

# Genetic Profiling and Characters Association in Wheat (*Triticum aestivum* L.) Genotypes: Implications for Breeding Programs

Satya Prakash<sup>1</sup>, Adesh Kumar<sup>2\*</sup> and Suhel Mehandi<sup>1</sup>

<sup>1</sup>Department of Genetics and Plant Breeding, School of Agriculture, Lovely Professional University, Phagwara, Punjab, India

<sup>2</sup>Department of Plant Pathology, School of Agriculture, Lovely Professional University, Phagwara, Punjab, India

## \*Correspondence to:

Adesh Kumar  
Department of Plant Pathology,  
School of Agriculture,  
Lovely Professional University,  
Phagwara, Punjab, India.  
E-mail: [adesh.19078@lpu.co.in](mailto:adesh.19078@lpu.co.in)

Received: October 11, 2023

Accepted: November 10, 2023

Published: November 15, 2023

**Citation:** Prakash S, Kumar A, Mehandi S. 2023. Genetic Profiling and Characters Association in Wheat (*Triticum aestivum* L.) Genotypes: Implications for Breeding Programs. *J Food Chem Nanotechnol* 9(S1): S524-S533.

**Copyright:** © 2023 Prakash et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY) (<http://creativecommons.org/licenses/by/4.0/>) which permits commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited.

Published by United Scientific Group

## Abstract

The investigation of breeding strategies targeted at improving the quality and efficiency of wheat germplasms is done methodically using research objectives, genetic variability and genetic diversity, correlation coefficient, path analysis, and principal component analysis. Examining characteristics like Spike length (SL), Number of tillers per plant (NTP), Plant height (PH), Days to 50% flowering (DFF), Days to maturity (DM), and Phenotypic coefficients of variation (PCV), the study finds important traits whose PCV and Genetic coefficients of variation (GCV) are significantly related to grain production. Except for Seed yield per plant (SYP) and Harvest index (HI), traits that demonstrate substantial heritability have been shown to be favorable for selective breeding. When desirable features in germplasms were identified using cluster analysis, several genetic subgroups were found, with Clusters III, IV, and IX showing the highest potential for production. Biplot analysis highlighted trait-genotype connections, indicating germplasm 13 and 26 as having very diverse genotypes, while principal component analysis demonstrated DFF, DM, PH, and Biological yield per plant (BYP) as the major influences of genetic variation. Findings reveal a noteworthy positive correlation between DFF, BYP, NTP, and PH with DM. In addition, traits DFF, DM, PH, NTP, and SL exhibit high heritability DM, PH, Total seed weight (TSW), and BY impacted quantity and quality of grain yield. To provide breeders with a larger genetic pool to develop selective breeding for robust wheat that is in line with sustainable agriculture, the research prioritized critical features and genotypes driving variability.

## Keywords

*Triticum aestivum*, Genetic diversity, Multivariate analysis, Principal component analysis

## Introduction

Wheat (*Triticum aestivum* L.) is the second most important cereal crop for the world's populations. In India wheat covers a total of 31.76 million hectares and production of 109.52 million tons. This self-pollinating crop is known for its relatively high niacin and thiamin content, which are crucial for producing the special protein called gluten. In recent years, India has produced a high amount of wheat, making it the world's second-largest producer of wheat. The production of wheat may still be improved, despite present efforts. This can be achieved by developing more efficient plant varieties that can adapt to changing environmental conditions and be planted at the appropriate time. Genetic variability is necessary to enhance any crop's genetic composition. Therefore, a crucial stage in wheat breeding is the identification of potential genotypes and associated attributes on variability [1]. One of the main goals of wheat breeding programs is to create

high-yielding wheat varieties [2]. However, yield is a complex trait, as it is determined by several important traits and how they interact with the environment [3].

To develop wheat cultivars with improved yield potential breeders have been working towards increasing the genetic yield potential of the crop while also improving its nutritional quality. With 67% of people suffering from malnutrition and 2 billion individuals experiencing hidden hunger caused by inadequate iron and zinc levels, it is vital to create new wheat genotypes with both high yield and nutritional value [4]. The assessment of genetic diversity is crucial in these breeding projects, especially with the complicated genetic architecture and low heritability of increasing grain production in wheat [5]. Correlation analysis is used to measure the degree and direction of the association between two variables. It is useful to determine whether and to what extent the variables are favorably or adversely associated [6], though path analysis uses regression analysis to examine the relationships between multiple variables simultaneously. It can be used to investigate complicated relationships and identify each variable direct and indirect effects on a particular trait [7].

In wheat breeding, clustering is an important technique for diversity research, Principal component analysis and Genetic divergence ( $D^2$ ) are used. Instead of using indices based on polygenic associations as was seen by Khare [8] for populations that can be differentiated,  $D^2$  statistics compute the generalized Mahalanobis distance. A statistical technique (PCA) is used to generate clusters for classifying and identifying divergent parents in wheat breeding [9]. Evaluating the type and extent of genetic diversity related to an inheritable characteristic is crucial for utilizing traits that enhance yield. Babar et al. [10] have suggested that utilizing  $D^2$  analysis to calculate the genetic diversity distance between two genotypes can aid in identifying superior parental candidates for hybridization. When genetically different genotypes interbreed, it is anticipated that they will generate superior hybrids and desired recombinants.

Genetic research plays a key role in overcoming these challenges, providing a means to describe the genetic determinants of grain yield and, allowing for specific enhancements tailored to different growing conditions worldwide [11]. However, the extensive monoculture of popular cultivars has led to a reduction in genetic diversity, making crops more susceptible to biotic and abiotic challenges. Plant breeding programs must

expand their genetic background to retain enough diversity and produce new, different varieties that can withstand various stress conditions [12, 13].

## Material and Methods

### Experimental site

The experiment was carried out at the farm of Department of Genetics and Plant Breeding, School of Agriculture, Lovely Professional University, Jalandhar, Punjab located 225 meters above mean sea level at latitude of 31.24 N and longitude of 75.69E. The trial was conducted during the rabi season of academic year 2021–2022.

### Experimental material

Thirty genotypes of wheat (*Triticum aestivum* L.) germplasms from Kapurthala district were used as experimental material with PBW-65 and C-306 check. Genotypes were collected from the department of genetics and plant breeding (Table 1). Germplasms were sown during the first fortnight of November.

### Layout and design

The experiment had thirty genotypes with three replications and was designed using Randomized complete block design (RCBD). Each plot was three meters long, with a 15 centimeter between plants and a 25 cm between rows distance (25 x 15). The crop was grown using every cultural operation.

