highlighted in several studies (Gebregergs and Mekbib, 2020; Adedugba *et al*., 2023; Somu *et al*., 2024), underscoring their role in identifying superior genotypes for future cultivation and breeding efforts.

Evaluation of 25 sweet sorghum genotypes revealed substantial variation in stalk and quality yield traits, indicating strong potential for genetic improvement. Among the studied genotypes, MN4080, Rex, Honey, Cukarcirak, AGSC2, and Sugar Drip consistently exhibited superior performance across key traits, highlighting their suitability for dual-purpose applications in bioethanol production and animal feed. Correlation analysis demonstrated strong positive associations among major yield-related traits, particularly between juice extraction percentage and juice yield, as well as between stalk yield and stalk diameter, juice yield, and juice extraction, supporting the efficiency of indirect selection in breeding programs. Furthermore, factor analysis extracted three principal components explaining over 80 % of the total variation, with stalk diameter, stalk yield, juice yield, and juice extraction identified as the most influential traits contributing to overall variability. Based on factor scores, genotypes were effectively differentiated into distinct groups, facilitating more precise selection using multivariate criteria that integrate both yield performance and key component traits. Overall, the integration of variability assessment, correlation analysis, and factor analysis provided a comprehensive understanding of trait relationships and genetic divergence, offering a strong foundation for the development of improved sweet sorghum genotypes with enhanced biomass and sugar productivity.

Author contributions

The main contributor to this article is FFBA. The study was conceptualized and data were collected by FFBA and KYMY. NAAM created the figures and tables, RNFA conducted data analysis, and contributed to writing the main content of the article. All authors made significant contributions to the development of the manuscript. All authors have reviewed and approved the final version of the manuscript.

ACKNOWLEDGMENT

The authors gratefully acknowledge and extend their sincere thanks to their institutions the Sugar Crops Research Institute, the Field Crops Research Institute, and the Central Laboratory for Design and Statistical Analysis Research all affiliated with the Agricultural Research Center.

REFERENCES

Abd El-Saber, A., Ghareeb, Z. E. and Ahmed, M. A. 2020. Multivariate analyses of yield and its components in some peanut genotypes. *Asian Journal of Advanced Agricultural Research,* **13**(4): 1-15. [Cross ref]

Abu-Ellail, F.F.B., Sadan, A. S.A. and Fares, W.M. 2023a. Cluster analysis and genetic variability of sweet sorghum (*Sorghum bicolor* L. Moench) genotypes using agro-morphological and juice quality traits. *Electronic Journal of Plant Breeding,* **14**(2): 371-382. [Cross ref]

Abu-Ellail, F. F.B., Ghareeb, Z. E. and Attia, A. E. 2023b. Genetic analysis for morphological, quality and biofuel related traits in sweet sorghum [*Sorghum bicolor* (L.) Moench]. *Indian Journal of Genetics and Plant Breeding,* **83**(4): 508-517. [Cross ref]

Abu-Ellail, F. F., Salem, E. A. S., Grad, W. E. and El-Sheikh, A. M. 2025. Heritability, genetic variability and performance of some sweet sorghum varieties for physiological, quality and yield parameters influenced by soil salinity. *Egyptian Sugar Journal*, **23**: 33-44. [Cross ref]

Adedugba, A. A., Adeyemo, O. A., Adetumbi, A. J., Amusa, O. D. and Ogunkanmi, L. A. 2023. Evaluation of genetic variability for major agro-morphological and stalk sugar traits in African sorghum genotypes. *Heliyon*, **9**(3). [Cross ref]

Allard, R. W. 1960. Principles of Plant Breeding. John Wiley and Sons, New York, pp. 92–94.

Alzahrani, Y., Abdulbaki, A. S. and Alsamadany, H. 2025. Genotypic variability in stress responses of Sorghum bicolor under drought and salinity conditions. *Frontiers in Genetics*, **15**, 1502900. [Cross ref]

AOAC. 2005. Association of Official Agricultural Chemist. Official Methods of Analysis. Published by the A.O.A.C., Washington, U.S.A.

Bakari, H., Djomdi and Ruben, Z. F. 2023. Sorghum (*Sorghum bicolor* L. Moench) and its main parts (By-products) as promising sustainable sources of value-added ingredients. *Waste and Biomass Valorization*, **14**: 1023–1044. [Cross ref]

Bartelett, M.S. 1937. Some examples of statistical methods of research in agricultural and applied biology. *Supplement to the Journal of the Royal Statistical Society,* **4**(2): 137-183. [Cross ref]

Bidhari, L. A., Santoso, S. B. and Syahruddin, K. 2025. Identification of high biomass sweet sorghum varieties using multivariate analysis. In IOP Conference Series: *Earth and Environmental Science*, **1471**, 012028. [Cross ref]

Boukrouh, S., Noutfia, A., Moula, N., Avril, C., Louvieaux, J., Hornick, J. L., Cabaraux, J.F. and Chentouf, M. 2023. Ecological, morpho-agronomical, and bromatological assessment of sorghum ecotypes in Northern Morocco. *Scientific Reports,* **13**(1): 15548. [Cross ref]

Cattell, R.B. 1965. Factor analysis: An introduction to essentials. The purpose and underlying models. *Biometrics*, **21:** 190–215. [Cross ref]

Everitt, B.S. 1993. Cluster analysis. Wiley, New York, USA.

