INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print) 2222-5234 (Online)
http://www.innspub.net
Vol. 26, No. 5, p. 209-220, 2025

RESEARCH PAPER OPEN ACCESS

Multivariate analysis-based screening of wheat genotypes for agronomic performance and yield potential in Bangladesh

Md. Hadiur Rahman[1], Nurunnaher Akter*[1], Md. Abdul Karim[1], Nasrin Akter Ivy[2], Md. Rafiqul Islam[1]

[1]*Department of Agronomy, Gazipur Agricultural University, Gazipur, Bangladesh*
[2]*Department of Genetics and Plant Breeding, Gazipur Agricultural University, Gazipur, Bangladesh*

Key words: Genetic diversity, Discriminant function analysis, Cluster analysis, Wheat genotypes, Grain yield

http://dx.doi.org/10.12692/ijb/26.5.209-220 Article published on May 09, 2025

Abstract

Identifying genetically diverse and high-yielding wheat genotypes is crucial for increasing productivity and addressing food security concerns in Bangladesh. The objective of this study was to assess the genetic diversity and agricultural potential of 49 wheat genotypes in Bangladesh using multivariate analyses. Genotypes were analyzed for 31 morpho-physiological, phenological, and yield-related parameters. Cluster analysis divided the genotypes into seven distinct clusters, which exhibited significant intercluster variation. Clusters V and VI are the most promising regarding higher grain yield, earlier physiological maturity, and efficient biomass accumulation. On the other hand, productivity rates were low in Clusters II and IV. Principal component analysis (PCA) explained 86.46% of the total variation through nine principal components, of which PC1 accounted for 24.29% of the variability, which was mainly related to grain yield and absorbed photosynthetically active radiation (PAR) and SPAD values. Discriminant Function Analysis (DFA) identified grain fresh weight and grain yield as the most discriminative traits, effectively separating high- and low-yielding genotypes. Mahalanobis D^2 analysis confirmed significant genetic divergence between clusters, especially between clusters IV and VII, indicating potential for hybridization. The classification accuracy from DFA was 96.4%, which verifies the robustness of the clustering results. Representative genotypes were identified for each cluster to guide future breeding strategies. Overall, this study shows that multivariate methods are effective tools for characterizing wheat germplasm and identifying elite genotypes for breeding development. These results provide a strong foundation for future breeding programs aimed at contributing to national food security in Bangladesh under changing climate conditions.

* **Corresponding Author:** Nurunnaher Akter ✉ lata.agr@gmail.com

Introduction

Wheat (*Triticum aestivum* L.) contributes to approximately 50% of global grain trade and around 30% of total grain production worldwide (Shavanov, 2021). As climate change intensifies, the importance of wheat as a staple food continues to grow. To meet the future global food demands, wheat production must rise from 642 million tons to 840 million tons by 2050 (Sharma *et al.*, 2015). In Bangladesh, wheat is the second most important food crop after rice, meeting the population's nutritional requirements (Nawaz *et al.,* 2019). Presently, the country produces 1.17 million metric tons of wheat, with an average yield of 3.7 t ha^{1} over a 316,832 ha area (BBS, 2023). However, this yield remains significantly lower compared to other major wheat-producing countries. Several factors contribute to this yield gap, including biotic and abiotic stresses, suboptimal management practices, limited availability of high-yielding varieties, improper use of fertilizers and seed rates, inadequate water management, and untimely sowing (Timsina *et al.,* 2018).

Despite the steady expansion of wheat acreage in recent years, concerns remain that wheat yields have reached either a plateau level or are declining (Martínez-Moreno *et al.,* 2022). In Bangladesh, although the climatic and edaphic conditions are generally favorable for wheat cultivation, the crop occupies only about 11% of the total area and a mere 4% of the country's total cropped land (Barma *et al.,* 2019). It also contributes just 7% to the nation's overall cereal production. One of the primary reasons for this limited acreage and low productivity is the scarcity of high-yielding wheat varieties (Rashid and Hossain, 2016).

An effort was undertaken from November 2014 to March 2015 at the Department of Agronomy, Gazipur Agricultural University (GAU), to identify superior wheat genotypes. The study evaluated 170 wheat accessions obtained from Advanced Chemical Industries (ACI), Bangladesh (Mustakim, 2016). It laid the groundwork for varietal characterization by applying various multivariate analytical techniques. However, there is a critical need for follow-up research to validate and further refine the selection of promising high-yielding lines. Such efforts are essential for advancing varietal improvement, narrowing the national wheat yield gap, and ensuring food security in Bangladesh in the face of increasing climatic uncertainties.

Recently, multivariate statistical techniques have emerged as powerful tools for screening large sets of genotypes and identifying key traits of interest. Among these, principal component analysis (PCA), discriminant function analysis (DFA), and cluster analysis are widely used for classifying and differentiating genotypes based on complex trait interactions, yielding more precise and meaningful results (Adilova *et al.,* 2020; Hussain *et al.,* 2024). These methods assist plant breeders in uncovering trait variability, elucidating genetic linkages, and prioritizing traits that significantly influence yield and adaptability (Rufati and Manasievska, 2022; Khalid *et al.,* 2023).