### Morphological data collection

Five randomly chosen representative plants from each genotype in each replication are used to collect data on morphological traits. Nine quantitative traits, including days to 50% flowering, days to maturity, plant height (cm), number of tillers per plant, spike length, 1000 grain weight (g), biological yield per plant (g), harvest index (%), and seed yield per plant (g), were used to collect information.

### Statistics analysis of data

#### Estimations of variance components

In germplasm the variation was quantified through the measurement of mean, phenotypic and genotypic variances. For a combined analysis of variance using the Agricolae package of R, the mean values for each genotype were used [14].

**Table 1:** Experimental material details.

Genotype code	Genotypes/germplasm	Genotype code	Genotypes/germplasm	Genotype code	Genotypes/germplasm
G1	PBW-226	G11	DBW-222	G21	WB-2
G2	WG-357	G12	WG-377	G22	DBW-327
G3	K-1006	G13	LOK-1	G23	DBW-168
G4	PBW-154	G14	K-9107	G24	RAJ-1482
G5	DBW-187	G15	UP-262	G25	CG-1018
G6	GW-322	G16	DBW-303	G26	WH-147
G7	NW-1014	G17	PBW-175	G27	NP-4
G8	PBW-65 Check	G18	8-A	G28	HD-3226
G9	K-68	G19	C-306 Check	G29	HD-3086
G10	TYPE-11	G20	K-65	G30	DBW-39

### Correlation coefficients analysis

The R Package 'metan' version 1.16.0 (GitHub-TiagoOlivoto/metan: Package for multi-environment trial analysis) was utilized to perform graphical and numerical visualization of a correlation matrix in R version 4.0.2 (Index of /src/base/R-4 (r-project.org)) [15].

### Multivariate analysis

Principal components analysis was calculated to estimate the eigenvalues, association between cultivars and traits, and contributions with quality of traits and genotypes towards principal components of the total variation. The multivariate analysis was performed utilizing R (v4.1.1) and R Studio software (v1.4.1717). The PCA-biplot was assembled using "ggplot2", "Factoextra", and "FactomineR" packages of R. The cluster plot with heatmap was visualized with the "heatmap" package of R [16].

## Results and Discussion

### Variability analysis among the studied traits

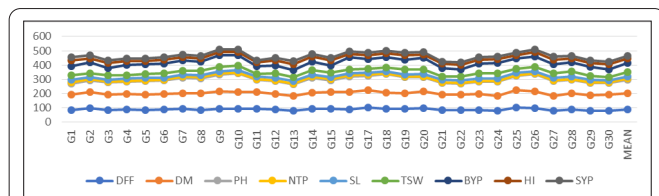
#### ANOVA using mean squares for yield

The Analysis of variance (ANOVA) results presented in (Table 2), there were highly significant differences observed among all variables, indicating an appropriate level of genetic variability among all genotypes. Which strongly suggests the presence of an appropriate level of genetic variability among the studied genotypes [17].

**Table 2:** Mean squares analysis of variance for grain yield.

Source	Replication	Treatment	Error
Degree of freedom (Df.)	2	29	58
Days to 50% flowering (DFF)	1.12	143.30**	3.11
Days of maturity (DM)	17.17	81.72**	2.82
Plant height (PH)	13.13	672.30**	24.39
Numbers of tillers per plant (NTP)	0.05	3.72**	0.09
Spike length (SL)	0.74	7.93**	0.70
Total seed weight (TSW)	11.00	50.65**	15.95
Biological yield per plant (BYP)	30.65	279.77**	80.57
Harvest index (HI)	63.65	87.21**	44.02
Seed yield per plant (SYP)	5.94	12.00**	6.19

Note: \*\*= Significant at a 5% level of probability.



**Figure 1:** Average genotype performance for different traits.

### Average mean performance of genotypes

Figure 1 and table 3 include the data as well as information on the general mean, range, standard error, and crucial difference (CD) for each attribute.

The DFF general mean has a mean value of 88.35 with

**Table 3:** Average performance, CV, S.E., and CD for different traits.

Genotype	Mean	C.V.	S.E.	C.D. @ 5%
DFF	88.35	2.00	1.02	2.88
DM	112.63	1.49	0.97	2.75
PH	96.89	5.10	2.85	8.07
NTP	5.84	5.13	0.17	0.49
SL	16.67	5.02	0.48	1.37
TSW	29.47	13.55	2.31	6.53
BYP	65.34	13.74	5.18	14.67
HI	29.22	22.71	3.83	10.84
SYP	18.37	13.54	1.44	4.07

a range of 77.00 to 103.67. The range of the mean value for DM was 102.44 to 122.89, with a mean value of 112.83. With a mean value of 96.89 cm, Plant height varied from 76.62 to 131.20 cm, and the average NTP was 5.84. The SL ranged from 10.96 to 20.64, with a mean value of 16.67. The TSW varied from 22.00 to 37.99 g, with a mean of 29.47 g. The BYP ranged from 49.84 to 82.91 g, with 65.34 g as the average, and the HI ranged from 21.83 to 41.96 g, with an average value of 29.22% and the average SYP is 18.37 g, with a range of 14.60 to 24.10 g.

The Coefficient of variation (CV) is a measure of the degree of variation among the genotypes. The CV ranged from 1.49 to 22.71, with a mean value of 9.14. According to the result, the wheat genotypes are good variables because they have a large coefficient of variation for parameters that contribute to yield, such as SL, TSW, BYP, and BYP is high. With a range of 0.49 to 14.67, the average CD value was 5.74. SL, TSW, BY, and SYP are the critical variations for the variable characteristics that contribute to yield.

### Genetic variations determination

#### Variance in phenotypic and genomic coefficients

The genetic variation within a population is measured by the GCV values, which vary from 4.553 to 18.834. The NTP (18.83) has the highest GCV value, while the DM has the lowest GCV value (4.55). The PCV values, which measure all genetic and environmental variation within a population, also range from 4.634 to 19.066. NTP is the trait with the highest PCV value, while DM is the trait with the lowest (Figure 2).

### Heritability and genetic advance

Heritability refers to the proportion of genetic variance relative to phenotypic variance, serving as a useful indicator of the transfer of traits from one generation to the next. The nine traits under study have heritability ranging from 49.50 to 97.80. DFF, DM, PH, NTP, and SL are the traits with the highest heritability values, while SYP and HI are the traits with the lowest heritability. The biological yield and TSW have moderate heritability values. The NTP and the DM are the two traits where the genetic advance mean, a measurement of the expected progress in a characteristic due to selection, is highest.