Gebregergs, G. and Mekbib, F. 2020. Estimation of genetic variability, heritability, and genetic advance in advanced lines for grain yield and yield components of sorghum [*Sorghum bicolor* (L.) Moench] at Humera, Western Tigray, Ethiopia. *Cogent Food and Agriculture*, **6**(1): 1764181. [Cross ref]

Gomez, K.A. and Gomez, A.A. 1984. Statistical procedures for agricultural research. John Wiley & Sons.

IBM Corp. 2020. IBM SPSS Statistics for Windows, Version 27.0. IBM Corp., Armonk, New York, USA.

Kaiser, H.F. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**(3): 187-200. [Cross ref]

Kanbar, A., Shakeri, E., Alhajturki, D., Horn, T., Emam, Y., Tabatabaei, S.A. and Nick, P. 2020. Morphological and molecular characterization of sweet, grain and forage sorghum (*Sorghum bicolor* L.) genotypes grown under temperate climatic conditions. *Plant Biosystems*, **154**(1): 49–58. [Cross ref]

Karthik, R. and Hanamaratti, N. G. 2025. Multivariate analysis and multitrait genotype-ideotype distance index (MGIDI) for selection of promising genotypes under drought stress in post rainy sorghum (*Sorghum bicolor* L. Moench). *Electronic Journal of Plant Breeding*, **16**(1). [Cross ref]

Khalifa, M. and Eltahir, E. A. 2023. Assessment of global sorghum production, tolerance, and climate risk. *Frontiers in Sustainable Food Systems*, **7**: 1184373. [Cross ref]

Khoddami, A., Messina, V., Vadabalija Venkata, K., Farahnaky, A., Blanchard, C. L. and Roberts, T. H. 2023. Sorghum in foods: Functionality and potential in innovative products. *Critical Reviews in Food Science and Nutrition*, **63**(9): 1170-1186. [Cross ref]

Mohiy, M. M. and El-Moselhy, O. 2022. Evaluation of grain yield and its components of some bread wheat promising lines under low input conditions in upper Egypt region. *Journal of Plant Production*, **13**(8): 543-548. [Cross ref]

Pattanashetti, B.M., Sajjanar, G.M., Maralappanavar, M.S., Swaroop, K.O. and Soregaon, C.D. 2025. Heterosis and combining ability analysis using landrace-based hybrids for yield and yield components in *rabi* sorghum. *Electronic Journal of Plant Breeding*, **16**(4): 489-495. [Cross ref]

R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Salem, K. F., Abu-Ellail, F. F.B., Ibrahim, A. A., Saleh, M. M., Abdel Razik, E. S. S., Abd-Ellatif, S. and Sakr, E. A. 2025. Genetic improvement of sorghum (*Sorghum bicolor* L. Moench) through biotechnology. In Breeding and Biotechnology of Grass and Bast Fiber Crops, pp. 197-236. Cham: Springer Nature Switzerland. [Cross ref]

Seiller, G.J. and Stafford, R.E. 1985. Factor analysis of components in Guar. *Crop Science*, **25:** 905–908. [Cross ref]

Shanalin, J., Chandirakala, R. and Chandrakumar, K. 2025. Profiling of proximate composition in red sorghum (*Sorghum bicolor* L. Moench) genotypes for nutritional enhancement. *Electronic Journal of Plant Breeding*, **16**(3): 381-387. [Cross ref]

Shrestha, N. 2021. Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, **9**(1): 4-11. [Cross ref]

Shukla, S., Felderhoff, T.J., Saballos, A. and Vermerris, W. 2017. The relationship between plant height and sugar accumulation in the stems of sweet sorghum (*Sorghum bicolor* (L.) Moench). *Field Crops Research*, **203:** 181-191. [Cross ref]

Snedecor, G.W. and Cochran, W.G. 1981. Statistical methods. 7th Ed. Iowa State University Press, Ames, Iowa, USA.

Somu, G., Meena, N. and Badigannavar, A. 2024. Exploring genetic variability for morphological and yield contributing traits in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Southern India. *Genetic Resources and Crop Evolution*, **72**: 5179–5194. [Cross ref]

Tukey, J. W. 1949. Comparing individual means in the analysis of variance. *Biometrics*, **5**(2): 99-114. [Cross ref]

VSN International Ltd. 2017. GenStat for Windows, 19th Edition. VSN International Ltd., Hemel Hempstead, UK.

Zaghloul, M.A., El Hosary, A.A., EL-Badawy, M.EL.M. and Khalifa, E.A. 2022. Combining ability for some quantitative traits in sorghum (*Sorghum bicolor* L. Moench) using 9x9 half diallel cross scheme. *Annals of Agricultural Science*, **60**(1): 1-8. [Cross ref]