RESEARCH ARTICLE

An integrative multivariate and clustering analysis dissecting genetic variability in bread wheat (*Triticum aestivum* L.)

Anamish Tyagi[1], Subhash Chandra Vimal[2], Ashwini Kumar[3], Ruchi Singh[4], Sharad Singh[1], Vinay Kumar[1], Poonam Sharma[5], Padmini Ganga[1], Sonali Srivastava[1], Nitin Kumar[6] & Aman Srivastava[1*]

[1]Department of Genetics and Plant Breeding, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India

[2]Department of Seed Science and Technology, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India

[3]Department of Molecular Biology & Biotechnology, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India

[4]Department of Genetics and Plant Breeding, Chandra Shekhar Azad University of Agriculture & Technology, Nawabganj, Kanpur 208 002, Uttar Pradesh, India

[5]Department of Genetics and Plant Breeding, Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur 176 061, Himachal Pradesh, India

[6]Department of Ag. Botany, Ch. Shivanath Singh Shandilya PG College (C C S University), Machhra, Meerut 250106, Uttar Pradesh, India

*Correspondence email - amansrivastava3542@gmail.com

Abstract

A detailed investigation was conducted to assess analysis of variance, genetic divergence, correlation and principal component analysis (PCA) among 381 wheat genotypes from the ICAR-NBPGR Core Collection with three local cultivated varieties. We analyzed 12 agro-morphological traits, yielding clear and significant results. The analysis of variance clearly shows that the treatments have a significant effect on almost all traits. High Genotypic Coefficient of variation observed for yield-related traits and high heritability observed among spike length, days to 50 % flowering and plant height. For grain yield and the number of grains per spike, a positive correlation was observed. The PCA results indicated that the first and second principal components (PC-1 and PC-2) have eigenvalues greater than one, accounting for an impressive 50.73 % of the total variability. Notably, for PC-2 exhibited strong positive loadings for both 1000-grain weight and grain yield. The cluster analysis distinctly categorized the genotypes into eight separate clusters using Euclidean distance, identifying the greatest genetic distance between Cluster V and Cluster III.

Keywords: principal component analysis; cluster analysis; bread wheat; grain yield

Introduction

Bread wheat (*Triticum aestivum* L.) in terms of global production is the second important cereal. In India, wheat is cultivated over 31.76 million hectares, producing 109.52 million tons (1). Wheat is a self-pollinating crop that is noted for its relatively high content of niacin and thiamine, essential for producing gluten, a special protein. In India, wheat is a primary food source, featuring in various forms such as chapattis, cakes and bread. The history of wheat cultivation in India spans over 5000 years and is deeply embedded in the nation's agricultural legacy. On a global scale, this vital grain has sustained civilisations for more than 10000 years, shaping diets and driving economic growth. By 2025, global wheat production is projected to reach 796 million tons, reflecting a modest year-on-year increase of nearly 1 %. India is expected to achieve a record wheat production of 115.4 million tons in the 2024-25 crop year, representing a crucial 2 % increase over the previous year (1).

The Economic Survey for 2024-25 highlights the vital role of wheat in the agriculture sector, contributing approximately 16 % of India's overall GDP. As the global population approaches ten billion by 2050 (2), it is imperative to enhance wheat production to meet future demands significantly. Wheat is a cornerstone of food security, supplying an impressive 55 % of carbohydrates and 21 % of food calories worldwide, outperforming all other grains in both production and area (3).

Enhancing germplasm and genetic variability is crucial for the consistent and sustainable production of food crops. To effectively estimate and utilize germplasm, measuring the extent of existing genetic divergence is of utmost importance (4). Correlation analysis reveals the interactions among different traits, equipping breeders with a precise roadmap for selecting combinations that can boost yield. Path analysis helps untangle these relationships, identifying both direct and indirect effects on overall plant yield (5). Given the complex interactions of traits

in crops like wheat, a robust multivariate approach is essential for managing multiple traits simultaneously and revealing hidden complexities among them (6). This comprehensive strategy enhances our understanding and strengthens our ability to make informed, impactful breeding decisions.

Multivariate statistics are crucial for classifying germplasm and analysing genetic relationships between breeding materials. The importance of PCA and hierarchical cluster analysis (HCA) proved instrumental in uncovering key insights. These techniques are essential for gaining deep insights into genetic diversity and the trait associations among the evaluated genotypes. PCA is a powerful tool for dimensionality reduction, enabling effective visualisation of trait variability across genotypes. By identifying the principal components that drive trait variation, PCA helps prioritise traits critical for selection in breeding programs, highlighting genotypes with desirable trait combinations (7).

Additionally, the Genotype by Trait (GT) biplot is an exceptionally effective biometrical technique for data resolution (8). This graphical method illustrates the performance capabilities of genotypes, facilitating the identification of traits suitable for indirect selection (9). The GT biplot is widely used for aligning accessions, providing a comprehensive overview of the interrelationships among agro-morphological traits in various crops (10).

Addressing genetic divergence and enhancing germplasm are foundational for ensuring consistent and sustainable production of food crops. A thorough understanding of the available genetic variability is vital for maximizing germplasm potential and guiding the selection of ideal parents to create superior varieties with exceptional nutritional quality (11). The greater the genetic diversity among parent plants, the more pronounced the heterotic expression in the first filial generation and the wider the array of traits within the segregating population. Research into genetic divergence is essential for identifying optimal genotypes for innovative breeding strategies aimed at cultivating the next generation of crops.

The primary objective of any breeding program is to produce high yielding that can adapt to diverse conditions. These goals can only be achieved in the presence of sufficient genetic diversity and variability. When the patterns of genetic variation in a community are recognised, the efficiency of genetic gain through random selection can be enhanced.