Highly heritable traits including DFF, DM, PH, NTP, and SL are more likely to respond to selection in breeding programs. Seed yield/plant and HI, on the other hand, have

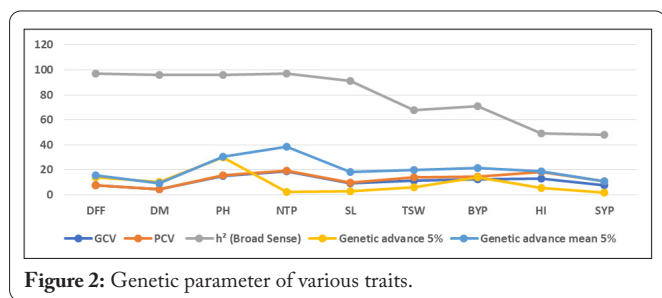


Figure 2: Genetic parameter of various traits.

lower heritability and might be less responsive to selection. On the traits, SYP and HI have lower heritability and may be less responsive to selection. Also, it is expected that selection for DM will be less effective due to genetic advancement than the mean for NTP at present. The high heritability means that a large proportion of the variation in the trait is due to genetic factors, making it more likely that the trait can be improved through selective breeding (Figure 2).

The genetic variability of different wheat genotypes in relation to multiple quantitative traits during the Kharif season Traits with higher genetic coefficient of variation (GCV) and phenotypic coefficient of variation (PCV) values are more likely to respond to selection, while traits with lower GCV and PCV values may be less susceptible to selection [18-21]. The NTP trait shows high GCV and PCV values, indicating a good response to selection. Traits such as DFF, DM, PH, NTP, and SL exhibit high heritability, suggesting a strong genetic influence and potential for improvement through selective breeding [22, 23]. Seed yield per plant and HI have low heritability, indicating they are less influenced by genetics and may be more challenging to improve through selective breeding [24, 25]. The TSW and BYP show moderate heritability, indicating a combination of genetic and environmental factors. The NTP trait also shows a high genetic advance mean, indicating a good response to selection, while the DM trait shows a low genetic advance mean, suggesting a poor response to selection. Similar findings have been reported by studies like Balkan and Rehman et al. [26, 27].

### Trait association

#### Genotypic correlation

Table 4 indicates that DFF, DM, PL, and TSW are highly positively correlated with BYP. Whereas DFF, DM, PH, TSW, and BYP are highly negatively correlated with the HI.

Additionally, DFF, DM, and PH are positively correlat-

ed with the NTP. Furthermore, DFF are positively correlated with PH. Finally, DFF are highly positively correlated to DM, and DM are highly positively correlated to PH.

This suggests that as DFF, DM, PH, and TSW increase, the BYP also increases. Similarly, as the DFF, DM, PH, TSW, and BYP increase, the HI decreases.

The positive correlation between DFF, DM, and PH with NTP suggests that as these traits increase, the NTP also increases. The positive correlation between DFF and PH suggests that as DFF increases, PH also increases.

The strong positive correlation between DFF and DM suggests that as DFF increases, DM also increases. Similarly, the strong positive correlation between DM and PH suggests that as DM increases, PH also increases.

#### Phenotypic correlation

Correlations between DM and various attributes were analyzed. Positive and significant correlations were found between DM and DFF (0.72\*\*\*), BYP (0.38\*\*\*), NTP (0.38\*\*\*), and PH (0.44\*\*), while negative and significant correlations were found with SYP (-0.24\*\*) and HI (-0.40\*\*\*). However, there were negative, non-significant correlations found between DM and SL (-0.12) and TSW (-0.16).

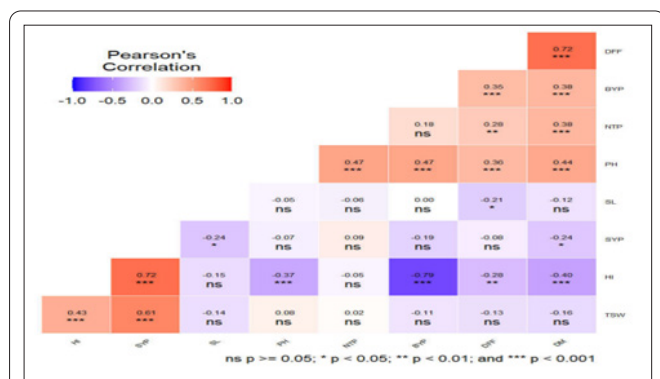
Positive and significant correlations were also found between DFF and BYP (0.35\*\*\*), NTP (0.28\*\*), and PH (0.36\*\*\*), while negative and significant correlations were found with SL (-0.21\*) and HI (-0.28\*\*). Non-significant negative correlations were found between DFF and SYP (-0.08) as well as TSW (-0.13).

The trait PH showed positive and significant correlations with BYP (0.47\*\*\*) and NTP (0.18), while significant negative correlation of -0.79\*\* was observed between BYP and HI. Non-significant negative correlations were found (-0.19) between BYP and SYP, as well as (-0.11) between BYP and TSW. Positive non-significant correlations were found between NTP and SYP (0.09) and TSW (0.02), while negative non-significant correlations (Figure 3).

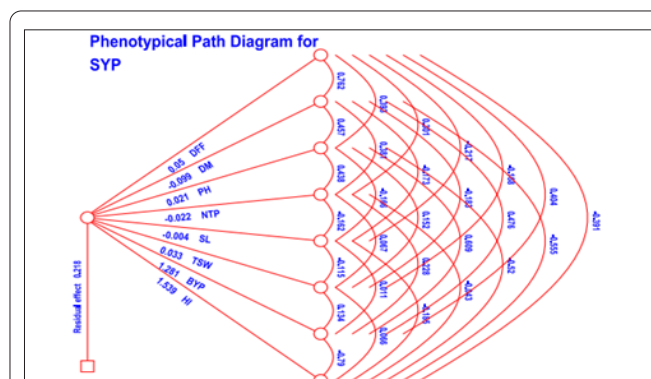
Positive and significant correlations were found between DM and DFF, BYP, NTP, and PH, indicating that as DM increases, these attributes also increase. The traits DFF, DM, PH, NTP, TSW, SL, BYP, and HI are identified as key factors influencing wheat SYP and can be used as selection criteria in breeding programs [28, 29]. The model used in the analysis has a high R-square value and residual effect, indicating a good fit

Table 4: Genotypic correlation on various traits.