PCA and DFA, in particular, offer deeper insights into the characteristics that distinguish high-performing genotypes under specific agro-ecological conditions. In this context, the present study was conducted to evaluate the agronomic performance of 49 wheat genotypes, including the standard Czech variety, using a comprehensive set of morphological, physiological, and yield-related traits. The objective was to identify genetically diverse, high-yielding genotypes possessing desirable characteristics and to classify those effectively using advanced multivariate analytical techniques. The findings of this research are expected to support the development of improved wheat varieties well-suited for widespread cultivation across Bangladesh.

Materials and methods

Experimental site

The experiment was conducted from November 2015 to March 2016 at the research field of the Department of Agronomy, Gazipur Agricultural University (GAU), located in Gazipur, Bangladesh. The site lies within the Madhupur Tract, which falls under Agro-Ecological Zone (AEZ) 28. The experimental area is situated at an elevation of 8.4 meters above sea level, with geographic coordinates of 24°05′ North latitude and 90°16′ East longitude. The study was carried out on upland and clay loam soil, typical of the region.

Planting materials

A total of 49 wheat accessions of the variety BARI Gom 26, developed by Advanced Chemical Industries (ACI), were used in this experiment. Additionally, three check varieties i.e. BARI Gom 25, BARI Gom 28, and BARI Gom 30 released by the Bangladesh Agricultural Research Institute (BARI) were included as controls. A list of genotypes with their identifiers and sources is given in Table 1. The accessions were selected based on their promising yield potential and desirable morphological characteristics, as reported in a previous study by Mustakim (2016). All wheat accessions were obtained from ACI, Bangladesh, and were chosen for their suitability in terms of yield performance, phenological traits, and adaptability to local agro-ecological conditions.

Table 1. Accession numbers and sources of the 49 wheat genotypes evaluated in this study

Genotype no. (G)	Accession identifier	Source
1	BARI Gom 25	BARI
2	2037-1	ACI, Bangladesh
3	456-6	ACI, Bangladesh
4	2220	ACI, Bangladesh
5	431-6	ACI, Bangladesh
6	1138-11	ACI, Bangladesh
7	1107-8	ACI, Bangladesh
8	2580-7	ACI, Bangladesh
9	28-11	ACI, Bangladesh
10	1258-1	ACI, Bangladesh
11	BARI Gom 28	BARI
12	315-2	ACI, Bangladesh
13	1073-23	ACI, Bangladesh
14	28-14	ACI, Bangladesh
15	1770-13	ACI, Bangladesh
16	1932-4	ACI, Bangladesh
17	2182-3	ACI, Bangladesh
18	65-3	ACI, Bangladesh
19	418-3	ACI, Bangladesh
20	429-2	ACI, Bangladesh
21	2462-2	ACI, Bangladesh
22	BARI Gom 30	BARI
23	64-8	ACI, Bangladesh
24	431-1	ACI, Bangladesh
25	2199-8	ACI, Bangladesh
26	432-2	ACI, Bangladesh
27	1271-1	ACI, Bangladesh
28	481-2	ACI, Bangladesh
29	582-34	ACI, Bangladesh
30	28-15	ACI, Bangladesh
31	582-26	ACI, Bangladesh
32	1838-9	ACI, Bangladesh
33	484-1	ACI, Bangladesh
34	1119-11	ACI, Bangladesh
35	1533-2	ACI, Bangladesh
36	2657-1	ACI, Bangladesh
37	627-3	ACI, Bangladesh
38	2019-4	ACI, Bangladesh
39	204-1	ACI, Bangladesh
40	2657-3	ACI, Bangladesh
41	261-1	ACI, Bangladesh
42	2043-11	ACI, Bangladesh
43	28-61	ACI, Bangladesh
44	444-7	ACI, Bangladesh
45	1833-5	ACI, Bangladesh
46	2121-1	ACI, Bangladesh
47	885-3	ACI, Bangladesh
48	1846-4	ACI, Bangladesh
49	432-2	ACI, Bangladesh

Experimental design

The experiment was conducted in a Randomized Complete Block Design (RCBD) with three replications to ensure robust statistical analysis. Each plot was 16.6 meters long and 14.7 meters wide, with a spacing of 20 cm between two lines of the same accession. There was a 40 cm gap between different accessions and a 60 cm drainage gap between replications. This design ensured proper spacing for growth and minimized interference between plots during the experimental period.

Data collection

A comprehensive dataset was collected encompassing a range of morpho-physiological traits, including leaf length, leaf breadth, leaf dry weight, stem dry weight, panicle dry weight, intercepted photosynthetically active radiation (IPAR), leaf temperature, and flag leaf SPAD values at various growth stages. Phenological observations such as days to germination, booting, heading, anthesis, grain filling, and physiological maturity were also recorded. Post-harvest evaluations included yield and its contributing components, such as plant height, number of fertile tillers per square meter, number of grains per spike, grain fresh weight, and thousand-grain weight.

Statistical analysis

Genetic variation among the 49 wheat genotypes was analyzed using SPSS version 24. Initial Pearson's correlation coefficients were calculated to explore relationships among 31 morpho-physiological, phenological, and yield-related traits. To classify the genotypes according to their agronomic performance, non-hierarchical K-means cluster analysis was performed, enabling grouping based on similarities across multiple traits. Principal Component Analysis (PCA) was subsequently applied to reduce data dimensionality and identify the most influential traits contributing to genetic variability, with components having eigenvalues greater than one retained for further interpretation. Discriminant Function Analysis (DFA), including stepwise DFA, was conducted to validate the clustering results and determine which traits most effectively discriminated among genotype groups. The significance and reliability of the DFA models were confirmed using Wilks' lambda, Chi-square tests, and assessments of group mean equality. Additionally, Mahalanobis D^2 distances were calculated to quantify genetic divergence between clusters, aiding in the identification of genetically distant groups suitable for hybridization. Classification accuracy of genotype assignment to clusters was evaluated to ensure the robustness of the multivariate approach.