Materials and Methods

Seed material

A diverse set of 381 wheat (*T. aestivum* L.) accessions, consisting of landraces, varieties, advanced breeding lines, elite lines, germplasm stocks and cultivars, obtained from ICAR-NBPGR Core, were used for this study (Supplementary Table 1) along with three commercial cultivated varieties, viz. PBW757, DBW187 and RAJ3777. The set was evaluated using an augmented design (10 × 40) over two consecutive years (2023-24 and 2024-25) at Acharya Narendra Deva University of Agriculture and Technology, Ayodhya, Uttar Pradesh, India, located at longitude 82°12' E and latitude 26°47' N during the *Rabi* seasons, with cultivation practices carried out throughout the entire growth period.

Phenotypic estimation of quantitative traits

The present study involved the systematic collection and analysis of key agronomic traits to assess crop performance. The following parameters were measured: Days to 50 % Flowering, which reflects the duration required for half of the plant population to achieve flowering; Days to Maturity, representing the time from planting until the crop reaches full maturity and is ready for harvest; and tiller number per plant, quantifying the number of tillers produced by each plant, an important aspect regarding yield potential. Additionally, plant height (cm) was measured from the soil surface to the apex of the spike, excluding any awns, while spike length (cm) was determined from the base of the spike to the tip of the terminal spikelet, also excluding awns. The number of spikelets per spike was counted to further characterize the spike's morphology. Biological Yield per Plot (g) was calculated to assess the total dry biomass produced, incorporating all plant constituents, including leaves, stems, roots and grains. Moreover, Grains per Spike were counted for each spikelet and grain yield (g) was determined to measure the total weight of grain produced over a specified area. Finally, the 1000 Grain Weight (g) was established by weighing a sample of 1000 kernels to evaluate kernel size and weight uniformity. These quantitative measures provide comprehensive insights into the crop's growth dynamics and yield potential, serving as a basis for further research and improvement of agronomic practices. Harvest index is the ratio that expresses the proportion of a plant's total biological yield that is harvested as grain or economic yield.

It is calculated using the formula:

$$\text{Grain yield} / \text{Biological yield} * 100.$$

Statistical analysis

Statistical analyses were conducted using the mean values collected from five randomly chosen plants. ANOVA analysis conducted with the R package Augmented Randomised Complete Block Design (RCBD) (12). To calculate correlation coefficients, the Corr plot package was used (13). We also performed Principal Component Analysis (PCA) and assessed genetic differences using functions from the stats package (14) and the FactoMineR package (15). To visualize and understand the genetic relationships among different genotypes, we conducted Hierarchical Cluster Analysis (HCA) using the cluster package (16), along with factoextra (17) and dendextend (18) packages.

Results

ANOVA, phenotypic variation and descriptive statistics in wheat genotypes

The ANOVA reveals that treatments have a significant impact on nearly all evaluated traits, as shown in Table 1. When comparing the Test and Check treatments, we observe significant differences, with one value at 762.96 days to 50 % flowering. However, the biological yield in the Test treatment was not significant. Block effects don't matter much and the low error variation suggests that the results are reliable.

In this study, we observed significant phenotypic variation among the examined wheat genotypes for all assessed traits. The time taken to reach 50 % flowering varied from about 54.3 to 77.3 days. Similarly, the duration until maturity ranged

Table 1. Block adjusted ANOVA (All traits)

Source of Variation	DF	DY50	DM	GFD	TN	PH	SL	SPP	GPS	GY	BY	HI	TGW
Treatment (ignoring Blocks)	383	23.46 **	49.17 *	26.02*	10.59*	124.78 **	1.82 **	5.18 **	47.18 **	3281.85 **	17932.23 *	47.82 **	31.44 **
Treatment: Check	2	2.63*	114.53 *	134.43 *	15.29*	17.86	2.69 **	16.04 **	150.03 **	3348.91*	31062.35 *	542.05 **	17.80 **
Treatment: Test	380	21.62 **	48.63 *	24.79*	10.54 *	125.11 **	1.80 **	5.07 **	45.95 **	3148.09 **	14718.26 ns	41.19 **	30.68 *
Treatment: Test vs. Check	1	762.96 **	120.29 *	277.35 **	23.18 *	214.23 *	10.55 **	24.90 **	307.64 **	53976.20 **	1212977.32 **	1580.08 **	348.14 **
Block (eliminating Treatments)	9	3.99 ns	19.61 ns	20.76 ns	20.84 ns	15.76 ns	1.07 ns	2.29 ns	22.26 ns	345.05 ns	14729.90 ns	45.26 ns	5.84 ns
Residuals	18	2.37	20.13	24.32	5.49	26.06	0.25	1.81	15.59	1038.84	8068.03	13.66	11.84

ns: Not significant (P > 0.05)*****: Significant (P ≤ 0.05)******: Highly significant (P ≤ 0.01)

DY50: Number of days to 50 % flowering, DTM: Number of days to crop maturity, GFD: Grain filling duration (days), TN: Tiller number per plant, PH: Plant height (cm), SL: Length of the spike (cm), SPS: Number of spikelets/spike, GPS: Number of grains/spike, , GY: grain yield/plot, BY: Biological yield/plot (BY), HI: Harvest index (%), TGW: 1000-grains weight (g)

from 93.4 to 127.1 days. Wide variation was observed in grain yield per plant, which ranged from 51.6 to 316.97 g, averaging 158.67 g. Biological yield showed even broader variability, spanning from 0 to 764.9 g, with an average of 365.58 g. Standard deviations for the yield traits were particularly high, recorded at approximately 55.6 g for grain yield and 144.8 g for biological yield. These variabilities are summarised in Table 2.

Genotypic and phenotypic variability

All the traits showed higher phenotypic coefficient of variation compared to genotypic coefficient of variation. This indicates that environmental factors had a strong influence on how these traits appeared. Some traits showed high GCV values (above 20 %), including tiller number (23.0 %), grain yield (28.9 %) and biological yield (22.3 %), suggesting considerable genetic variability. Other traits showed moderate GCV values, between 10 % and 20 %. These included plant height (11.1 %), spike length (13.2 %), spikes per plant (10.4 %), grains per spike (10.6 %), harvest index (11.9 %) and thousand-grain weight (12.4 %). On the other hand, traits related to growth timing had low GCV values, indicating they were less influenced by genetic differences. For example, days to 50 % flowering had a GCV of 6.6 %, days to maturity had 4.2 % and grain filling duration showed the lowest GCV at only 1.5 %. Other Results are summarised in Table 2.