Traits	DFF	DM	PH	NTP	SL	TSW	BYP	HI	SYP
DFF	1.000	0.781**	0.411*	0.313	-0.219	-0.203	0.460*	-0.542**	-0.163
DM		1.000	0.463**	0.382*	-0.202	-0.216	0.569**	-0.796**	-0.436
PH			1.000	0.422*	-0.236	0.221	0.738**	-0.727**	0.013
NTP				1.000	-0.220	0.111	0.273	-0.039	0.330
SL					1.000	-0.095	0.019	-0.242	-0.328
TSW						1.000	0.410*	-0.485**	-0.135
BYP							1.000	-0.807**	0.412
HI								1.000	0.222
SYP									1.000



**Figure 3:** Correlation coefficient (corrplot) among grain yield and yield-related traits.



**Figure 4:** Direct and indirect effects at phenotypic levels of nine traits in wheat.

and explaining most of the data variation. Similar results have been reported in studies by Thapa et al. & Abo-Elwafa et al. [30, 31].

### Path coefficient analysis

The results of path analysis indicate that SYP had a positive direct effect on DFF (0.049), PH (0.021), TSW (0.033), BYP (1.281) and HI (1.539). However, negative direct effects were recorded for DM (-0.099), NTP (-0.022), and SL (0.004). BYP and HI have very high direct effects.

Nine characters studied have positive indirect effects on DFF, including DM, PH, NTP and BYP and negative indirect effects including TSW, SL and HI. The DM character has positive indirect effects on SL, TSW, HI and negative indirect effects on DFF, PH, NTP, and BYP. The PH character has positive indirect effects on DFF, DM, NTP, TSW, and BYP, and negative indirect effects on SL and HI. Number of tillers/plants has positive indirect effects on SL and HI, and negative indirect effects on DFF, DM, PH, BYP and TSW. The SL character has low positive indirect effects on all characters. 1000 seed weight has positive indirect effects on PH, NTP, BYP, and HI and negative indirect effects on DFF, DM, and SL. Biological yield has high positive indirect effects on DFF, DM and PH, medium positive indirect effects on NTP, and very high negative indirect effects on HI. Harvest index has a low positive indirect effect on TSW, and very High negative indirect effects on BYP, high negative indirect effects on DFF, DM and PH, as well as medium negative indirect effects on SL. The phenotypic path correlation R-square value is 0.9527 and the residual effect is 0.2175 (Table 5).

This suggests that SYP has a positive direct effect on DFF, PH, TSW, BYP, and HI and negative direct effect on DM, NTP, and SL. Biological yield and HI are showing very high direct effects. The characters studied have positive and negative indirect effects on DFF, DM, PH, NTP, TSW, SL, and HI. The DM, PH, NTP, TSW, SL, and HI have positive and negative indirect effects on each other. The BYP has high positive indirect effects on DFF, DM, and PH and very high negative indirect effects on HI. The HI has negative indirect effects on DFF, DM, PH, NTP, TSW, SL, and BYP (Figure 4).

The 1000 seed weight had positive indirect effects on PH, NTP, BYP, and HI but negative indirect effects on DFF, DM, and SL. Biological yield had strong positive indirect effects on DFF, DM, and PH, and medium positive effects on NTP, but it had very high negative effects on HI. The harvest index had a low positive effect on TSW but had very high negative effects on BYP, high negative effects on DFF, DM, and PH, and medium negative effects on SL. Similar results have been reported in studies by Baye et al. and Khanal et al. [32, 33].

### Genetic divergence

Characters like DFF, NTP, SL, PH, TSW and SYP showed the high percent contribution towards the overall genetic divergence, whereas characters like DM, BYP and HI made a low contribution to the total genetic divergence according to the percent contribution of each character towards total diversity (Table 6).

### Cluster analysis by Torcher's method

The results, cluster III and IX have the highest genetic diversity among all the clusters, which corresponds to the gen-

Table 5: Phenotypic path analysis.

Traits	DFF	DM	PH	NTP	SL	TSW	BYP	HI
DFF	0.0497	0.0379	0.0196	0.015	-0.0108	-0.0084	0.0201	-0.0194
DM	-0.0756	-0.0992	-0.0454	-0.0378	0.0172	0.0182	-0.0472	0.0551
PH	0.0083	0.0097	0.0211	0.0093	-0.0035	0.0032	0.0129	-0.011
NTP	-0.0065	-0.0083	-0.0095	-0.0217	0.0035	-0.0014	-0.0049	0.0009
SL	0.0009	0.0007	0.0007	0.0007	-0.0043	0.0005	0.0000	0.0008
TSW	-0.0055	-0.006	0.0049	0.0022	-0.0037	0.0325	0.0043	0.0021
BYP	0.5175	0.6099	0.7806	0.2916	0.0137	0.1713	1.2813	-1.0117
HI	-0.6013	-0.8539	-0.7999	-0.0658	-0.2853	0.1008	-1.215	1.5388

**Table 6:** Percent contribution of traits towards the divergence.

Source	Times ranked 1st	Contribution %
DFF	10	2.30%
DM	4	1.28%
PH	12	8.76%
NTP	20	16.67%
SL	18	22.53%
TSW	11	1.30%
BYP	2	8.23%
HI	9	11.23%
SYP	35	9.05%

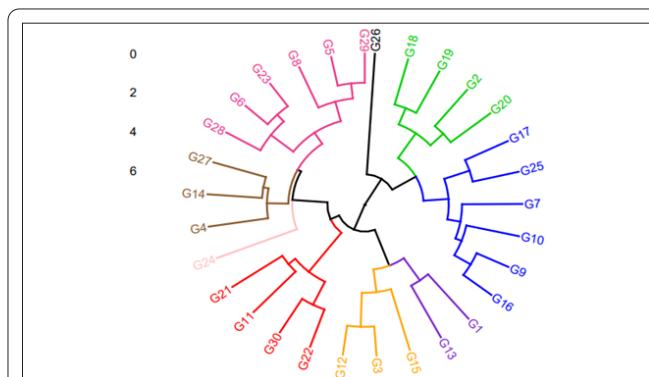
otypes that are most like one another. Cluster II and VI have the second-highest diversity among the clusters, with four genotypes. Cluster IV has the fourth-highest diversity among the clusters, with two genotypes. The remaining I and VII clusters each had a single genotype. This suggests that these genotypes differ from others within their respective clusters. This information may help to guide breeding decisions, such as selecting parents for crosses or identifying genotypes that may be useful in different environments (Figure 5).

#### Inter and intra cluster distances among nine clusters

The lower the distance between two clusters, the more similar the varieties in those clusters are to each other. Cluster I and Cluster III are 44.8 units apart, indicating that the varieties in these two clusters are relatively dissimilar. On the other hand, Cluster VI and Cluster VII are only 19.75 units apart, suggesting that the varieties in these two clusters are very similar. This information can be used to group similar varieties together and potentially make decisions regarding breeding, cultivation, and selection (Table 7).