Results and discussion

Cluster-based multivariate analysis of wheat genotypes

A multivariate approach, specifically k-means clustering, was employed to classify 49 wheat genotypes based on 23 morpho-physiological, phenological, and yield-related traits (Table 2). Such analytical techniques are instrumental in capturing the complex interactions among plant traits, facilitating the identification of genetic variability and enhancing the efficiency of genotype selection in breeding programs (Awan *et al.,* 2015; Ali *et al.,* 2021; Farokhzadeh *et al.,* 2022).

Table 2. Comparison profile of the 7 groups of wheat genotypes classified by k-means clustering

Plant characters	Cluster						
	I	II	III	IV	V	VI	VII
Number of genotypes	7	9	14	6	1	4	8
Morpho-physiological							
Leaf length(cm)	23.06	24.41	25.10	24.99	22.55	23.15	23.75
Leaf breadth(cm)	1.62	1.59	1.66	1.71	1.70	1.55	1.62
Plant components DW (g/plant)							
Leaf	0.41	0.42	0.44	0.38	0.44	0.43	0.44
Stem	0.88	0.73	0.81	0.59	0.95	0.83	0.84
Panicle	2.33	2.31	2.27	2.00	2.73	2.37	2.35
Total	3.62	3.46	3.52	2.97	4.12	3.63	3.63
IPAR (%)	79.64	65.45	69.17	56.16	82.80	81.18	76.22
Leaf temperature at							
Booting	15.58	15.01	15.09	15.47	17.07	14.97	15.31
Heading	16.20	15.64	15.76	16.04	17.76	15.63	15.97
Anthesis	16.84	16.35	16.38	16.76	18.42	16.36	16.68
Flag leaf SPAD value at							
Booting	48.49	45.90	48.51	44.16	49.13	48.10	47.83
Heading	49.70	47.78	50.19	45.61	50.83	50.06	49.41
Anthesis	49.60	47.52	50.33	45.46	50.45	50.09	49.39
1 week after anthesis	44.01	41.82	43.81	40.32	44.62	44.01	44.20
2 weeks after anthesis	39.22	37.15	39.16	35.01	40.08	39.48	38.56
3 weeks after anthesis	29.36	28.50	29.25	27.08	29.50	29.96	29.07
4 weeks after anthesis	20.07	20.57	20.91	19.11	19.45	19.84	23.21
Phenological characters							
Days to germination	6.19	6.67	6.02	6.89	4.00	6.42	5.04
Days to booting	53.38	54.26	54.19	53.50	49.00	55.75	54.13
Days to heading	62.48	63.74	64.00	64.50	59.17	64.58	64.48
Grain filling duration	33.17	32.63	32.90	32.28	32.17	32.33	31.75
Days to physiol. maturity	95.64	96.37	96.90	96.78	91.33	96.92	96.23

Yield and yield attributes							
Plant height (cm)	85.29	77.98	79.82	73.36	86.40	88.66	83.28
Fertile tillers/m2 (no.)	750.60	416.90	501.10	191.10	882.50	847.90	634.30
Spike length (cm)	11.47	12.28	12.01	11.36	11.36	12.31	12.09
Spikelets per spike (no.)	20.50	20.29	20.67	19.21	20.20	20.39	20.21
Grains per spike (no.)	54.57	49.22	53.48	49.06	55.16	48.92	51.52
Grain fresh weight (g m^{-2})	1091.30	539.30	684.80	273.30	1426.6	1269.6	888.30
Initial seed moisture (%)	13.76	14.29	14.56	13.26	13.35	13.02	13.88
1000-grain weight (g)	40.63	41.19	39.65	37.57	42.55	37.93	40.72
Grain yield (g m^{-2})	1069.70	525.56	664.85	269.23	1404.8	1254.8	869.3

Clustering analysis revealed significant intercluster variability, indicating extensive genetic diversity among genotypes. Cluster I contained seven genotypes that were characterized by tall plant height, high grain yield, and the highest number of fertile tillers, although leaf dry weight was low. Cluster II included nine genotypes, which were composed of shorter plants with relatively low yield potential, yet these genotypes exhibited long spikes, indicating a closed exchange between yield components. Cluster III contained 14 genotypes that were highest in terms of leaf length and dry weight, indicating vigorous plant growth, but were limited by low 1,000-grain weight and less fertile tillers, which could compromise grain yield. Cluster IV included six genotypes with the worst overall performance, including the lowest plant height, biomass accumulation, and grain yield. In contrast, cluster V (single genotype) emerged as the most promising, exhibiting superior performance in multiple traits including highest total dry weight, highest grain yield, and early physiological maturity – highly desirable traits in resource-limited and stress-prone environments. Four genotypes in cluster VI also performed well, including the tall genotype with high yield and biomass, although its spikes had lighter grains and fewer grains, which may affect grain quality. Cluster VII included eight genotypes with early grain-filling that have average yield performance but favorable dry matter partitioning, making them potentially suitable for environments with short growing seasons or terminal stress.