Some traits showed high GCV values (above 20 %), including tiller number (23.0 %), grain yield (28.9 %) and biological yield (22.3 %), suggesting considerable genetic variability. Other traits showed moderate GCV values, between 10 % and 20 %. These included plant height (11.1 %), spike length (13.2 %), spikes per plant (10.4 %), grains per spike (10.6 %), harvest index (11.9 %) and thousand-grain weight (12.4 %). On the other hand, traits related to growth timing had low GCV values, indicating they were less influenced by genetic differences. For example, days to 50 % flowering had a GCV of 6.6 %, days to maturity had 4.2 % and grain filling duration showed the lowest GCV at only 1.5 %. Other Results are summarised in Table 2.

Some traits showed high GCV values (above 20 %), including tiller number (23.0 %), grain yield (28.9 %) and biological yield (22.3 %), suggesting considerable genetic variability. Other traits showed moderate GCV values, between 10 % and 20 %. These included plant height (11.1 %), spike length (13.2 %), spikes per plant (10.4 %), grains per spike (10.6 %), harvest index (11.9 %) and thousand-grain weight (12.4 %). On the other hand, traits related to growth timing had low GCV values, indicating they were less influenced by genetic differences. For example, days to 50 % flowering had a GCV of 6.6 %, days to maturity had 4.2 % and grain filling duration showed the lowest GCV at only 1.5 %. Other Results are summarised in Table 2.

Heritability and genetic advance

The estimates for broad-sense heritability showed a lot of differences among the wheat traits studied. High heritability was found in days to maturity (89.0 %), spike length (86.2 %) and plant height (79.2 %). which means that these traits are mainly controlled by genetics and making them good options for selection. Moderate heritability was found for traits such as grain yield (67.0 %), harvest index (66.8 %), spike length (64.3 %) and thousand-grain weight (64.4 %). This indicates that while genetics has a role in these traits, other factors such as environment also affect their development. On the other hand, some traits had lower heritability values. For instance, days to maturity showed 58.6 %, tiller number had 47.9 % and biomass yield was at 45.2 %. Grain-filling duration had a very low heritability of only 1.9 %. analysis that biological yield (113.09) and grain yield (77.55) can show significant improvement through selection, as they represent about 31 % and 49 % of their average values, respectively. Additionally, plant height (GA 18.27) and harvest index (8.85) also show good potential for genetic gain.

Table 2. Estimates of descriptive statistics and genetic variability

Trait	Mean	Std. Error	Std. Deviation	Min	Max	GCV	PCV	hBS	GA
DY50	66.75	0.25	4.91	54.27	77.27	6.57	6.97	89.02	8.54
DM	113.04	0.38	7.49	93.4	127.07	4.72	6.17	58.62	8.43
GFD	46.29	0.28	5.53	24.47	64.47	1.48	10.76	1.89	0.19
TN	9.77	0.16	3.23	1.03	22.32	22.99	33.23	47.87	3.21
PH	89.31	0.57	11.08	50.63	120.63	11.14	12.52	79.17	18.27
SL	9.41	0.07	1.39	6.02	17.57	13.22	14.24	86.18	2.38
SPP	17.39	0.13	2.52	7.29	25.29	10.38	12.95	64.31	2.99
GPS	52.24	0.39	7.66	21.97	75.97	10.55	12.98	66.07	9.24
GY	158.67	2.84	55.63	51.63	316.97	28.94	35.36	67	77.55
BY	365.58	7.39	144.83	0	764.87	22.31	33.19	45.18	113.09
HI	43.92	0.38	7.53	25.9	67	11.94	14.61	66.83	8.85
TGW	35.11	0.29	5.69	19.4	51	12.36	15.78	61.42	7.02

DY50: Number of days to 50 % flowering, DTM: Number of days to crop maturity, GFD: Grain filling duration (days), TN: Tiller number per plant, PH: Plant height (cm), SL: Length of the spike (cm), SPS: Number of spikelets/spike, GPS: Number of grains/spike, , GY: grain yield/plot, BY: Biological yield/plot (BY), HI: Harvest index (%), TGW: 1000-grains weight(g)

Correlation patterns amongst traits

The analysis of simple correlation coefficients has uncovered valuable insights that will inform future breeding strategies. Notably, we found a strong positive correlation between number of grains/spike and several key traits like the number of spikelets/spike (0.986), the spike length (0.221), grain yield/plot (0.182), weight of 1000 grains (0.147), biological yield per plot (0.107) and harvest index (0.106). other hand, plant height shows a significant negative correlation with grain yield (-0.121) and harvest index (-0.112). harvest index negative correlations were also found for (-0.086), grain filling duration (-0.077), days to crop maturity (-0.074), biological yield (-0.066), grain per spike (-0.053) and spiklets per spike (-0.052), these were not statistically significant (Table 3).

Harvest index and spike length were found to have significant positive connections with spiklets per spike (0.233), biological yield (0.229), grain per spike (0.221), grain yield (0.194), weight of thousand grains (0.167) and the number of days to 50 % flowering (0.104). Additionally, spiklets per spike had a strong positive relation with harvest index, with grain/spike (0.986), spike length (0.233), grain yield (0.192), weight of thousand grains (0.156), biological yield (0.119) and harvest index (0.11).

Furthermore, the biological yield/plot (biological yield) shows significant positive correlation with grain yield (0.871), weight of thousand grains (0.78), spike length (0.229) and the number of tillers per plant (number of tillers per plant) (0.156). Similarly, harvest index has significant positive associations with grain yield (0.209), weight of thousand grains (0.203), number of tillers per plant (0.174), spiklets per spike (0.11) and grain per spike (0.106).