#### Mean values of different clusters

The result represents the cluster mean for various parameters of wheat genotypes, grouped into 9 different clusters (Table 8). The mean values of the parameters differ among the clusters, indicating the presence of different subgroups or varieties of wheat. The highest mean values for DM, PH, TSW, BYP and SYP are observed in clusters III, IV, and IX, while the lowest values are found in clusters VII and VIII. Clusters



**Figure 5:** The dendrogram depicting the distribution of 30 wheat genotypes into 6 diverse clusters.

III, IV and IX have the highest mean values for DM, PH, and SYP, indicating that they could be the most productive subgroups of wheat. Clusters VII and VIII have the lowest mean values for NTP and SYP, suggesting that they might not be as productive as the other clusters. The mean values for SL, HI and SYP indicate that cluster IX might be the most promising subgroup for seed production. The results suggest that the different subgroups of wheat present in the clusters have varying agronomic characteristics and potential for seed production.

Several traits, including DFF, NTP, SL, PH, TSW, and SYP, were significant role in the genetic divergence [9]. Clusters III and IX showed the highest genetic diversity, followed by Cluster II, Cluster VI, and Cluster IV, each with a unique genotype. The distance between clusters indicates the similarity between genotypes within those clusters, with Cluster I and Cluster III showing significant dissimilarity among their varieties [34, 35].

High mean values for several characteristics were present in Clusters III, IV, and IX, indicating the potential for these subgroups to be fruitful. Based on its mean values for SL, HI, and SYP, Cluster IX demonstrated favorable traits for seed production [36].

#### Principal component analysis (PCA)

Eigenvalues represent the amount of variation in the result that is explained by each principal component. Days to 50% flowering has the highest eigenvalue of 3.385, indicating that it is the most important variable in the analysis. The

**Table 7:** Estimates of average inter and intra- cluster distances for 9 clusters in wheat.

Cluster	I	II	III	IV	V	VI	VII	VIII	IX
I	30.98	43.99	44.8	43.78	49.16	58.67	46.78	64.56	78.24
II		0.00	62.8	73.64	50.93	91.89	26.58	83.32	104.35
III			0.00	59.75	24.32	55.11	48.06	94.22	43.39
IV				0.00	75.68	31.64	79.04	39.32	76.34
V					0.00	76.77	30.09	104.82	63.29
VI						19.75	90.35	66.8	53.96
VII							0.00	98.54	89.84
VIII								0.00	114.9
IX									0.00

**Table 8:** Cluster mean among all cluster group.

Cluster	DFE	DM	PH	NTP	SL	TSW	BYP	HI	SYP
I	87.02	112.33	93.48	5.79	16.88	29.43	64.24	29.28	18.09
II	85.07	106.33	79.86	6.4	17.8	33.05	62.4	39.27	24.1
III	90.67	117.44	131.2	8.2	17.88	29	72.96	23.58	16.9
IV	96	117.22	101.3	4.6	17.44	30	82.34	22.94	18.55
V	91.67	110.89	106.96	7.8	15.18	27.99	70.08	24.4	16.8
VI	92.83	112.17	119.38	5.1	16.18	29.5	71.42	27.08	19
VII	92.33	118.78	81.92	7	15.07	29	50.73	37.68	18.8
VIII	85	108.67	80.06	3	19.02	25.02	58.47	28.55	16.4
IX	96.67	116.11	123.62	6.4	10.96	33.01	71.13	31.2	21.8

first PC explains 37.611% of the overall variation, while the first three components (PC1, PC2 and PC3) together explain 72.122% of the total variation.

This result provides information on which variables are most important in explaining the variation in the data and how much of the total variation is explained by each component. These results can be useful in identifying underlying patterns or relationships among the variables included in the analysis (Table 9).

#### Principal component analysis of different yield traits in wheat

Principal component first has relatively high positive loadings for DFF (0.403), DM (0.459), PH (0.406), BYP (0.426) and a negative loading for HI (-0.439), suggesting that these variables have the highest contribution to PC1. Principal component second has relatively high positive loadings

for SYP (0.595), TSW (0.392) and NTP (0.380), indicating that these variables are most strongly related to PC2. Principal component three has high positive loading for PC1 and negative loadings for PC2, indicating that it captures some of the variance unique to PC1 that is not captured by PC2 (Table 10). These loadings suggest that there is a complex relationship among these variables, with different groups of variables contributing to different Principal Component (PCs). The interpretation of these PCs will depend on the specific context and research question, but they can be used to reduce the dimensionality of the data and identify the most important variables for each PC.

#### Graphical presentation of the scree plot

Kaiser's Rule for PCA is a method for determining the optimal number of principal components to retain based on the eigenvalues, with a suggested threshold of retaining only those components with eigenvalues greater than or equal to 1. The eigenvalue is a measure of the amount of variance explained by each principal component, with larger eigenvalues indicating a greater amount of variance explained.

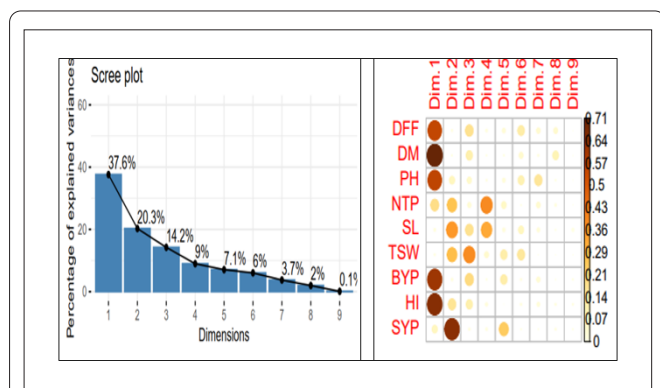
The scree plot (Figure 6) showed the percentage of variation for each of the principal components. Principal component first through PC9 explained 100% of the variance. Principal component first showed the largest portion of the total variance (37.6%), while PC2 showed 20.3%. The remaining PCs 3 to 9 showed, in order, 14.2, 9, 7.1, 6, 3.7, 2, and 0.19% of the overall variance respectively. As a result, three PCs can be used to demonstrate the contribution and quality of top-performer genotypes and variables to total variability.

**Table 9:** Principal components indicating eigen values, proportion of variation and total variation (%) of different traits.