Overall, Clusters V and VI represent genotypes with high agronomic potential and are valuable candidates for inclusion in wheat improvement programs. In contrast, Clusters II and IV demonstrated limited productivity and may be less suitable for breeding objectives. These results underscore the significance of integrated trait-based selection, particularly plant height, spike morphology, dry matter allocation, and yield components, in enhancing wheat productivity and guiding future breeding strategies (Beral *et al.*, 2020; Kaleri *et al.*, 2023).

Principal component analysis (PCA)

Principal Component Analysis (PCA) is an effective multivariate statistical technique used to reduce dimensionality and identify the most influential traits contributing to genetic variation among genotypes (Gewers *et al.*, 2021). It reveals patterns of association among traits, thereby facilitating the selection of superior genotypes for breeding programs (Nayana *et al.*, 2022; Weraikat *et al.*, 2024). In the present study, PCA was conducted on 49 wheat genotypes across 31 morpho-physiological, phenological, and yield-related traits. The analysis generated 31 principal components (PCs) through linear transformation, each associated with an eigenvalue that indicates the proportion of total variance it explains, and a latent vector representing the contribution of each original variable (Table 3). Of these, the first nine PCs had eigenvalues greater than 1 and collectively accounted for approximately 86.46% of the total genetic variation. The individual contributions of these PCs were: PC1 (24.29%), PC2 (17.88%), PC3 (12.74%), PC4 (8.81%), PC5 (5.35%), PC6 (4.92%), PC7 (4.41%), PC8 (4.20%), and PC9 (3.86%).

The latent vectors corresponding to the first three principal components (PCs) are presented in Table 4. PC1 explained the largest proportion of variability (24.29%) and was predominantly associated with yield-contributing traits such as grain yield, grain fresh weight, fertile tillers per square meter, intercepted photosynthetically active radiation (IPAR), and SPAD values at various reproductive stages. Other positively associated traits included spikelets per spike, spike length, plant height, and total dry biomass. In contrast, traits such as leaf length, leaf breadth, panicle dry weight, and leaf temperatures at booting, heading, and anthesis contributed negatively, along with days to germination, heading, and physiological maturity. This indicates that PC1 captures variation related to photosynthetic efficiency and yield potential. PC2, accounting for 17.88% of the variation, was largely influenced by phenological traits such as days to booting, heading, and physiological maturity, as well as leaf size (length and breadth). Positive associations were also observed with SPAD values during early reproductive stages. However, traits like high leaf temperatures at booting and heading stages and SPAD values during later grain filling stages had negative loadings, suggesting PC2 captures variation related to maturity and leaf physiology. PC3, which explained 12.74% of the total variation, was driven by biomass-related traits such as stem dry weight, panicle dry weight, total dry weight, plant height, grain fresh weight, and grain yield. Conversely, early-stage traits including SPAD values at booting, heading, and anthesis, as well as days to heading and leaf dimensions, had negative contributions, indicating that PC3 represents variation associated with post-anthesis biomass accumulation and grain development.

Table 3. Initial and extracted Eigenvalues and % of variation in respect of 23 plant characters of 49 wheat genotypes

Principal component	Initial Eigen values			Extraction sums of squared loadings		
	Total	% of variation	Cumulative %	Total	% of variation	Cumulative %
1	7.53	24.29	24.29	7.53	24.29	24.29
2	5.54	17.88	42.17	5.54	17.88	42.17
3	3.95	12.74	54.91	3.94	12.74	54.91
4	2.73	8.81	63.72			
5	1.66	5.35	69.07			
6	1.53	4.92	73.99			
7	1.37	4.41	78.40			
8	1.30	4.20	82.61			
9	1.19	3.86	86.46			

Table 4. Latent vectors associated with the first two principal components

Pant character	Principal components		
	1st	2nd	3rd
Leaf length (cm)	-0.165	0.457	-0.084
Leaf breadth (cm)	-0.150	0.440	0.043
Leaf dry weight (g plant^{-1})	0.143	0.219	0.656
Stem dry weight (g plant^{-1})	0.227	-0.075	0.863
Panicle dry weight (g plant^{-1})	-0.060	-0.145	0.823
Total dry weight (g plant^{-1})	0.049	-0.092	0.919
IPAR (%)	0.657	-0.409	0.435
Leaf temperature at booting (^{o}C)	-0.194	-0.817	0.192
Leaf temperature at heading (^{o}C)	-0.160	-0.824	0.214
Leaf temperature at anthesis (^{o}C)	-0.160	-0.816	0.233
SPAD value at booting	0.847	-0.002	-0.156
SPAD value at heading	0.894	0.049	-0.077
SPAD value at anthesis	0.907	0.022	-0.090
SPAD value at 1 week after anthesis	0.918	-0.006	0.007
SPAD value at 2 weeks after anthesis	0.922	-0.096	0.055
SPAD value at 3 weeks after anthesis	0.807	-0.093	0.098
SPAD value at 4 weeks after anthesis	0.170	-0.481	0.012
Days to germination	-0.323	0.336	-0.305
Days to booting	0.156	0.775	0.129
Days to heading	-0.096	0.783	-0.040
Days to grain filling period	0.922	-0.096	0.055