Finally, it's important to note that grain yield (0.901), biological yield (0.78), harvest index (0.203), numbers of tiller per plant (0.186), spike length (0.167), Spikelets per spike (0.156) and grain per spike (0.147) all shows positive correlations with thousand grains weight.

Principal component analysis

Out of twelve principal components, two PC-1, PC-II, possessed eigenvalues greater than two (2.01), contributing 50.73 % of the entire cumulative variability amongst diverse genotypes (Table 4). Contribution of PC-I towards variability is highest (33.72 %), followed by PC-II, which contributed 17 % variability correspondingly. The PC-II revealed positive factor loadings for all characteristics exclusively related to yield and yield-related components. PC-II indicates positive factor loading for 100 grain weight and grain yield (kg per hectare), while negative factor loading for plant height, spike length, spikelets per spike and grains per spike. The first two principal components, which account for 50.73 % of overall variance, were plotted on PC-I (x-axis) and PC-II (y-axis) to identify the relationship between diverse clusters (Fig. 1). Out of 12 PC computed in the analysis, 6 have eigenvalues greater than 2. The principal components PC3 and PC4 explained variation of 11.59 % and 9.31 % (Supplementary file).

Cluster analysis

Grouping of genotypes based on the analysed traits is depicted in Fig. 2. Cluster analysis has successfully divided 381 wheat genotypes into 8 separate clusters. Cluster I contains 135 genotypes, followed by Cluster III with 111 genotypes, Cluster II

Table 3. Correlation among agronomic traits for 381 genotypes of wheat

Variables	DY50	DM	GFD	TN	PH	SL	SPP	GPS	GY	BY	HI	TGW
DY50	1											
DM	0.673	1										
GFD	-0.015	0.729	1									
TN	-0.068	-0.121	-0.1	1								
PH	-0.024	-0.074	-0.077	0.025	1							
SL	-0.033	-0.087	-0.088	-0.047	0.104	1						
SPP	-0.042	-0.057	-0.037	-0.082	-0.052	0.233	1					
GPS	-0.053	-0.06	-0.031	-0.092	-0.053	0.221	0.986	1				
GY	-0.459	-0.67	-0.481	0.22	-0.121	0.194	0.192	0.182	1			
BY	-0.3	-0.523	-0.43	0.156	-0.066	0.229	0.119	0.107	0.871	1		
HI	-0.244	-0.221	-0.073	0.174	-0.112	-0.07	0.11	0.106	0.209	-0.258	1	
TGW	-0.435	-0.61	-0.422	0.186	-0.086	0.167	0.156	0.147	0.901	0.78	0.203	1

significant at the 5 % level of probability

DY50: Number of days to 50 % flowering, DTM: Number of days to crop maturity, GFD: Grain filling duration (days), TN: Tiller number per plant, PH: Plant height (cm), SL: Length of the spike (cm), SPS: Number of spikelets/spike, GPS: Number of grains/spike, , GY: grain yield/plot, BY: Biological yield/plot (BY), HI: Harvest index (%), TGW: 1000-grains weight (g)

Table 4. Eigenvalues and coordinates of the eigenvector of 12 principal components for agro-morphological traits

| PC	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
Eigen Value	4.05	2.04	1.39	1.12	9.97	9.37	7.86	5.31	1.22	1.69	1.29	0.19
Variance %	33.719	17.003	11.593	9.311	8.309	7.81	6.555	4.425	1.025	0.141	0.108	0
Cumulative %	33.719	50.722	62.315	71.626	79.935	87.746	94.301	98.726	99.751	99.892	100	100
DY50	0.28	-0.09	0.33	-0.19	0.58	-0.24	-0.16	0.33	0.01	-0.03	0.00	-0.48
DM	0.42	-0.15	0.10	-0.30	0.09	0.25	-0.02	0.37	0.04	-0.02	0.00	0.70
GFD	0.30	-0.13	-0.17	-0.23	-0.41	0.56	0.12	0.19	0.04	-0.01	0.00	-0.52
TN	-0.11	0.18	-0.21	-0.08	0.66	0.57	0.19	-0.34	-0.06	-0.01	0.00	0.00
PH	0.02	0.06	0.23	0.80	0.08	0.24	0.28	0.40	0.05	-0.01	0.00	0.00
SL	-0.12	-0.23	0.35	0.18	-0.05	0.39	-0.76	-0.20	-0.03	-0.01	-0.01	0.00
SPP	-0.14	-0.65	-0.06	0.03	0.09	-0.03	0.19	-0.05	0.00	0.34	0.62	0.00
GPS	-0.13	-0.65	-0.06	0.04	0.08	-0.04	0.20	-0.06	-0.02	-0.33	-0.62	0.00
GY	-0.47	0.04	0.01	-0.18	0.00	0.07	-0.02	0.27	0.41	-0.63	0.33	0.00
BY	-0.41	0.06	0.36	-0.26	-0.04	0.11	0.16	0.12	0.41	0.56	-0.31	0.00
HI	-0.10	-0.03	-0.71	0.15	0.16	-0.05	-0.40	0.37	0.24	0.25	-0.15	0.00
TGW	-0.44	0.05	0.00	-0.17	-0.02	0.09	-0.02	0.41	-0.77	0.05	-0.02	0.00
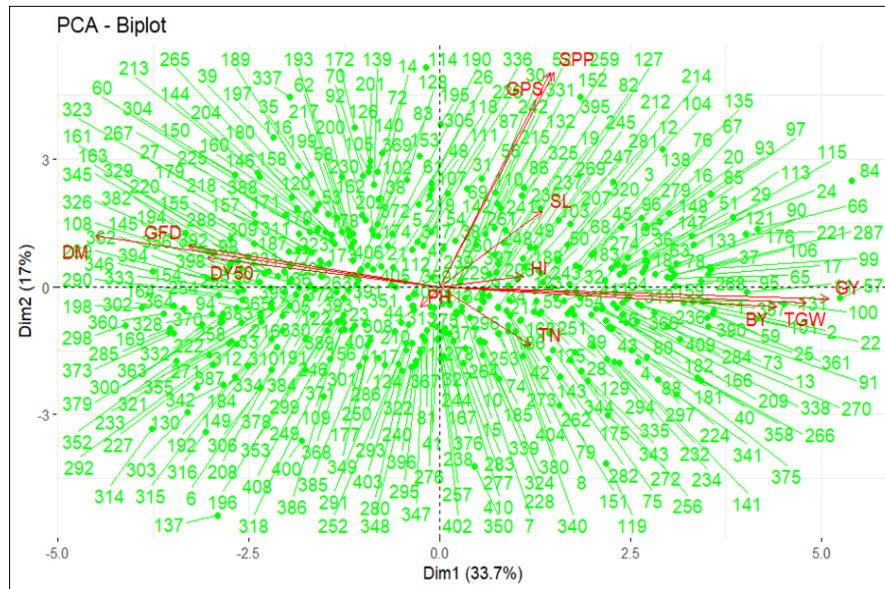
Fig. 1. Biplot of wheat genotypes for 1st two principal components (PC1 and PC2 50.73 %).