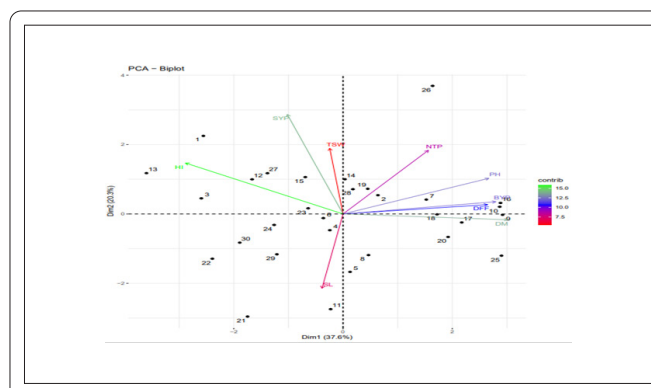
Traits	Eigen value	Proportion of variance (%)	Cumulative variance (%)
DFE	3.385	37.611	37.611
DM	1.820	20.275	57.887
PH	1.281	14.235	72.122
NTP	0.808	8.979	81.101
SL	0.636	7.067	88.169
TSW	0.542	6.029	94.199
BYP	0.335	3.728	97.927
HI	0.177	1.972	99.900
SYP	0.008	0.099	100.000

**Table 10:** Principal component analysis of different yield traits in wheat.

Traits	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
DFE	0.403	0.055	0.376	-0.114	0.212	-0.498	0.351	0.512	-0.016
DM	0.459	-0.034	0.307	-0.039	-0.070	-0.300	-0.138	-0.759	0.029
PH	0.406	0.213	-0.205	0.079	-0.192	0.447	0.695	-0.137	-0.011
NTP	0.237	0.380	0.166	0.693	-0.319	0.050	-0.360	0.234	0.018
SL	-0.058	-0.444	-0.373	0.634	0.173	-0.382	0.260	-0.111	-0.004
TSW	-0.037	0.392	-0.546	-0.245	-0.445	-0.534	-0.010	0.005	-0.008
BYP	0.426	0.073	-0.395	-0.064	0.445	0.103	-0.323	0.051	-0.579
HI	-0.439	0.304	0.305	0.137	-0.001	-0.114	0.261	-0.193	-0.694
SYP	-0.155	0.595	-0.075	0.109	0.619	-0.050	0.047	-0.187	0.424



**Figure 6:** A scree plot of variability explained by each component of 30 wheat germplasm.



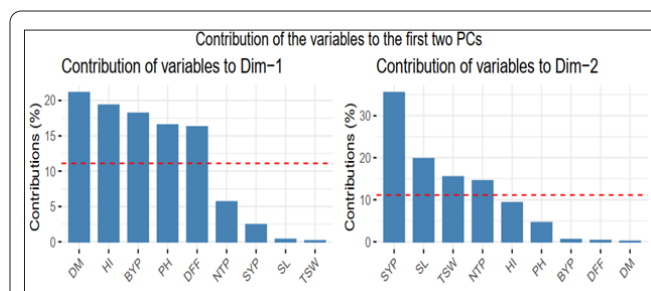
**Figure 7:** Biplot indicating the best genotype for the target traits.

### PCA-Biplot Analysis combined with cos2 and contribution

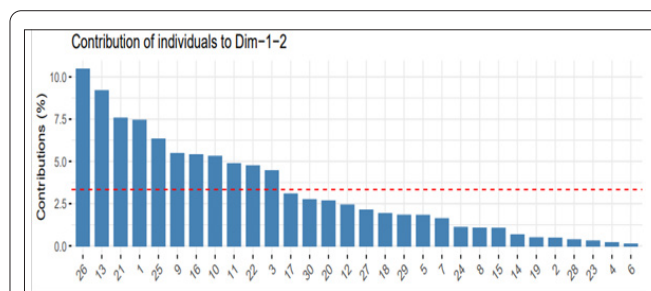
The first two main principal components of the biplot analysis, which shows a 57.9% variation, depict the relationship between traits, attributes, and genotypes (Figure 7). The two first principal components were represented by four coordinates (Coordinate-1, Coordinate-2, Coordinate-3 and, Coordinate-4) along the x- and y-axes in the biplot. Coordinate-1 is comprised of eight genotypes: G2, G7, G10, G14, G16, G19, G26 and G28, with the first PC having a positive value. The genotypes of these coordinates were related to significant factors, including the DFF, BYP, PH, and NTP. The most diverse genotype, G16, is far from the origin, followed by G10, whereas other genotypes are located near the center. The BYP, which showed the most variance, was followed by the PH, which had the longest vector. Coordinate-2 is comprised of 7 genotypes, namely, G1, G3, G12, G13, G15, G23 and G27, with the positive and negative values of the second PCs. G1, G3, G12, G13, G15, G23, and G27 are the seven genotypes that make up Coordinate-2, along with the positive and negative values of the second PCs for each. The most diverse genotype is G13, which is positioned far from the center and is followed by G1, while other genotypes are located near the center. The HI exhibited the longest vector, followed by the SYP, which reflected the greatest variance. Coordinate-3, which is comprised of genotypes such as G4, G11, G21, G22, G24, G29 and G30, is plotted in the direction of negative values, where the percent of the SL trait is negatively associated with those genotypes. Coordinate-4, which is comprised of genotypes such as G5, G8, G9, G17, G18, G20, and G25, is plotted in the direction of the positive and negative value of coordinate-4, where the percentage of the DM trait is strongly associated with those genotypes.

### Contribution along with the variables and genotypes

In PC1, which had positive and negative values in the first PCs, respectively, there are a maximum of 11 genotypes. The 11 genotypes largely contribute to the characteristics of DM, HI, BYP, PH, and DFF. In the first PC, DM exhibited the longest line, indicating the highest variance among the five variables. Principal component second exhibited greater variability among the four traits, as evidenced by the higher variance of SYP. There was some genetic variability across the various wheat breeding lines, according to the biplot genotype distribution (Figure 8).



**Figure 8:** Graphical representation of PC1 and PC2 percent of the contribution of the 5 variables.



**Figure 9:** Contribution of individuals to PC1 and PC2.

Among the 30 genotypes, G26 exhibited the highest percentage of contribution towards the overall diversity, following where genotypes G13, G21, G1, G25, G9, G16, G10, G11, G22, and G3 are located respectively (Figure 9).

The analysis revealed significant variation among genotypes, with DFF being the most significant variable. The principal component, influenced by DFF, DM, PH, and BYP, explained a significant percentage of overall variation [37]. The scree plot analysis verified that the first nine principal components PC1 and PC2 explained the entire variance. The biplot analysis revealed the contribution of various genotypes to overall diversity, with G26 being the most varied genotype [38]. This investigation allows for the identification of crucial variables and the assessment of genetic variability in wheat breeding lines and breeding programme [39].