Days to physiological maturity	-0.183	0.804	0.081
Plant height (cm)	0.188	-0.173	0.775
Fertile tillers m^{-2} (no.)	0.677	-0.418	0.484
Spike length (cm)	0.209	0.045	0.128
Spikelets per spike (no.)	0.334	-0.159	0.234
Grains per spike (no.)	0.285	-0.274	0.089
Grain fresh weight (g m^{-2})	0.618	-0.395	0.557
Initial seed moisture (%)	0.161	-0.016	0.104
Thousand-grain weight (g)	0.053	-0.015	0.468
Grain yield (g m^{-2})	0.612	-0.395	0.555

Overall, the PCA results effectively identified key traits contributing to genetic variability among wheat genotypes. The first three principal components, explaining more than 54% of the total variance, revealed strong associations between yield and physiological attributes, including SPAD values, radiation interception, and biomass production. These findings provide valuable insights for selecting high-performing genotypes and align with earlier studies by Poudel *et al.* (2017) and Zewdu *et al.* (2024), who demonstrated the utility of PCA in understanding genetic diversity in wheat.

Discriminant function analysis (DFA)

Discriminant Function Analysis (DFA) is a powerful multivariate technique used to identify the traits that most effectively differentiate among predefined groups (Abdolshahi *et al.,* 2015). In plant breeding, DFA aids in grouping genetically similar genotypes based on shared trait expressions, thereby enhancing the precision and efficiency of selection processes (Mustakim, 2016). In the present study, a stepwise DFA was conducted to evaluate the discriminatory power of 23 morpho-physiological, phenological, and yield-related traits in separating 49 wheat genotypes previously classified into seven clusters via k-means clustering. The objective was not only to assess the traits' ability to distinguish among clusters but also to provide a visual representation of the relationships among genotypes in multivariate space.

The analysis identified two significant discriminant functions that collectively accounted for 100% of the total variance among clusters (Table 5). The first function alone explained 98.1% of the total variation, while the second function accounted for the remaining 1.9%. Both functions were statistically significant at $p < 0.001$, as confirmed by Wilks' lambda and chi-square tests, indicating their robust capacity to discriminate among the genotype clusters. These results demonstrate that a small number of trait combinations can effectively capture the underlying genetic variability, offering a valuable tool for genotype selection and classification in wheat breeding programs.

Table 5. Discriminant functions that differentiate between clusters of 49 wheat genotypes

Function	Latent root	Variance %		r^2 coefficient	Walk's λ	χ^2	df	*p*
		Function	Cumulative					
1	32.717^a	98.1	98.1	0.985	0.018	174.741	12	.000
2	0.647^a	1.9	100.0	0.627	0.607	21.707	5	.001

^{a} First two canonical discriminant functions were used in the analysis.

Table 6. Standardized canonical discriminant function coefficients of the plant characters that mostly contributed in the grouping of 49 genotypes of wheat

Discriminating variables	Discriminant function	
	1	2
Grain yield (g m^{-2})	-0.445	10.385
Grain fresh weight (g m^{-2})	1.442	-10.294

Table 6 presents the variables that most significantly influenced the discriminant functions along with their respective coefficients. The analysis reveals that, among the 49 genotypes evaluated, grain yield (g m^{-2}) and grain fresh weight (g m^{-2}) were the primary contributing traits. Together, these two traits accounted for nearly 100% of the total variance explained by Functions 1 and 2. Grain fresh weight showed a higher coefficient (1.442) in Function 1, indicating that it was the dominant contributor, explaining 98.1% of the variance in this function. In contrast, grain yield had a greater influence in Function 2, with a coefficient of 10.385 g m^{-2}, accounting for the remaining 1.9% of the variance.

Table 7. The structure matrix represents the correlation between 31 discriminating variables and the standardized canonical discriminant functions of 49 genotypes of wheat

Discriminating variables	Discriminant function	
	1	2
Grain fresh weight	0.999*	0.043
Grain yield	0.990*	0.139
Fertile tiller	0.523*	0.239
IPAR(%)	0.435*	0.243
Plant height	0.364*	-0.081
Days to heading	-0.272*	-0.026
Leaf temperature at booting	0.238*	0.077
Days to germination	-0.237*	-0.009
Days to physiological maturity	-0.232*	0.105
Days to booting	-0.223*	0.128
Leaf temp. at heading	0.223*	0.076
Flag leaf SPAD value 4 weeks after anthesis	0.220*	0.074
Leaf temp. at anthesis	0.204*	0.046
Flag leaf SPAD value 3 weeks after anthesis	0.122*	-0.093
Total dry weight	0.101*	-0.037
Panicle dry weight	0.088*	0.043
Number of grains per spike	0.079*	-0.037
Thousand-grain weight	0.071*	-0.065
Initial seed moisture	-0.064	-0.849*
Spikelets per spike	-0.235	-0.287*
Days to grain filling period	0.065	0.241*
Leaf length	-0.067	0.232*
Leaf dry weight	0.125	-0.228*
Spike length	-0.196	-0.207*
Leaf breadth	-0.094	0.203*
Flag leaf SPAD value 2 weeks after anthesis	0.072	-0.166*
Stem dry weight	0.073	-0.121*
Flag leaf SPAD value 1 week after anthesis	0.064	-0.113*
Flag leaf SPAD value at booting	-0.031	-0.111*
Flag leaf SPAD value at heading	0.054	-0.111*
Flag leaf SPAD value at anthesis	0.011	-0.089*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation within function.