with 92 genotypes and thereafter Clusters V (31), IV (29), VIII (6), VI (5) and VII (2). Cluster I distinctly shows the shortest time to flowering, indicating a strong early growth potential and considerable adaptability to stressful conditions. It also features the highest grain yield (GY) at 232.17 g/plot and the top thousand-grain weight (TGW) at 41.60 g, confirming its exceptional grain quality.

On the other hand, Cluster III has the tallest plants at 94.21 cm, likely improving photosynthetic efficiency. The harvest index (HI) shows considerable variation across the clusters, with values ranging from 32.54 % in Cluster VII to an outstanding 48.87 % in Cluster III, highlighting remarkable differences in biomass allocation efficiency among the genotypes.

Cluster II has many genotypes with increased plant height, making these genotypes ideal for enhancing this desirable characteristic. There is significant variability in plant height (PH), with Cluster VI representing the shortest height at 75.02 cm, which offers notable advantages in terms of resistance to lodging. Genotypes in Cluster VIII have the highest days to

maturity (DM) at 120.17 days and the longest grain filling duration (GFD) at 51.95 days. This prolonged vegetative and reproductive period ensures yield consistency across diverse environmental situations, although Cluster VIII experiences the lowest yield at 97.75 g/plant, underscoring important genetic variability in productivity.

Table 5 presents the mean values for various agro-morphological traits of the wheat genotypes derived from the cluster analysis. Pairwise Mahalanobis distances (D² statistics) as described (19) can be found in Table 6. Genotypes in Cluster V demonstrate significantly high divergence values alongside a distinctive genetic profile, marking them as vital candidates for the introduction of innovative traits into the breeding program. Additionally, it is important to observe that the smallest differences were recorded between Clusters III and V, as evidenced by the minimal genetic divergence.
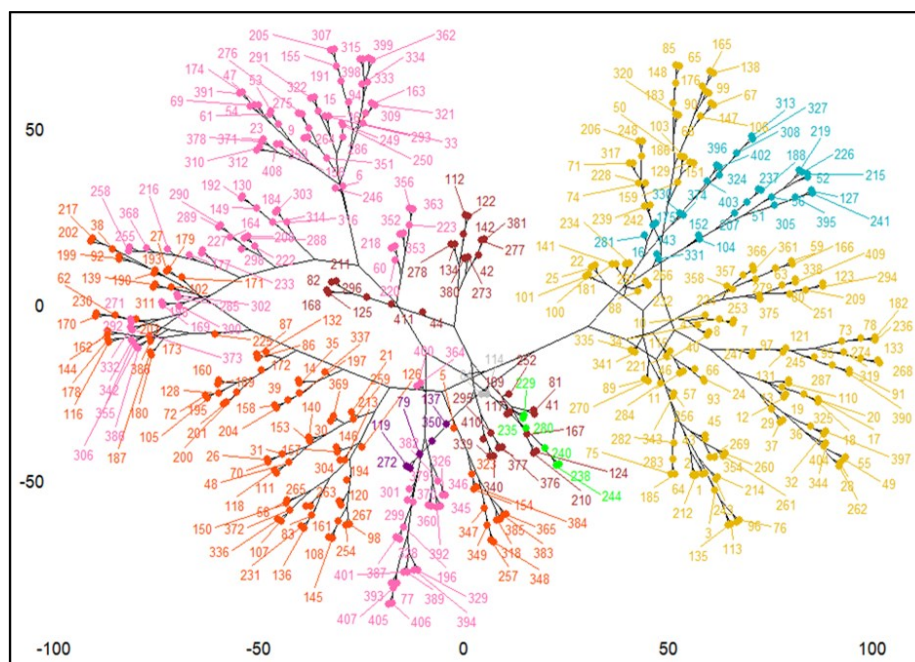

Fig. 2. Phylogenetic cluster dendrogram based on thirteen traits of different genotypes.