### Conclusion

The study showed significant genetic variability among genotypes, which is beneficial for breeding purposes. High heritability values were observed for traits such as DFF, DM, PH, NTP, and SL, making them more responsive to selection



in breeding programs. Key factors that influence the yield and quality of wheat were identified as DM, DM, PH, TSW, and BYP. Seed yield per plant was found to have positive direct effects on several other traits and negative direct effects on DM, NTP, and SL. Genetic diversity analysis revealed the presence of different subgroups of wheat, with the highest genetic diversity observed in Cluster 1 and the most productive subgroups in Clusters III, IV, and IX. Overall, the results of the study provide useful information for wheat breeding programs. The Principal Component Analysis revealed that DFF is the most important variable, with the highest eigenvalue of 3.385. The analysis also revealed that the first three components, DFF, DM and PH, together explain 72.122% of the total variation which can help reduce the dimensionality of the data and identify the most significant variables for each PC.

The biplot analysis showed the relationship among the various traits and genotypes and identified the eight genotypes that largely contributed to the characteristics such as DM, DFF, BYP, PH, and NTP. The study found that DM, HI, BYP, PH, and FF variables had the highest contribution and quality of representation towards the total variability, based on principal components 1 to 5. In terms of genotypes, G26 exhibited the highest percentage of contribution towards the entire diversity, followed by G13, G21, G1, G25, G9, G16, G10, G11, G22, and G3. Through gene editing, the use of cutting-edge technologies like CRISPR-Cas9 may improve the effectiveness of breeding programmes. Managing and analyzing the massive volume of genetic data in the big data timings for advanced computational methods and cross-disciplinary cooperation. The application of artificial intelligence and machine learning to breeding programmes has the potential to revolutionize the speed and accuracy of trait selection as we continue to solve the mysteries of wheat genetics. By tackling these issues, agricultural researchers will be able to push the limits of scientific innovation while also maintaining global food security.

## Acknowledgements

The authors are highly thankful to the Department of Genetics and Plant Breeding, School of Agriculture, Lovely Professional University, Phagwara, Punjab, India for providing all the necessary facilities for conducting the research work.

## Conflict of Interest

None.

## References

- Hajjar R, Jarvis DI, Gemmill-Herren B. 2008. The utility of crop genetic diversity in maintaining ecosystem services. *Agric Ecosyst Environ* 123(4): 261-270. <https://doi.org/10.1016/j.agee.2007.08.003>
- Ehdaie B, Waines JG. 1989. Genetic variation, heritability, and path-analysis in landraces of bread wheat from southwestern Iran. *Euphytica* 41: 183-190. <https://doi.org/10.1007/BF00021584>
- Fragley N, Gardner KA, Bentley AR, Howell P, Mackay IJ, et al. 2023. Multi-trait ensemble genomic prediction and simulations of recurrent selection highlight importance of complex trait genetic architecture for long-term genetic gains in wheat. *In silico Plants* 5(1): 002. <https://doi.org/10.1093/insilicoplants/diad002>
- Welch RM, Graham RD. 2002. Breeding Crops for Enhanced Micronutrient Content. In Adu-Gyamfi JJ (ed) *Food Security in Nutrient-Stressed Environments: Exploiting Plants' Genetic Capabilities*. Developments in Plant and Soil Sciences, Springer, Dordrecht, pp 267-276.
- Dragov R, Uhr Z, Dimitrov E. 2022. Genetic variability, heritability and genetic advance for important quantitative traits of durum wheat: Part I. *Bulg J Agric Sci* 28(4): 691-698.
- Hinson PO, Adams CB, Dong X, Xue Q, Thapa S, et al. 2022. Path analysis of phenotypic factors associated with grain protein in dryland winter wheat. *J Crop Improv* 36(6): 892-918. <https://doi.org/10.1080/15427528.2022.2042882>
- Pirzado AA, Sutahar SK, Sutahar V, Qurashi, NA, Jatoi IK, et al. 2021. Checking the significance of correlation coefficient from the regression analysis using wheat yield. *J Appl Res Plant Sci* 2(2): 132-141. <https://doi.org/10.38211/joarps.2021.2.2.17>
- Khare V. 2022. Multivariate analysis and role of direct-indirect effect for yield and its component traits in bread wheat (*Triticum aestivum* L.). *Electron J Plant Breed* 13(2): 447-454. <https://doi.org/10.37992/2022.1302.045>
- Khalid A, Hameed A, Tahir MF. 2023. Estimation of genetic divergence in wheat genotypes based on agro-morphological traits through agglomerative hierarchical clustering and principal component analysis. *Cereal Res Commun* 51(1): 217-224. <https://doi.org/10.1007/s42976-022-00287-w>
- Babar M, Khan A, Khan N, Akbar F, Ullah N, et al. 2023. Evaluation of wheat genotypes based on morphological characters in bread wheat. *Pure and Appl Biol* 12(1): 219-228. <http://doi.org/10.19045/bspab.2023.120024>
- Ortiz R, Braun HJ, Crossa J, Crouch JH, Davenport G, et al. 2008. Wheat genetic resources enhancement by the international maize and wheat improvement center (CIMMYT). *Genet Resour Crop Evol* 55: 1095-1140. <https://doi.org/10.1007/s10722-008-9372-4>
- Bohra A, Kilian B, Sivasankar S, Caccamo M, et al. 2022. Reap the crop wild relatives for breeding future crops. *Trends Biotechnol* 40(4): 412-431. <https://doi.org/10.1016/j.tibtech.2021.08.009>
- Mahpara S, Bashir MS, Ullah R, Bilal M, Kausar S, et al. 2022. Field screening of diverse wheat germplasm for determining their adaptability to semi-arid climatic conditions. *Plos one* 17(3): 0265344. <https://doi.org/10.1371/journal.pone.0265344>
- Package 'Agricolae': Statistical Procedures for Agricultural Research.
- Debsharma SK, Syed M, Ali M, Maniruzzaman S, Roy PR, et al. 2023. Harnessing on genetic variability and diversity of rice (*Oryza sativa* L.) Genotypes based on quantitative and qualitative traits for desirable crossing materials. *Genes* 14(1): 10. <https://doi.org/10.3390/genes14010010>
- Olivoto T, Lúcio ADC. 2020. Metan: an R package for multi-environment trial analysis. *Methods Ecol Evol* 11(6): 783-789. <https://doi.org/10.1111/2041-210X.13384>
- Farshadfar E, Sutka J. 2003. Multivariate analysis of drought tolerance in wheat substitution lines. *Cereal Res Commun* 31: 33-40. <https://doi.org/10.1007/BF03543247>
- Semahegn Y, Shimelis H, Laing M, Mathew I. 2021. Genetic variability and association of yield and yield components among bread wheat genotypes under drought-stressed conditions. *Aust J Crop Sci* 15(6): 863-870. <https://doi.org/10.21475/ajcs.21.15.06.p2987>
- Haydar FMA, Ahamed MS, Siddique AB, Uddin GM, Biswas KL, et al. 2020. Estimation of genetic variability, heritability and correlation for some quantitative traits in wheat (*Triticum aestivum* L.). *J Bio-Sci* 28: 81-86. <https://doi.org/10.3329/jbs.v28i0.44713>
- Lamara A, Fellahi ZEA, Hannachi A, Benniou R. 2022. Assessing the phenotypic variation, heritability and genetic advance in bread wheat (*Triticum aestivum* L.) candidate lines grown under rainfed semi-arid region of Algeria. *Rev Fac Nac Agron Medellín* 75(3): 1-12. [https://doi.org/10.1007/978-3-319-22000-0\\_1](https://doi.org/10.1007/978-3-319-22000-0_1)