* Largest absolute correlation between each variable and any discriminant function.

Table 7 presents the correlation coefficients between the two discriminant functions and thirty-one discriminating traits. The results indicate that grain fresh weight (g m^{-2}) had the strongest association with Function 1, exhibiting a correlation coefficient of 0.99, making it the most influential variable in distinguishing among the 49 genotypes. Stepwise Discriminant Function Analysis (DFA) further confirmed that grain fresh weight was the most significant factor in explaining the variation across genotypes.

According to the X-axis (Function 1) in Fig. 1, genotypes located on the left produced the lowest grain yield per square meter, while those on the right yielded the highest. Thus, Function 1 effectively separated clusters V and VI (high-yielding genotypes) from clusters II and IV (low-yielding genotypes) based on grain yield. On the other hand, the Y-axis (Function 2) revealed that genotypes positioned in the upper part of the diagram exhibited delayed germination, whereas those in the lower part germinated earlier. Consequently, Function 2 clearly differentiated clusters II, IV, and VI from clusters I, III, V, and VII based on days to germination. The latter group showed earlier germination compared to the former. Additional traits such as number of fertile tillers, intercepted photosynthetically active radiation (IPAR %), plant height, and SPAD readings also contributed to genotype differentiation.

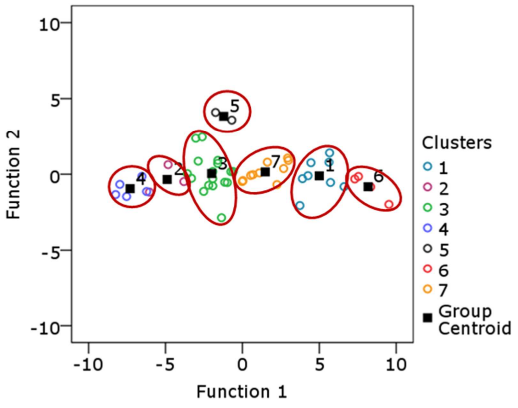

Fig. 1. Graphical presentation of the discriminant function analysis of seven groups of 49 wheat genotypes. The encircled accessions show the groups (clusters) obtained through cluster analysis

Overall, the DFA validated the cluster groupings and identified grain fresh weight and grain yield as the most critical traits for distinguishing among wheat genotypes. These findings provide valuable insights for enhancing wheat breeding and selection strategies (Tshikunde *et al.*, 2019; Chachar *et al.*, 2024).

Mahalanobis distance

The Mahalanobis D^2 analysis revealed highly significant differences ($p < 0.001$) among the seven clusters, confirming the presence of substantial genetic diversity (Table 8). The greatest genetic divergence was observed between Cluster VII and Cluster IV ($D^2 = 967.13$), followed by Cluster VI and Cluster III ($D^2 = 900.35$), and Cluster II and Cluster I ($D^2 = 844.12$). These large distances indicate distinct genetic backgrounds among these clusters, making them promising candidates for parental selection in wheat breeding programs. Selecting genotypes from genetically distant clusters may enhance the potential for developing high-yielding hybrid wheat varieties (Adilova *et al.*, 2020; Mohammadi *et al.*, 2023).

Table 8. Pairwise Mahalanobis distance (D^2) between seven clusters of wheat genotypes

Cluster	I	II	III	IV	V	VI	VII
I	0.00						
II	844.12^{a}	0.00					
III	625.78^{a}	218.48^{a}	0.00				
IV	127.30^{a}	433.05^{a}	649.77^{a}	0.00			
V	492.14^{a}	133.33^{a}	111.50^{a}	176.03^{a}	0.00		
VI	274.93^{a}	111.90^{a}	900.35^{a}	154.80^{a}	220.34^{a}	0.00	
VII	308.10^{a}	536.17^{a}	317.93^{a}	967.13^{a}	798.92^{a}	582.82^{a}	0.00

^{a} Distances differing from zero at a 99% confidence interval.

Classification accuracy

The classification results from Discriminant Function Analysis (Table 9) demonstrated a high level of precision. More than 75% of the genotypes were correctly assigned to their respective clusters, with an overall classification accuracy of 96.4%. Clusters I, II, III, IV, V, and VII showed perfect classification accuracy (100%), while Cluster VI had 75% accuracy, with one genotype misclassified into Cluster V. This level of accuracy is considered highly reliable in biological research.

Representative genotypes

Fig. 1 illustrates the spatial distribution of genotypes across the seven clusters based on their cumulative responses to Functions 1 and 2. Each cluster's group centroid represents the average position of its genotypes, reflecting their collective response to the discriminating traits. Genotypes closest to their cluster centroid can be considered the most representative of that group, though not necessarily the highest performing. Accordingly, the following genotypes may be regarded as the most indicative representatives of their respective clusters (as listed in Table 10): Group I: G1 (BARI Gom 25), Group II: G35 (1533-2), Group III: G23 (64-8), Group IV: G32 (1838-9), Group V: G11 (BARI Gom 28), Group VI: G37 (627-3), Group VII: G41 (261-1). These representative genotypes reflect the typical trait profiles of their groups and can serve as benchmarks for future selection and breeding efforts.