Table 5. Mean value of different agro-morphological traits of wheat genotypes

CLUSTER	DY50	DM	GFD	TN	PH	SL	SPP	GPS	GY	BY	HI	TGW
I	62.33	107.08	44.74	10.39	88.98	9.90	19.52	58.50	232.17	507.9	46.22	41.60
II	70.84	107.88	37.04	12.18	91.02	9.54	17.10	51.40	227.5	505.6	45.72	40.18
III	69.22	115.07	45.84	11.59	94.21	8.79	16.18	48.44	112.3	232.4	48.87	30.94
IV	69.92	117.92	48.00	8.07	89.43	9.97	18.96	57.15	135.61	327.0	41.62	33.25
V	62.03	105.82	43.78	10.11	89.91	9.44	16.49	49.63	189.92	426.0	44.63	38.13
VI	69.48	118.90	49.42	8.52	75.02	8.45	17.46	52.39	127.2	279.8	45.88	31.91
VII	71.65	117.08	45.43	9.61	85.37	10.25	15.63	46.36	184.02	579.2	32.54	36.82
VIII	68.22	120.17	51.95	9.34	92.26	8.76	14.87	44.63	97.75	260.0	38.12	29.55

Table 6. D^2 Statistics among eight different clusters

CLUSTER	I	II	III	IV	V	VI	VII	VIII
I	0							
II	7.05	0						
III	5.22	4.93	0					
IV	4.19	8.53	9.91	0				
V	10.76	9.67	13.30	12.20	0			
VI	4.85	5.86	7.69	5.60	12.93	0		
VII	3.38	7.50	6.12	6.94	12.66	8.47	0	
VIII	5.65	6.19	3.69	5.63	11.58	5.93	8.03	0

Discussion

A breeding program depends on genetic diversity, which is very important for getting high yields, good quality, adaptability to various conditions and resistance to pests and diseases. This knowledge helps breeders make better choices when selecting different parent plants for creating high-yielding varieties. Our study shows that environmental factors largely affect agronomic traits in wheat. We found that the treatments significantly affect almost all traits and the PCV is consistently higher than the GCV for all the traits we measured (20), traits like days to 50 % flowering and maturity are more influenced by the environment and show less genetic variation. However, we noticed high GCV for yield-related traits such as number of tillers, grain yield and biological yield. This indicates a lot of genetic variety which is good for breeding. Other studies have shown that traits with high genetic variation are usually easier to improve by selective breeding.

The study also found that traits like days to maturity, plant height and spike length are mostly determined by genetic factors. This supports the earlier findings (21), confirming that these traits are reliable for selection in breeding programs. We also observed that grain yield, harvest index, spike potential and thousand-grain weight have moderate heritability, which means breeders need to be careful with environmental variations since these traits are influenced by both genetics and the environment (22).

On the other hand, the low heritability of grain-filling duration shows that it relies heavily on environmental factors and may not improve much through selection. However, the significant genetic advancement seen for biological yield, grain yield and harvest index indicate that these traits have strong genetic potential and can improve positively through selection (23). Stress the importance of focusing on traits that show high heritability and genetic gain to boost wheat productivity. For traits like grain-filling duration, which have low heritability, different breeding strategies or better environmental management could be necessary for effective improvement. Biological yield and grain yield can potentially increase by about 31 % and 49 % of their average values, respectively, with careful breeding. This suggests that targeted breeding can lead to real improvements in these yield components. Additionally, plant height and harvest index also shows significant genetic progress, indicating they can be improved through selective breeding (24). Traits with high heritability and genetic gain can be targeted for better yield enhancement (25).

The present study elucidates a significant positive correlation between plant height and both spike length and the number of spikelets per spike. Former researchers evaluated bread wheat genotypes and established that the number of productive tillers is significantly and positively associated with key traits such as spike length, spikelets per spike, grain yield and thousand-grain weight, each of which also contributes indirectly to biological yield and harvest index (26).

Conversely, a negative correlation between plant height and grain yield also reported previously (27). Furthermore, spike length exhibits a highly significant and positive correlation with both spikelets per spike and grains per spike (28). Additionally, spikelets per spike show a significant positive correlation with plant height and a strong positive association with spike length and grains per spike. Former studies conducted an extensive investigation that confirmed the positive association of grains per spike with spikes per plant, spike length, grain yield, thousand-grain weight, biological yield and harvest index. The relationship between yield (kg/hectare) and grains per spike is likewise significant, which underscores the critical role of grain yield in enhancing overall yield. noted a negative association between grain yield and plant height was reported in earlier studies (29). Prior research has established that optimising the expression of yield-contributing traits, which exhibit a robust positive correlation with yield, can effectively increase the thousand-grain weight of wheat (29-31).

The current study revealed that, out of twelve principal components, only two-PC-I and PC-II exhibited eigenvalues greater than one and collectively accounted for 50.73 % of the total variability among the analysed genotypes. Notably, PC-I showed positive correlations with traits such as days to maturity, grain filling duration, days to 50 % maturity and plant height, whereas PC-II demonstrated positive loadings on tiller number, biological yield, Plant height, thousand-grain weight and grain yield. However, PC-II also highlighted negative loadings for traits such as plant height, spike length, spikelets per spike and grains per spike. This dual-axis structure, where

PC-I delineates a vegetative axis and PC-II a reproductive axis (32), enhances the clarity of the findings. A similar approach was employed in their 2 of agro-morphological traits in Mediterranean durum wheat landraces, which demonstrated the capability of PCA to cluster genotypes based on distinct agronomic characteristics (33) . Although their study allocated variation into more groups, the underlying theme remained consistent; key yield components were primarily associated with one or two principal components representing vegetative versus reproductive contributions.

Current investigation indicates that PCA, combined with genetic resource characterisation, highlights essential traits required for strategic selection in breeding programs. The first two (PC-I, PC-II) principal components, accounting for 50.73 % of the total variability, provide a framework for illustrating the relationships among diverse clusters. Notably, the results indicate that the thousand-grain weight is significantly and positively correlated with yield (kg/hectare). In contrast, plant height is positively correlated with spike length, spikelets per spike and grains per spike.

In parallel, earlier reports states that the first principal component was predominantly loaded with traits such as spike length, spikelets per spike and grain yield within wheat genotypes (32). This observation is consistent with our findings, as PC-I primarily reflects traits associated with vegetative growth, while yield-related characteristics are distributed across both PCs. Moreover, yield-associated traits such as spike length, spikelets per spike and grain yield contributed significantly to PC-I, while thousand-grain weight and overall grain yield were major contributors to PC-II was reported (34). These consistent outcomes across studies underscore the validity of PCA as a robust multivariate tool for capturing genetic divergence among wheat germplasms.