- org/10.15446/rfnam.v75n3.100638
21. Ahmed HGMD, Zeng Y, Shah AN, Yar MM, Ullah A, et al. 2022. Conferring of drought tolerance in wheat (*Triticum aestivum* L.) genotypes using seedling indices. *Front Plant Sci* 13: 961049. <https://doi.org/10.3389/fpls.2022.961049>
  22. Barry CJS, Walker VM, Cheesman R, Smith GD, Morris TT, et al. 2023. How to estimate heritability: a guide for genetic epidemiologists. *Int J Epidemiol* 52(2): 624–632. <https://doi.org/10.1093/ije/dyac224>
  23. Dvivedi RS, Singh B, Pathak V, Verma S, Pandey K. 2023. Genetic variability, character association and component analysis in wheat: genetic variability in wheat. *J AgriSearch* 10(3): 151–157.
  24. Atsbeha G, Mekonnen T, Kebede M, Haileselassie T, Tesfaye K. 2023. Multivariate analyses, heritability and genotype environment interaction of bread wheat genotypes in Ethiopia. *Ecol Genet Genom* 29: 100209. <https://doi.org/10.1016/j.egg.2023.100209>
  25. Huertas-García AB, Tabbita F, Alvarez JB, Sillero JC, Ibba MI, et al. 2023. Genetic variability for grain components related to nutritional quality in spelt and common wheat. *J Agric Food Chem* 71(28): 10598–10606. <https://doi.org/10.1021/acs.jafc.3c02365>
  26. Balkan A. 2018. Genetic variability, heritability and genetic advance for yield and quality traits in M2–4 generations of bread wheat (*Triticum aestivum* L.) genotypes. *Turkish J Field Crop* 23(2): 173–179. <https://doi.org/10.17557/tjfc.485605>
  27. Rehman AU, Khan MF, Abbas A, Javed, M, Abbas MZ, et al. 2023. Evaluation of genetic variability and heritability of wheat genotypes under late sowing effects. *Biol Clin Sci Res J* 2023(1): 268. <https://doi.org/10.54112/bcsrj.v2023i1.268>
  28. Jaiswal R, Gaur SC, Jaiswal SK, Kumar A. 2020. An estimate of variability, heritability and genetic advance for grain yield and yield components in bread wheat (*Triticum aestivum* L.). *Curr Appl Sci Technol* 39(12): 1–6. <https://doi.org/10.9734/CJAST/2020/v39i1230657>
  29. Shankar S, Singh G, Prasad BK. 2023. Estimation of genetic variability, correlation and path coefficient analysis for yield and its attributes in wheat (*Triticum aestivum* L.). *Int J Agric Sci* 8: 272–281.
  30. Thapa RS, Sharma PK, Kumar A, Singh T, Pratap D. 2019. Assessment of genetic variability, heritability and genetic advance in wheat (*Triticum aestivum* L.) genotypes under normal and heat stress environment. *Indian J Agric Res* 53(1): 51–56. <http://doi.org/10.18805/IJAR.E.A-5095>
  31. Abo-Elwafa A, Taib ABA, Abo-Sapra HM, Bakheit BR. 2023. Path coefficient analysis for grain yield and some of its attributes in bread wheat (*Triticum aestivum* L.). *Assiut J Agric Sci* 54(1): 1–18. <http://doi.org/10.21608/ajas.2023.183395.1216>
  32. Baye A, Berihun B, Bantayehu M, Derebe B. 2020. Genotypic and phenotypic correlation and path coefficient analysis for yield and yield-related traits in advanced bread wheat (*Triticum aestivum* L.) lines. *Cogent Food Agric* 6(1): 1752603. <https://doi.org/10.1080/23311932.2020.1752603>
  33. Khanal D, Thapa DB, Dhakal KH, Pandey MP, Kandel BP. 2020. Correlation and path coefficient analysis of elite spring wheat lines developed for high temperature tolerance. *Environ Ecosys Sci* 4(2): 73–76. <http://doi.org/10.26480/ees.02.2020.73.76>
  34. Laala Z, Oulmi A, Fellahi ZEA, Benmahammed A. 2021. Studies on the nature of relationships between grain yield and yield-related traits in durum wheat (*Triticum durum* Desf.) populations. *Rev Fac Nac Agron Medellin* 74(3): 9631–9642. <https://doi.org/10.15446/rfnam.v74n3.92488>
  35. Sharma P, Kamboj MC, Singh N, Chand M, Yadava RK. 2018. Path coefficient and correlation studies of yield and yield associated traits in advanced homozygous lines of bread wheat germplasm. *Int J Curr Microbiol App Sci* 7(2): 51–63. <https://doi.org/10.20546/ijcmas.2018.702.008>
  36. Renu O, Aditi S, Asmita A, Rahul KC, Sabina T, et al. 2018. Correlation and path coefficient analysis of wheat (*Triticum aestivum* L.) genotypes. *Farm Manage* 3(2):136–141. <http://dx.doi.org/10.31830/2456-8724.2018.0002.19>
  37. Shyam C, Chandrakar PK, Rastogi NK, Banjare U. 2018. Evaluation of genetic divergence analysis in wheat for yield and its component characters. *Int J Agric Environ Biotechnol* 11(6): 829–834.
  38. Divya P, Panwar IS, Saharan RP, Vikram S, Anuradha G. 2017. Genetic diversity and association studies for yield attributing traits in bread wheat [*Triticum aestivum* (L.) em. Thell]. *Res on Crop* 18(1): 139–144. <http://dx.doi.org/10.5958/2348-7542.2017.00024.9>
  39. Nayana BM, Kumar KR, Chesneau C. 2022. Wheat yield prediction in india using principal component analysis-multivariate adaptive regression splines (PCA-MARS). *Agri Eng* 4(2): 461–474. <https://doi.org/10.3390/agriengineering4020030>