Table 9. Classification matrix (at the precision level) for seven clusters of wheat genotypes

Group	% correct	I	II	III	IV	V	VI	VII	Total no. observed
I	100	7	0	0	0	0	0	0	7
II	100	0	9	0	0	0	0	0	9
III	100	0	0	14	0	0	0	0	14
IV	100	0	0	0	6	0	0	0	6
V	100	0	0	0	0	1	0	0	1
VI	75	0	0	0	0	1	3	0	4
VII	100	0	0	0	0	0	0	8	8
Total no. predicted	96.4	7	9	14	6	2	3	8	49

Rows represent the actual observed categories, and columns correspond to the predicted categories.

Table 10. Major characteristics of seven wheat genotypes that mostly represent their respective groups

Cluster	Genotypes	Acc. no.	Grain yield (g m^{-2})	No. of fertile tillers m^{-2}	Plant height (cm)	1000-grain weight (g)	Days to physiological maturity	Days to grain filling period
I	G1	BARI Gom 25	1123.35	819.2	85.86	44.26	91.5	32.2
II	G35	1533-2	489.61	355.0	76.23	42.93	96.3	32.0
III	G23	64-8	613.10	463.0	71.70	41.40	97.0	34.0
IV	G32	1838-9	192.88	126.7	81.23	42.57	97.3	32.6
V	G11	BARI Gom 28	1404.76	882.5	86.40	42.55	91.3	32.2
VI	G37	627-3	1317.04	875.0	84.73	41.67	96.0	33.0
VII	G41	261-1	782.61	561.7	85.13	48.77	96.3	30.7

Conclusion

This study identified significant genetic variability among 49 wheat genotypes using multivariate analytical techniques including cluster analysis, principal component analysis (PCA), and discriminant function analysis (DFA). Genotypes grouped in clusters V and VI exhibited superior agronomic traits such as high grain yield, favorable dry matter partitioning, early physiological maturity, and robust biomass accumulation. PCA revealed that traits like grain yield, fertile tillers, intercepted PAR, and SPAD values were the most influential in determining genotype performance. DFA validated the clustering pattern, identifying grain fresh weight and grain yield as key discriminative traits. The observed genetic divergence, especially between Clusters IV and VII, highlights the potential for selecting diverse parental lines in breeding programs. These results provide a broad basis for future wheat development strategies in Bangladesh that aim to reduce the national yield gap and increase resilience to climate stress conditions.

Recommendations

Based on the results of this study, genotypes in clusters V and VI should be prioritized as parental lines in breeding programs due to their superior performance in many agronomic traits. Breeders should use a trait-based selection approach that prioritizes key yield-affecting traits such as grain fresh weight, grain yield, intercepted photosynthetically active radiation (PAR), and SPAD values. High Mahalanobis D² values were noted among genetically distant clusters, particularly clusters IV and VII, suggesting the potential for developing heterotic hybrids through deliberate crossbreeding. Multi-location experiments across different agro-ecological regions are needed to establish the adaptation and performance of these genotypes. Furthermore, the elite genotypes identified through multivariate analysis should be integrated into national variety release pipeline to assist in ongoing efforts to increase wheat yields and ensure food security.

References

Abdolshahi R, Nazari M, Safarian A, Sadathossini TS, Salarpour M, Amiri H. 2015. Integrated selection criteria for drought tolerance in wheat (*Triticum aestivum* L.) breeding programs using discriminant analysis. Field Crops Research **174,** 20–29. https://doi.org/10.1016/j.fcr.2015.01.009

Adilova SS, Qulmamatova DE, Baboev SK, Bozorov TA, Morgunov AI. 2020. Multivariate cluster and principal component analyses of selected yield traits in Uzbek bread wheat cultivars. American Journal of Plant Sciences **11**(06), 903. https://doi.org/10.4236/ajps.2020.116066

Ali N, Hussain I, Ali S, Khan NU, Hussain I. 2021. Multivariate analysis for various quantitative traits in wheat advanced lines. Saudi Journal of Biological Sciences **28**(1), 347–352. https://doi.org/10.1016/j.sjbs.2020.10.011

Ali Y, Khan MA, Hussain M, Atiq M, Ahmad JN. 2019. An assessment of the genetic diversity in selected wheat lines using molecular markers and PCA-based cluster analysis. Applied Ecology & Environmental Research **17**(1), 931-950. http://dx.doi.org/10.15666/aeer/1701_931950

Awan SI, Ahmad SD, Ali MA, Ahmed MS, Rao A. 2015. Use of multivariate analysis in determining characteristics for grain yield selection in wheat. Sarhad Journal of Agriculture **31**(2), 139–150. http://dx.doi.org/10.17582/journal.sja/2015/31.2.139.150

Bangladesh Bureau of Statistics (BBS). 2023. Estimates of wheat, 2022-23. Government of the People’s Republic of Bangladesh, Bangladesh Bureau of Statistics, Agriculture Wing, Agargaon, Dhaka 1207.

Barma NCD, Hossain A, Hakim MA, Mottaleb KA, Alam MA, Reza MAA, Rohman MM. 2019. Progress and Challenges of Wheat Production in the Era of Climate Change: A Bangladesh Perspective. In: Hasanuzzaman M, Nahar K, Hossain M, Eds. Wheat Production in Changing Environments. Springer, Singapore. https://doi.org/10.1007/978-981-13-6883-7_24

Beral A, Rincent R, Le Gouis J, Girousse C, Allard V. 2020. Wheat individual grain-size variance originates from crop development and from specific genetic determinism. PLoS One **15**(3), e0230689. https://doi.org/10.1371/journal.pone.0230689

Chachar Z, Fan L, Chachar S, Ahmed N, Narejo MUN, Ahmed N, Lai R, Qi Y. 2024. Genetic and Genomic Pathways to Improved Wheat (*Triticum aestivum* L.) Yields: A Review. Agronomy **14**(6), 1201. https://doi.org/10.3390/agronomy14061201

Farokhzadeh S, Hassani HS, Zinati Z, Rajaei M. 2022. Evaluation of triticale lines compared to wheat cultivars using supervised learning. Philippine Agricultural Scientists **105**(4), 369–389.

Gewers FL, Ferreira GR, Arruda HFD, Silva FN, Comin CH, Amancio DR, Costa LDF. 2021. Principal component analysis: A natural approach to data exploration. ACM Computing Surveys **54**(4), 1–34. https://doi.org/10.1145/3447755

Hussain MA, Hameed MU, Ahmad N. 2024. Principal component analysis in *Triticum aestivum* under field conditions for food security. Journal of Biological and Agricultural Advancements **2**(1), 11-21.

Kaleri GS, Memon S, Kaleri AA, Magsi MA, Kumar V, Dhiloo AJ, Awan MH, Banbhan GM, Kaleri IA, Kalhoro MH, Kaleri MK, Soomro G, Kakar N, Shar AR. 2023. Genetic Diversity Studies in Durum Wheat. Journal of Arable Crops and Marketing **5**(2), 49–54. https://doi.org/10.33687/jacm.005.02.4800

Khalid A, Hameed A, Tahir MF. 2023. Estimation of genetic divergence in wheat genotypes. Cereal Research Communications **51**(1), 217–224. https://doi.org/10.1007/s42976-022-00287-w

Martínez-Moreno F, Ammar K, Solís I. 2022. Global changes in cultivated area and breeding activities of durum wheat. Agronomy **12**(5), 1135. https://doi.org/10.3390/agronomy12051135

Mohammadi R, Jafarzadeh J, Armion M, Hatamzadeh H, Roohi E. 2023. Clustering stability methods. Euphytica **219**(10), 109. https://doi.org/10.1007/s10681-023-03237-7

Mohi-Ud-Din M, Hossain MA, Rohman MM, Uddin MN, Haque MS, Ahmed JU, Hossain A, Hassan MM, Mostofa MG. 2021. Multivariate analysis of drought tolerance. Plants **10**(5), 879. https://doi.org/10.3390/plants10050879

Mustakim AAMM. 2016. Morphological and agronomical characterization of wheat accessions. MS Thesis, Department of Agronomy, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Salna, Gazipur-1706.

Nawaz A, Farooq M, Nadeem F, Siddique KH, Lal R. 2019. Rice–wheat cropping systems in South Asia. Crop and Pasture Science **70**(5), 395–427. https://doi.org/10.1071/CP18383

Nayana BM, Kumar KR, Chesneau C. 2022. Wheat yield prediction using PCA-MARS. AgriEngineering **4**(2), 461–474. https://doi.org/10.3390/agriengineering4020030

Poudel A, Thapa DB, Sapkota M. 2017. Assessment of genetic diversity of bread wheat (*Triticum aestivum* L.) genotypes through cluster and principal component analysis. International Journal of Experimental Research and Review **11**(1), 1-9.

Rashid MA, Hossain TM. 2016. Adoption of wheat varieties in Bangladesh. Bangladesh Journal of Agricultural Research **41**(3), 491–505. https://doi.org/10.3329/bjar.v41i3.29721

Rufati LQ, Manasievska S. 2022. Principal component and cluster analysis for wheat species. Journal of Agricultural, Food and Environmental Sciences **76**(2), 47–54.

Sharma I, Tyagi BS, Singh G, Venkatesh K, Gupta OP. 2015. Enhancing wheat production – A global perspective. The Indian Journal of Agricultural Sciences **85**(1), 03–13. https://doi.org/10.56093/ijas.v85i1.45935

Shavanov MV. 2021. The role of food crops in Poaceae and Fabaceae families. IOP Conference Series: Earth and Environmental Science **624**(1), 012111. IOP Publishing. https://doi.org/10.1088/1755-1315/624/1/012111

Timsina J, Wolf J, Guilpart N, Van Bussel LGJ, Grassini P, van Wart J, Hossain A, Rashid H, Islam S, van Ittersum MK. 2018. Can Bangladesh produce enough cereals? Agricultural Systems **163**, 36–44. https://doi.org/10.1016/j.agsy.2016.11.003

Tshikunde NM, Mashilo J, Shimelis H, Odindo A. 2019. Agronomic traits affecting wheat yield. Frontiers in Plant Science **10**, 1428. https://doi.org/10.3389/fpls.2019.01428

Weraikat D, Šorič K, Žagar M, Sokač M. 2024. Data analytics in agriculture for crop yield optimization. Sustainability **16**(17), 7331. https://doi.org/10.3390/su16177331

Zewdu D, eMekonnen F, Geleta N. 2024. Cluster and principal component analysis for yield and yield related traits of bread wheat (*Triticum aestivum* L.) genotypes. Agricultural and Biological Research **40**(2), 962–967. https://doi.org/10.35248/0970-1907.24.40.962-967