In genetic divergence Cluster I is a standout group because it flowers early and has the highest grain yield and thousand-grain weight when compared to other clusters. In contrast, Cluster VIII includes plants that mature very late, which means they have a longer grain-filling duration, but their yield potential is low. Cluster III is known for having the tallest plants and the highest harvest index, meaning they invest a lot of their growth into producing grains. other clusters display various traits. For example, Cluster VI consists of notably short plants, while Cluster II features plants that are taller than average. These differences highlight the significant genetic diversity among the genotypes, which is important for breeding wheat varieties that are both high-yielding and adaptable.

Breeding programs should focus on crossing high-yielding lines from Cluster I, recognised for their early maturity, with lines from the other clusters to take advantage of their useful traits. Identifying these trait clusters helps improve selection strategies by organizing genotypes based on multiple important traits. Previous studies support this idea, showing that grouping plants by traits can enhance breeding efforts (35). Additionally, crossing lines from different inter-cluster distances are crossed new combinations are formed that can result in superior plants exhibiting traits from both parents which leads to the development of superior new varieties (36).

Conclusion

The bread wheat genotypes exhibited a wide range of noticeable genetic variance for all traits studied which indicates substantial environmental influence on trait expression. However, some important traits related to yield, like grain yield and biological yield, had a high amount of genetic variation. This suggests there's good potential for selecting better varieties. On other hand, traits related to growth stages, like grain-filling duration and days to maturity have low genetic variation and heritability, indicating environmental factors mostly influence them. Traits such as plant height and spike length showed both high heritability and potential for genetic improvement, making them good choices for selection. Correlation analysis demonstrated strong positive relationships among key yield components, meaning that improving one trait generally boosts others. Principal component analysis revealed two main axes of variation, accounting for over 50 % of trait differences. The first component focused on growth and yield traits, while the second included growth stage and spike traits, clearly separating vegetative and reproductive characteristics. Crossing lines from significantly different clusters which leads to the development of superior new varieties.

In conclusion, the results suggest that breeders should focus on high-heritability traits that offer good potential for improvement. Future wheat breeding programs can efficiently create high-yielding and adaptable varieties by prioritizing these traits and using principal components for selecting parent plants.

Authors' contributions

AT played a pivotal role in executing the experiment by seamlessly managing analytical work and data collection with VK, SS, meticulously organising the resulting data and spearheading the drafting and revision of the manuscript with support from AS. AS and SS were instrumental in the initial phases, contributing significantly to the planning, design and refinement of the core experimental concept. SCV and PS provided essential oversight during the analytical process to ensure methodological rigour, while PG and NK performed the statistical analysis using RStudio, which informed key revisions that greatly enhanced the clarity and precision of the manuscript.

Compliance with ethical standards

Conflict of interest: The authors affirm that there are no financial, personal, or professional conflicts of interest that could have influenced the outcome or interpretation of this study. This research was conducted with integrity, transparency and a shared commitment to advancing scientific knowledge in the field of plant science. All authors have reviewed the manuscript and agree with its content. Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. FAO. Publications. 2024. https://www.fao.org/publications

2. Kormoczi P, Toth B, Nagy-Gyorgy A, Kocsis K, Ovari J, Szabo B, et al. SNP based genetic diversity assessment among Hungarian bread wheat (*Triticum aestivum* L.) genotypes. Cereal Res Commun. 2020;48(1):1–7. https://doi.org/10.1007/s42976-019-00005-z

3. Ahmad F, Khan S, Latif A, Khan H, Khan A, Nawaz A. Genetics of yield and related traits in bread wheat over different planting dates using diallel analysis. Afr J Agric Res. 2011;6(8):1564–71. https://doi.org/10.5897/AJAR11.1801

4. Khalid A, Hammed A. Genetic divergence in wheat genotypes based on seed biochemical profiles appraised through agglomerative hierarchical clustering and association analysis among traits. Pak J Bot. 2021;53(4):1281–6. https://doi.org/10.30848/PJB2021-4(7)

5. Singh PK, Singh L, Singh SV, Nath S, Yadav V, Mishra A, et al. Estimation of genetic variability, correlation and path analysis for yield and some yield contributing traits in bread wheat (*Triticum aestivum* L.). Int J Environ Climate Change. 2023;13:2429–40. https://doi.org/10.9734/ijecc/2023/v13i113409

6. Krishna S, Upadhyay P, Mishra VK, Yadav PKM, Singh P, Ashutosh SS, et al. Evaluation of terminal heat tolerance in bread wheat (*Triticum aestivum* L.). Indian J Hum Genet. 2020;8(4):468–70.

7. Upadhyay P, Gupta M, Sra SK, Sharda R, Sharma S, Sardana VK, et al. Genome-wide association studies for acid phosphatase activity at varying phosphorus levels in *Brassica juncea* L. Front Plant Sci. 2022;13:105602. https://doi.org/10.3389/fpls.2022.105602

8. Khan FUZ, Rehman SU, Ali MA, Waqas M, Chaudhry MH, Bilal M, et al. Exploitation of germplasm for plant yield improvement in cotton (*Gossypium hirsutum* L.). J Green Physiol Genet Genomics. 2015;1:1–10.

9. Khodarahmpour Z, Choukan R, Bihamta MR, Hervan E. Determination of the best heat stress tolerance indices in maize (*Zea mays* L.) inbred lines and hybrids under Khuzestan province conditions. J Agric Sci Technol. 2011;13:111–21.

10. Malik R, Sharma H, Sharma I, Kundu S, Verma A, Sheoran S, et al. Genetic diversity of agro-morphological characters in Indian wheat varieties using GT biplot. Aust J Crop Sci. 2014;8:1266.

11. Aharizad S, Sabzi M, Mohammadi SA, Khodadadi E. Multivariate analysis of genetic diversity in wheat (*Triticum aestivum* L.) recombinant inbred lines using agronomic traits. Ann Biol Res. 2012;3(5):2118–26.

12. Khodadadi M, Fotokian MH, Miransari M. Genetic diversity of wheat (*Triticum aestivum* L.) genotypes based on cluster and principal component analyses for breeding strategies. Aust J Crop Sci. 2011;5(1):17–24.

13. Aravind J, Sankar SM, Wankhede DP, Kaur V. augmentedRCBD: Analysis of augmented randomised complete block designs [software]. 2023. https://doi.org/10.5281/zenodo.8015094

14. Wei T, Simko V. corrplot: Visualization of a Correlation Matrix. R package version 0.92; 2021. https://CRAN.R-project.org/package=corrplot

15. R Core Team. stats: Statistical Functions. R package version 4.3.2; 2023. In: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

16. Husson F, Josse J, Le S, Mazet J. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining. R package version 2.4; 2020. https://CRAN.R-project.org/package=FactoMineR

17. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.4; 2023. https://CRAN.R-project.org/package=cluster

18. Kassambara A, Mundt F. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7; 2020. : https://CRAN.R-project.org/package=factoextra

19. Galili T. dendextend: Extending 'dendrogram' Functionality in R. R package version 1.13.4; 2015. https://CRAN.R-project.org/package=dendextend

20. Mahalanobis PC. On the generalized distance in statistics. Proc Natl Inst Sci (Calcutta). 1936;2:49–55.

21. Sharma G, Singh V, Singh KP, Tiwari A, Singh V, Patel A. Assessing Genetic Variability and Heritability in Wheat [*Triticum aestivum* L.]. Int J Curr Microbiol App Sci. 2020;11:477–82.

22. Sewore BM, Abe A. Genetic variability and trait associations in bread wheat (*Triticum aestivum* L.) genotypes under drought stressed and well watered conditions. CABI Agric Biosci. 2024;5 (1):64. https://doi.org/10.1186/s43170-024-00259-6

23. Zubair M, Akhtar LH, Minhas R, Jahan Bukhari MS, Hussain S, Rehman M, et al. The performance of potential wheat genotypes across diverse environments of Pakistan. Pak J Agric Sci. 2021;58 (6). https://doi.org/10.21162/PAKJAS/21.1226

24. Desheva GN, Deshev MG. Correlation and path coefficient analyses of grain yield and its related components for some common winter wheat genotypes. Agric Sci/Agrarni Nauki. 2021;13(31). https://doi.org/10.22620/agrisci.2021.31.005

25. Mecha B, Alamerew S, Assefa A, Dutamo D. Correlation and path coefficient studies of yield and yield associated traits in bread wheat (*Triticum aestivum* L.) genotypes. Adv Plants Agric Res. 2017;6(5):128–36. https://doi.org/10.15406/apar.2017.06.00226

26. Das A, Das A, Muntaha S, Akhter S, Sagor GS. Genetic variability, correlation and path analysis of yield and yield contributing characters in wheat (*Triticum aestivum* L.) under normal and terminal heat stress condition. Fundam Appl Agric. 2024;9(2):65–71. https://doi.org/10.5455/faa.187752

27. Uddin F, Mohammad F, Ahmed S. Genetic divergence in wheat recombinant inbred lines for yield and yield components. Acad J Agric Res. 2015;3:303–7.

28. Baloch AW, Baloch M, Baloch IA, Mari SN, Mandan DK, Abro SA. Association and path analysis in advance Pakistani bread wheat genotypes. Pure Appl Biol. 2014;3:115–20. https://doi.org/10.19045/bspab.2014.33005

29. Pradhan B, Upadhyay P, Singh SK, Yadav PK, Krishna B. Dissecting the genetic variability in Wheat (*Triticum aestivum* L.) genotypes using univariate and multivariate statistical approaches. Biopestic Int. 2024;20(2). https://doi.org/10.59467/BI.2024.20.295

30. Zewdu D, Mekonnen F, Geleta N. Correlation and path analysis for yield and yield component traits on bread wheat (*Triticum aestivum* L.) genotypes. Int J Bio-Res Stress Manag. 2024;15(1):1–3. https://doi.org/10.23910/1.2024.5040

31. Tiwari S, Singh B. Correlation and path coefficient analysis among yield and yield attributing traits of wheat (*Triticum aestivum* L.) genotypes. J Cereal Res. 2024;16(3):325–30. https://doi.org/10.25174/2582-2675/2024

32. Wani SH, Sheikh FA, Najeeb S, et al. Genetic variability study in bread wheat (*Triticum aestivum* L.) under temperate conditions. Curr Agric Res J. 2018;6(3):268–77. https://doi.org/10.12944/CARJ.6.3.06

33. Goel S, Singh M, Grewal S, Razzaq A, Wani SH. Wheat proteins: a valuable resource to improve nutritional value of bread. Front Sustain Food Syst. 2021;5:425. https://doi.org/10.3389/fsufs.2021.769681

34. Ajmal SU, Minhas NM, Hamdani A, Shakir A, Zubair M, Ahmad Z. Multivariate analysis of genetic divergence in wheat (*Triticum aestivum*) germplasm. Pak J Bot. 2013;45:1643–8.

35. Adilova S, Qulmamatova D, Baboev S, Bozorov T, Morgunov A.

Multivariate cluster and principle component analyses of selected yield traits in Uzbek bread wheat cultivars. Am J Plant Sci. 2020;11:903–12. https://doi.org/10.4236/ajps.2020.116066

36. Fellahi ZEA, Boubellouta T, Hannachi A, Belguet H, Louahdi N, Benmahammed A, et al. Exploitation of the genetic variability of diverse metric traits of durum wheat (*Triticum turgidum* L. ssp. *durum* Desf.) cultivars for local adaptation to semi-arid regions of Algeria. Plants. 2024;13(7):934. https://doi.org/10.3390/plants13070934

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy