

EVALUATION OF GENETIC DIVERSITY IN WHEAT (*TRITICUM AESTIVUM* L.) USING MORPHOLOGICAL TRAITS

Arya V.¹, Priyanka^{1*}, Singh A.¹, Vikram P.², Yadav S.¹ and Kataria P.¹

¹Faculty of Agricultural Sciences, SGT University, Gurugram, Haryana, India

²Division of Seed Science and Technology, Indian Agricultural Research Institute, New Delhi, India

Corresponding author: Priyanka, E-mail: guleriapriyanka033@gmail.com

ABSTRACT

Genetic diversity plays a vital role in wheat improvement by providing the variability required for selection and the development of superior cultivars. The present study was conducted to assess the genetic diversity among 60 wheat (*Triticum aestivum* L.) genotypes using agro-morphological traits and multivariate statistical approaches. The genotypes were evaluated under an augmented block design at the Research Farm of the Faculty of Agricultural Sciences, SGT University, Gurugram, Haryana. Data were recorded for key agronomic traits and analyzed using analysis of variance, correlation analysis, path coefficient analysis, K-means clustering, hierarchical clustering, heatmap visualization and multidimensional scaling (MDS). Analysis of variance revealed significant variability for leaf rust response, tiller number, and weight per spike, indicating the presence of exploitable genetic variation within the germplasm. Correlation analysis showed strong positive associations of grain yield with tiller number ($r = 0.867$) and weight per spike ($r = 0.865$). Path coefficient analysis further identified these traits as the major direct contributors to grain yield, with direct effects of 0.528 and 0.536, respectively. Multivariate analyses consistently grouped the genotypes into two distinct clusters, demonstrating substantial genetic divergence among the evaluated materials. Heatmap and MDS analyses validated the clustering pattern and identified several genetically unique genotypes. Genotypes such as HI1531, MP3224, PBW590, Kharchia65, WH1021, IC42, NP866, and GW451 were identified as promising parental lines for future breeding programmes. The results demonstrate that agro-morphological characterization combined with multivariate analysis is an effective approach for identifying genetically diverse and high-performing wheat genotypes for crop improvement.

KEYWORDS: Agro-morphological traits; Cluster analysis; Genetic diversity; Grain yield; Multidimensional scaling; Path coefficient analysis; *Triticum aestivum*; Wheat breeding.

INTRODUCTION

Wheat (*Triticum aestivum* L.) is one of the most important cereal crops in the world and serves as a staple food for a substantial proportion of the global population. As a member of the family Poaceae, wheat occupies a central position in global agriculture owing to its wide adaptability, high productivity, ease of cultivation, and diverse end-use applications. Bread wheat is a hexaploid species ($2n = 6x = 42$; AABBDD genome) and accounts for nearly 90–95% of global wheat production (Dubcovsky and Dvorak, 2007; Shewry and Hey, 2015). Wheat is a staple food and a source of protein for people globally, accounting for around 20% of the calories and protein eaten by humans (Shewry and Hey, 2015). Wheat is essential to feed the world's population, which has been increasing at a rapid pace, with recent years seeing production surpass 780 million tonnes globally, according to the Food and Agriculture Organization (FAO, 2024).

The demand for wheat is projected to grow significantly over the next several decades, driven by fast population growth, urbanization, changing food habits and pressure on food production systems. It is expected that the world population will reach 10 billion by 2050, while the demand for food is expected to rise accordingly to guarantee food security in the future (United Nations, 2022). But, climate change, unpredictable rainfall, higher temperatures, falling soil fertility, water shortages, and the spread of new pests and diseases are all becoming an increasing problem for wheat production (Tester and Langridge, 2010; Xiong and Reynolds, 2024; Lopes et al., 2015). In such a scenario, generation of high yielding, climate resilient and genetically diverse wheat cultivars has emerged as one of the priority areas in modern wheat breeding programmes.

Genetic diversity is the basic source of crop improvement and the raw material needed for creation of improved cultivars. Expanding wheat's genetic diversity is critical for continued genetic improvement and increased resilience to new challenges and changing environments (Reynolds and Braun, 2022). The effectiveness of selection within a breeding population and the potential to improve economically important traits is dependent on the level of genetic variability in the population (Hill and Mackay, 2004; Govindaraj et al., 2015; Fu, 2015). In the past, wheat breeding efforts have effectively increased productivity by the use of wide range of germplasm

and introgression of beneficial alleles for yield, quality and stress tolerance. However, the limited genetic diversity of the few elite varieties that are planted has led to a narrowing of the genetic base of cultivated wheat over time, with concerns about genetic erosion and increased susceptibility to environmental stressors and disease epidemics (Reif et al., 2005; Fu, 2015).

There are recent studies showing that modern wheat cultivars have a limited genetic diversity within the genus *Triticum* and its related species (Lopes et al., 2015). Much of the value alleles are still conserved in landraces and traditional cultivars, as well as other wild relatives that have been poorly considered in current breeding programmes (Tanksley and McCouch, 1997; Zhang et al., 2017; Rasheed et al., 2018). The benefits of utilising these underused genetic resources lie in their unique adaptive characteristics and as reservoirs of genes for biotic and abiotic stress resistance, grain quality improvement and yield enhancement. Hence, systematic characterization and assessment of germplasm collections is crucial to expanding the genetic diversity of cultivated wheat and to sustainable genetic improvement (Jaradat, 2013; Broccanello et al., 2023).

Genetic diversity can be estimated with morphological, biochemical and molecular methods. In these, agro-morphological characterization is one of the most practical, economical and universally applicable methods to assess phenotypic variation between genotypes (Mohammadi & Prasanna, 2003). Morphological traits are the manifestation of genetic and environmental factors and give direct information about agronomic performance, adaptation and yield potential. Days to flowering, days to maturity, plant height, productive tillers, spike length, spikelets per spike, grains per spike, grain weight and harvest index are extensively used traits to assess diversity and characterization of wheat varieties (Govindaraj et al., 2015). Additionally, morphological characterization can be useful as an initial screening step to find potential genotypes for further molecular investigations.

Multivariate statistical analyses methods are excellent tools that allow the simultaneous analysis of multiple traits, which have significantly improved the evaluation of genetic diversity. Many of the traditional univariate analyses used to assess agronomic characters do not reflect the interactions between characters and can underestimate the diversity found in a germplasm collection. These multivariate tools like cluster, principal component analysis (PCA), multidimensional scaling (MDS) and heatmap visualization have therefore become essential for germplasm characterization and diversity assessment (Crossa and Franco, 2004; Govindaraj et al., 2015). The following methods are useful for identifying genetically divergent genotypes, uncovering cryptic variation and for selecting appropriate hybridisation combinations.

Wheat genotypes have been extensively used for classification into groups or clusters by cluster analysis, using overall similarity or divergence. Genotypes in distant clusters are frequently selected as good potential parents for crossing because they are likely to produce a wide range of variability and transgressive segregants in their subsequent generations (Mohammadi and Prasanna, 2003). Likewise, multidimensional scaling can be used to display the genetic similarity/dissimilarity between the different genotypes and it can be used to visualize the patterns of genetic diversity that may not be observed using traditional statistical tools (Borg and Groenen, 2005). Heatmap analysis has become a very effective visualization tool that not only provides an overview of genotype clustering but also of the performance of these traits, which will help identify superior genotypes and important agronomic traits (Wilkinson and Friendly, 2009).

In addition to diversity assessment, knowledge of the relationships between yield and its component traits is also crucial for successful crop improvement. Grain yield is a polygenic character influenced by many interacting morphological and physiological characters. Selection for grain yield alone is sometimes difficult due to the low heritability and high environmental genetic effects (Hill and Mackay, 2004). Breeders have often used associated traits which are directly and indirectly involved in yield improvement. Correlation analysis gives useful information about the size and the direction of association between the characters and helps in determining which of them can be used as reliable selection characters for breeding program (Singh and Chaudhary, 1985). Correlation coefficients, however, cannot tell direct and indirect effects of personal characteristics.

In order to overcome this drawback, path coefficient analysis has been extensively applied in wheat breeding program. This method separates yield-related traits into direct and indirect effects and gives a more complete picture of the causal relationships between the traits (Dewey and Lu, 1959). A number of investigators have found that productive tillers, grains per spike, biological yield, harvest index and thousand-grain weight have significant direct effects on grain yield and thus could serve as useful selection criteria in wheat improvement programmes (Mohammadi et al., 2012; Kumar et al., 2024). Correlation and path analyses can be used together to pinpoint important yield components and develop better breeding programs.

Although the genetic diversity of wheat has been the subject of a number of studies, the ongoing losses of genetic resources, and the need for cultivars that are more resilient to climate change more broadly, require the assessment of more collections of wheat germplasm with comprehensive analytical approaches. Moreover, few studies combined morphological characteristic with cluster analysis, MDS visualization, heatmap analysis, correlation studies, and path coefficient analysis in one study. These co-ordinated efforts can give an integrated view of genetic variation and aid in the development of better breeding stocks in the future. The current study was conducted to assess the genetic differences between 60 wheat genotypes based on 11 agro-morphological characters in field environment.

The study was aimed at evaluating phenotypic variability, grouping genotypes into divergent categories by multivariate statistical tools, finding better genotypes with desired agronomic traits, and elucidating the role played by morphological characters in increasing the grain yield by correlation and path coefficient analysis.

Results from the study should offer information for wheat germplasm utilization, parent selection and high yielding, genetically diverse wheat cultivars.

MATERIALS AND METHODS

Plant material and experimental site

The present study was conducted during the rabi season at Kaliawas Research Farm, Gurugram, Haryana, India. The experimental material comprised 60 bread wheat (*Triticum aestivum* L.) genotypes, including 57 test entries and three standard check varieties (HD 3271, HD 3086, and HD 3237). The genotypes represented a diverse collection of wheat germplasm obtained from different breeding programmes and were evaluated for genetic variability, diversity, and yield-related traits.

Experimental design and crop management

The experiment was laid out in an Augmented Randomized Block Design (ARBD) following the procedure of Federer (1956). The augmented design was adopted because replication of all test entries was impractical due to the large number of genotypes under evaluation. The three standard check varieties were repeated across all blocks to estimate experimental error and adjust genotype means. Each genotype was grown in a single-row plot under recommended agronomic practices for wheat cultivation. Uniform crop management practices, including irrigation, fertilization, weed control, and plant protection measures, were followed throughout the growing season to minimize environmental variation.

Recording of agro-morphological traits

To assess the extent of morphological variation among the wheat genotypes, a total of eleven agro-morphological traits representing phenological, vegetative, spike-related, and yield-related characteristics were recorded during the field experiment. These traits were selected because they are widely used for evaluating genetic diversity and identifying superior genotypes in wheat breeding programmes. The details of the recorded traits, their abbreviations, and units of measurement are presented in Table 1.

Plant height was measured from the soil surface to the tip of the spike excluding awns. Spike length was recorded from the base to the tip of the spike excluding awns. Thousand grain weight was determined by weighing 1000 randomly selected grains using an electronic balance. Harvest index was calculated as the ratio of grain yield to biological yield and expressed as a percentage.

Table 1. Agro-morphological traits recorded for the evaluation of wheat (*Triticum aestivum* L.) genotypes under field conditions.

Category	Trait	Abbreviation	Unit
Phenological traits	Days to 50% flowering	DFF	Days
	Days to maturity	DM	Days
Vegetative traits	Plant height	PH	cm
	Effective tillers per plant	TN	Number
Spike-related traits	Spike length	SL	cm
	Number of spikelets per spike	NSPS	Number
	Number of grains per spike	GPS	Number
Yield-related traits	Thousand grain weight	TGW	g
	Weight per spike	WPS	g
	Grain yield per plant	GY	g
	Harvest index	HI	%

Statistical analysis

Analysis of variance

Analysis of variance (ANOVA) appropriate for the augmented design was performed to determine the existence of significant differences among wheat genotypes for the studied traits. Adjusted means were estimated using the repeated check varieties as controls for block effects.

Correlation analysis

Pearson's correlation coefficients were computed among all agro-morphological traits to determine the magnitude and direction of associations between grain yield and its component characters. The analysis facilitated identification of traits exhibiting significant relationships with grain yield and other agronomic traits.

Path coefficient analysis

Path coefficient analysis was conducted according to the method of Dewey and Lu (1959) to partition correlation coefficients into direct and indirect effects. This analysis enabled identification of traits contributing directly to grain yield and those influencing yield indirectly through other associated characters.

Cluster analysis

Hierarchical cluster analysis was performed using standardized trait data to classify genotypes into distinct groups based on their overall similarity and genetic divergence. Euclidean distance matrices were used to assess dissimilarities among genotypes, and clustering patterns were visualized through dendrogram construction.

K-means clustering

K-means clustering was employed as a complementary multivariate approach to classify wheat genotypes into homogeneous groups. The analysis facilitated the identification of genetically distinct clusters and provided additional insights into diversity patterns within the germplasm collection.

Multidimensional scaling (MDS)

Multidimensional Scaling (MDS) analysis was performed to visualize genetic relationships among genotypes in reduced-dimensional space. The technique provided a graphical representation of similarities and divergences among genotypes and served as a validation tool for cluster analysis results.

Heatmap analysis

A heatmap integrated with hierarchical clustering was generated using standardized trait values to visualize variation among genotypes and relationships among traits simultaneously. The heatmap facilitated the identification of superior genotypes and trait-specific performance patterns within the germplasm collection.

Software

All statistical analyses were performed using R statistical software (R Core Team, Vienna, Austria). Graphical outputs including cluster dendrograms, multidimensional scaling plots, correlation matrices, path diagrams, and heatmaps were generated using appropriate R packages.

RESULTS AND DISCUSSION

Analysis of variance

Analysis of variance (ANOVA) under the augmented block design revealed significant differences among genotypes for several agro-morphological traits, indicating the presence of exploitable genetic variability within the evaluated wheat germplasm. Significant differences between checks and test entries were observed for leaf rust score ($F = 10.74$, $p = 0.008$), tiller number ($F = 7.04$, $p = 0.024$), and weight per spike ($F = 5.85$, $p = 0.036$), demonstrating substantial divergence between the breeding material and the standard cultivars (Table 2). In addition, significant variation among test genotypes was detected for leaf rust ($F = 3.39$, $p = 0.020$) and tiller number ($F = 3.51$, $p = 0.018$), shows the existence of useful genetic diversity for these traits.

Table 2. Summary of analysis of variance (ANOVA) showing significant and non-significant differences among wheat (*Triticum aestivum* L.) genotypes for selected agro-morphological traits evaluated under an Augmented Randomized Block Design (ARBD).

Trait	Source of Variation	F-value	P-value	Interpretation
Leaf rust (LR)	Checks vs. Tests	10.74	0.008	Highly significant
Leaf rust (LR)	Among Tests	3.39	0.020	Significant
Tiller number (TN)	Checks vs. Tests	7.04	0.024	Significant
Tiller number (TN)	Among Tests	3.51	0.018	Significant
Weight per spike (WPS)	Checks vs. Tests	5.85	0.036	Significant
DFE, PH, TGW and GY	Among Tests	< 1.65	> 0.10	Non-significant

The significant variation observed for tiller number and leaf rust resistance suggests that these traits possess considerable potential for selection and genetic improvement. Genetic variability for yield-contributing traits and disease resistance is essential for achieving sustained genetic gains in wheat breeding programs, as effective selection depends on the availability of heritable variation within the breeding population (Hill and Mackay, 1996; Mohammadi and Prasanna, 2003). Similar levels of variability for agronomic and adaptive traits have been reported in diverse wheat germplasm collections by Fu (2015), Rasheed et al. (2018), and Lopes et al. (2015), who emphasized the importance of exploiting genetic diversity for enhancing breeding efficiency and resilience. The significant differences observed for weight per spike further indicate opportunities for improving grain

productivity through indirect selection, as spike-related traits have frequently been identified as major contributors to grain yield in wheat (Mohammadi et al., 2012; Acquaah, 2020). In contrast, days to 50% flowering, plant height, thousand grain weight, and grain yield exhibited non-significant variation among test genotypes ($p > 0.10$), suggesting relatively lower genetic differentiation or greater environmental buffering for these traits within the studied germplasm. Similar observations have been reported in wheat populations with relatively narrow genetic backgrounds, where intensive selection and cultivar development have reduced variability for major agronomic traits (Reif et al., 2005; Lopes et al., 2015; Fu, 2015). Environmental effects may further obscure the expression of complex quantitative traits such as grain yield, thereby reducing the detection of significant genotypic differences under field conditions (Blum, 2011). Nevertheless, the presence of significant variation for leaf rust resistance, tiller number, and weight per spike demonstrates that the evaluated germplasm retains sufficient diversity to support future breeding efforts. Traits showing non-significant variation among test genotypes (DFF, PH, TGW, and GY) suggest relatively lower genetic differentiation or greater environmental influence on their expression within the evaluated germplasm. Consequently, breeding efforts may be more effectively directed toward traits such as leaf rust resistance (LR), tiller number (TN), and weight per spike (WPS), which exhibited significant genetic variation and greater potential for selection.

Correlation analysis

Pearson's correlation analysis revealed considerable variation in the association of agro-morphological traits with grain yield (Figure 1; Table 3). Effective tillers per plant (TN) exhibited the strongest positive correlation with grain yield ($r = 0.867$), followed closely by weight per spike (WPS) ($r = 0.865$). A moderate positive correlation was observed between TN and WPS ($r = 0.590$), indicating that genotypes with more productive tillers tended to produce heavier spikes. In contrast, days to 50% flowering (DFF) and thousand grain weight (TGW) showed negative correlations with grain yield ($r = -0.315$ and -0.394 , respectively). Plant height (PH), grains per spike (GNPS), and leaf rust response (LR) exhibited weak positive correlations with grain yield.

Table 3. Pearson's correlation coefficients showing the relationship between agro-morphological traits and grain yield in wheat (*Triticum aestivum* L.) genotypes.

Trait	Correlation with Grain Yield (r)	Direction of Association
Effective tillers per plant (TN)	+0.87	Strong positive
Weight per spike (WPS)	+0.86	Strong positive
Effective tillers per plant (TN) vs. Weight per spike (WPS)	+0.59	Moderate positive
Days to 50% flowering (DFF)	-0.32	Moderate negative
Thousand grain weight (TGW)	-0.39	Moderate negative
Plant height (PH), Grains per spike (GNPS), and Leaf rust response (LR)	$r \approx +0.09$ to $+0.14$	Weak positive

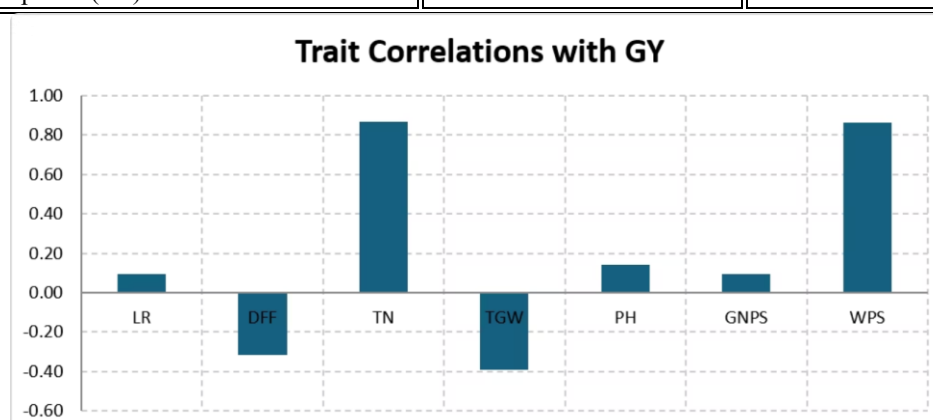


Figure 1. Pearson's correlation coefficients of selected agro-morphological traits with grain yield (GY) in wheat.

The strong positive correlations of effective tillers per plant and weight per spike with grain yield suggest that these traits are important components influencing wheat productivity. Similar relationships have been reported by Mohammadi et al. (2012), who identified spike-related traits as major contributors to grain yield. The positive association between tiller number and spike weight further indicates that simultaneous improvement of these traits

could enhance yield potential. Reynolds et al. (2009) also emphasized that improvements in productive tillers and spike productivity are essential for increasing wheat yield.

The negative association of days to 50% flowering with grain yield suggests that relatively early-flowering genotypes may be advantageous under the experimental conditions, possibly by escaping terminal stress. Comparable observations were reported by Lopes et al. (2015) in wheat adapted to stress-prone environments. The weak correlations of plant height, grains per spike, and leaf rust response indicate that these traits contributed less directly to yield variation in the present germplasm. Overall, the correlation analysis identified effective tillers per plant and weight per spike as the most informative traits for indirect selection in wheat breeding programmes.

Path coefficient analysis

Path coefficient analysis partitioned the correlation coefficients into direct and indirect effects on grain yield (Figure 2; Table 4). Among the evaluated traits, weight per spike (WPS) exhibited the highest positive direct effect (0.536), followed by effective tillers per plant (TN) (0.528). Both traits also showed strong positive correlations with grain yield ($r = 0.865$ and 0.867 , respectively). In contrast, days to 50% flowering (DFF) and thousand grain weight (TGW) exhibited negative direct effects and negative correlations with grain yield. The coefficient of determination ($R^2 = 0.951$) indicated that the evaluated traits explained 95.1% of the variation in grain yield, while the residual effect (0.222) suggested a relatively small contribution of unmeasured factors.

Table 4. Direct and indirect effects of agro-morphological traits on grain yield in wheat (*Triticum aestivum* L.) estimated through path coefficient analysis.

Trait	Correlation with Grain Yield (r)	Direct Effect	Indirect Effect	Selection Priority
Effective tillers per plant (TN)	+0.867	+0.528	+0.339	★★★
Weight per spike (WPS)	+0.865	+0.536	+0.329	★★★
Days to 50% flowering (DFF)	−0.315	−0.059	−0.256	★☆☆
Thousand grain weight (TGW)	−0.394	−0.021	−0.373	★☆☆
Plant height (PH)	—	−0.019	—	☆☆☆
Grains per spike (GNPS)	—	+0.005	—	☆☆☆
Leaf rust response (LR)	—	+0.044	—	☆☆☆

The high positive direct effects of weight per spike and effective tillers per plant indicate that these traits are the major determinants of grain yield in the evaluated wheat germplasm. Similar findings were reported by Mohammadi et al. (2012), who identified spike-related traits as important contributors to grain yield. The strong agreement between correlation and path coefficient analyses further supports the use of these traits as reliable selection criteria. Reynolds et al. (2009) also emphasized that improvements in spike productivity are essential for increasing wheat yield potential. The high coefficient of determination suggests that the selected agro-morphological traits adequately explained the observed variation in grain yield, consistent with the findings of Dewey and Lu (1959), who demonstrated the usefulness of path coefficient analysis in identifying key yield components.

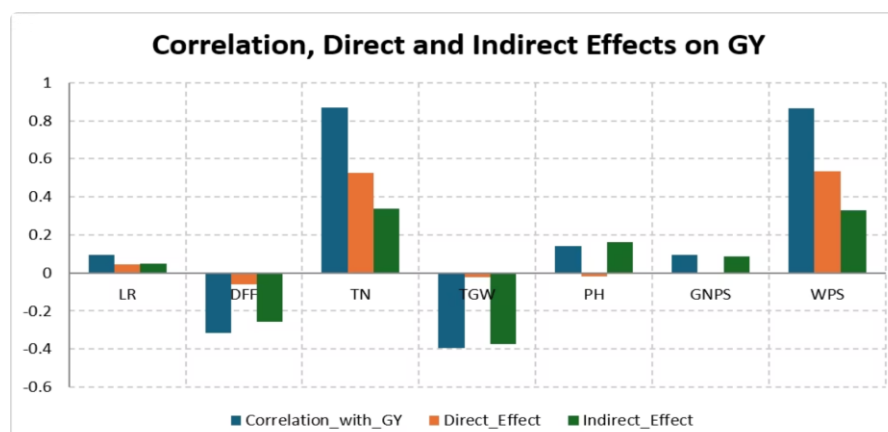


Figure 2. Comparison of Pearson's correlation coefficients and direct and indirect effects of agro-morphological traits on grain yield obtained through path coefficient analysis in wheat genotypes

The coefficient of determination ($R^2 = 0.343$) indicated that the evaluated traits collectively explained approximately 34.3% of the variation in grain yield. These findings suggest that additional physiological, biochemical, or environmental factors may contribute substantially to yield performance under saline conditions. Similar observations have been reported by Mohammadi et al. (2012) and Khan and Naqvi (2012); Kumar et al. (2023), who demonstrated that grain yield in wheat is a complex quantitative trait influenced by multiple yield components and environmental interactions. Furthermore, Kannaujiya et al. (2020) reported that the number of grains per spike exerts a positive direct effect on grain yield and can therefore be used as an effective selection criterion in wheat breeding programmes.

K-means clustering analysis

K-means clustering classified the 60 wheat genotypes into two distinct groups, with the optimal number of clusters determined as $K = 2$ based on the highest silhouette score (0.298). Although the silhouette value indicated moderate separation, it was substantially higher than those obtained for $K = 3$ to $K = 8$, suggesting the presence of two biologically meaningful genetic groups within the germplasm collection (Table 5).

Table 5. Summary of K-means clustering analysis of 60 wheat (*Triticum aestivum* L.) genotypes based on eight agro-morphological traits

Parameter	Value
Number of genotypes	60
Traits used	LR, DFF, TN, TGW, PH, GNPS, WPS and GY
Clustering algorithm	K-means
Initialization method	K-means++
Number of clusters	2
Number of re-runs	10
Maximum iterations	300
Optimal silhouette score	0.298

Cluster I included genotypes such as WH1021, IC42, GW451, NP866, and PBW343, whereas Cluster II comprised HI1531, MP3224, PBW590, HD3090, Kharchia65, and several other genotypes (Table 6). The clustering pattern was largely consistent with the hierarchical dendrogram, indicating the robustness of the genetic structure identified through multivariate analysis. Similar clustering behaviour has been reported in wheat germplasm, where different multivariate approaches, consistently identified genetically distinct groups and supported the selection of diverse parents for hybridization (Mohammadi and Prasanna, 2003; Khodadadi et al., 2011; Ali et al., 2020). The concordance among clustering methods enhances confidence in the classification of genotypes and provides a reliable basis for exploiting genetic diversity in wheat breeding programmes.

Table 6. Characteristics and breeding significance of the two clusters identified through K-means clustering.

Cluster	Representative genotypes	Major characteristics	Breeding significance
Cluster I	WH1021, IC42, GW451, NP866, PBW343 and related entries	Relatively lower grain yield and weight per spike; genetically divergent material	Useful as a source of genetic diversity for hybridization
Cluster II	HI1531, MP3224, PBW590, HD3090, Kharchia65 and related entries	Higher grain yield and weight per spike	Promising source of high-yielding parents for wheat improvement

Comparison of cluster mean values revealed that Cluster II exhibited higher grain yield and weight per spike than Cluster I, indicating the presence of superior agronomic performance within this group. In contrast, Cluster I contributed greater genetic divergence and may serve as a valuable source of novel alleles. Therefore, hybridization between genotypes belonging to different clusters may facilitate the development of broad-based segregating populations and improve the efficiency of selection for yield-related traits.

Hierarchical cluster analysis

Hierarchical clustering grouped the 60 wheat genotypes into two major clusters. Cluster I comprised genotypes such as HI1531, MP3224, PBW590, and Kharchia65, whereas Cluster II included WH1021, NP866, IC42, and GW451. The large inter-cluster distance observed between the two groups indicates substantial genetic divergence among the evaluated genotypes, suggesting that crosses between genotypes from different clusters are likely to

generate greater genetic recombination and transgressive segregants (Figure 3). Such divergent parental combinations are valuable for broadening the genetic base and improving yield and other economically important traits in wheat breeding programmes. Similar observations have been reported by Mohammadi and Prasanna (2003), Reif et al. (2005), Khodadadi et al. (2011), and Rasheed et al. (2018), who demonstrated that cluster analysis is an effective tool for identifying genetically diverse parental lines for hybridization.

Genotypes belonging to different clusters are expected to possess distinct allelic combinations and may produce greater heterosis and wider segregation upon hybridization. Therefore, crosses involving WH1021, NP866, IC42, and GW451 from Cluster II with HI1531, MP3224, PBW590, and Kharchia65 from Cluster I may generate highly variable segregating populations suitable for selecting superior recombinants. The identified divergent genotypes could thus serve as valuable parental resources for future wheat improvement programmes.

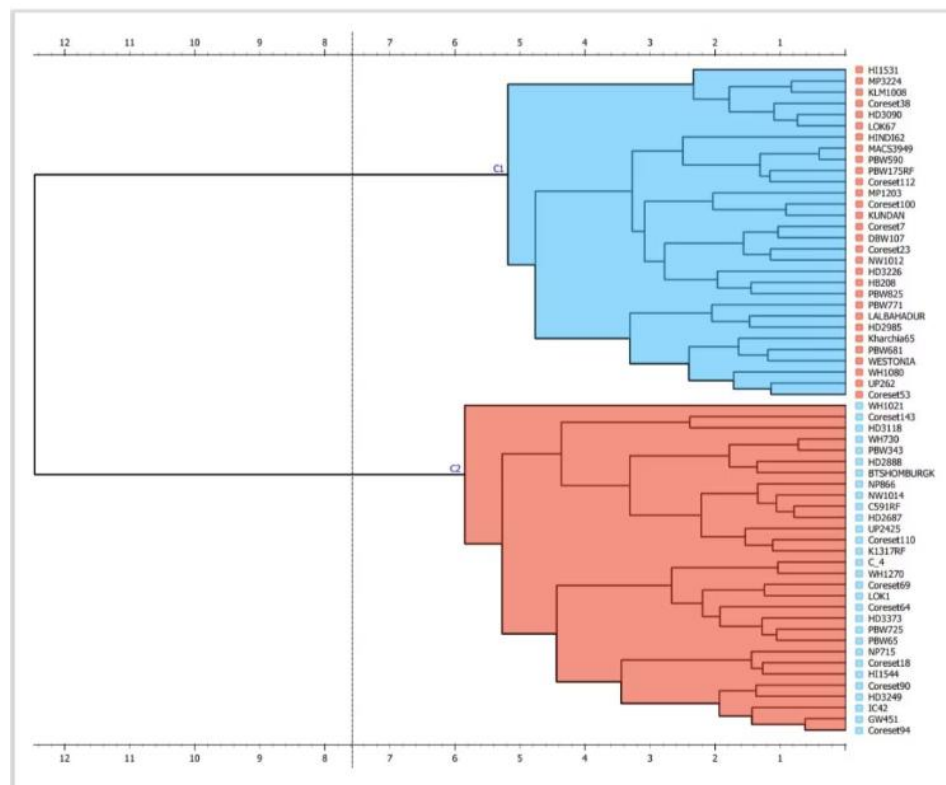


Figure 3. Hierarchical dendrogram illustrating the genetic relationships among 60 wheat (*Triticum aestivum* L.) genotypes based on Euclidean distance using Ward's minimum variance clustering method. Two major clusters (C1 and C2) were identified, indicating the presence of substantial genetic diversity within the evaluated germplasm.

Heatmap analysis

Heat map analysis of the standardized agro-morphological traits showed that the 60 wheat genotype can be grouped into distinct pattern of similarity and dissimilarity (Figure 4). Two distinct genotype groups were well defined and the grouping was very similar to both hierarchical dendrogram and K-means and multidimensional scaling (MDS) analysis. The higher tiller number and weight per spike were grouped together and core-set accessions were spread out in the different clusters, suggesting high phenotypic diversity within the evaluated germplasm. This is a good indication of the robustness of the genetic structure found among the wheat genotypes, as shown by the concordance between the heatmap and other multivariate analyses.

Mohammadi and Prasanna (2003) and Govindaraj et al. (2015) also reported that heatmap-based visualization and cluster analysis can be very useful to determine the genetic diversity and to group the genotypes into distinct clusters. The results of the grouping of the genotypes and the better number of tillers and weight of each tillers is confirmed by the results of correlation and path coefficient analysis, which shows that the traits are important factors in grain yield (Mohammadi et al., 2012; Reynolds et al., 2009). Furthermore, the distribution of core-set accessions in several clusters indicates that the germplasm assessed in this study had a good representation of the genetic diversity available and therefore serves as a valuable resource for selection in breeding programmes and as parents for wheat improvement efforts (Lopes et al., 2015; Rasheed et al., 2018).

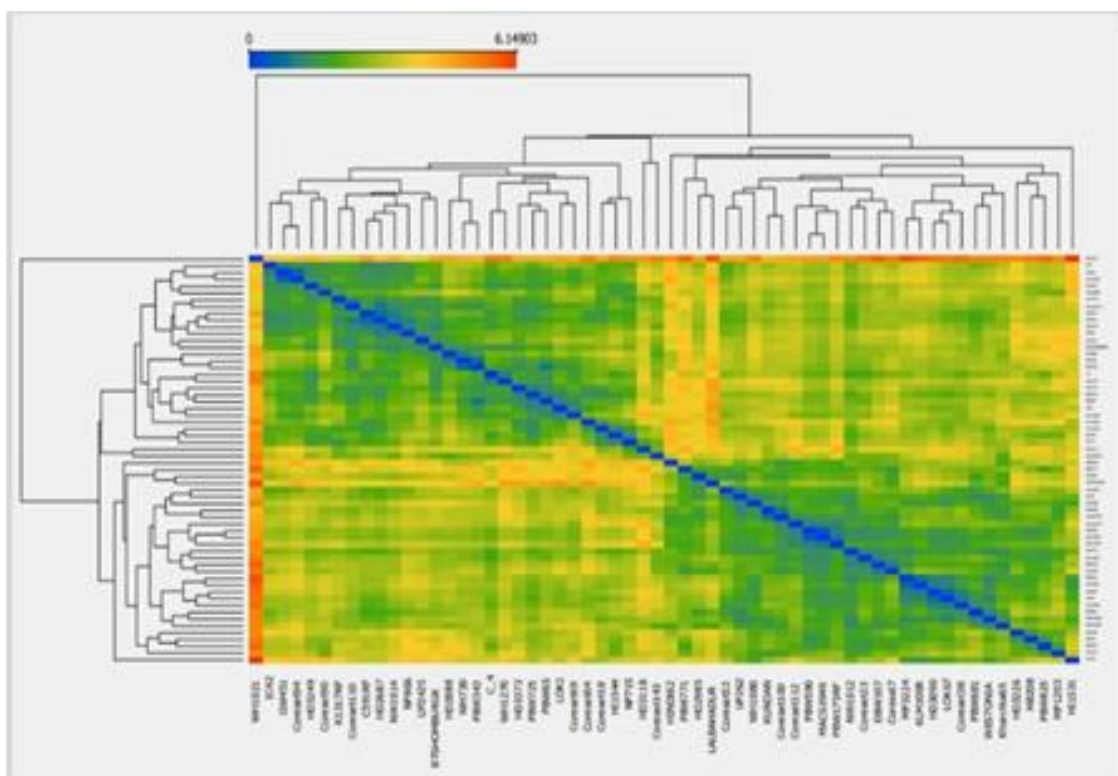


Figure 4. Hierarchical clustered heatmap of the genetic relationships among 60 wheat (*Triticum aestivum* L.) genotypes, based on Euclidean distance, computed from eight agro-morphological traits. Color intensity is a measure of similarity or dissimilarity between the genotypes whereas the clustering pattern shown in the corresponding dendrograms represents a hierarchical grouping.

Multidimensional scaling (MDS) analysis

The 60 wheat genotypes were projected in the two-dimensional space based on agro-morphological similarity using multidimensional scaling (MDS) (Figure 5). The genotypes were distinctly grouped in two clusters (C1 and C2) and there was little overlap between the clusters, suggesting a high degree of phenotypic divergence. Some of the high performing genotypes from Cluster II were HI1531, MP3224, PBW590, HD3090 and Kharchia65 while some of the high performing genotypes from Cluster I were WH1021, NP866, IC42 and GW451. The network connections revealed close relationship between the genotypes within a cluster, and some of the genotypes on the periphery were in isolated positions indicating increased genetic differentness.

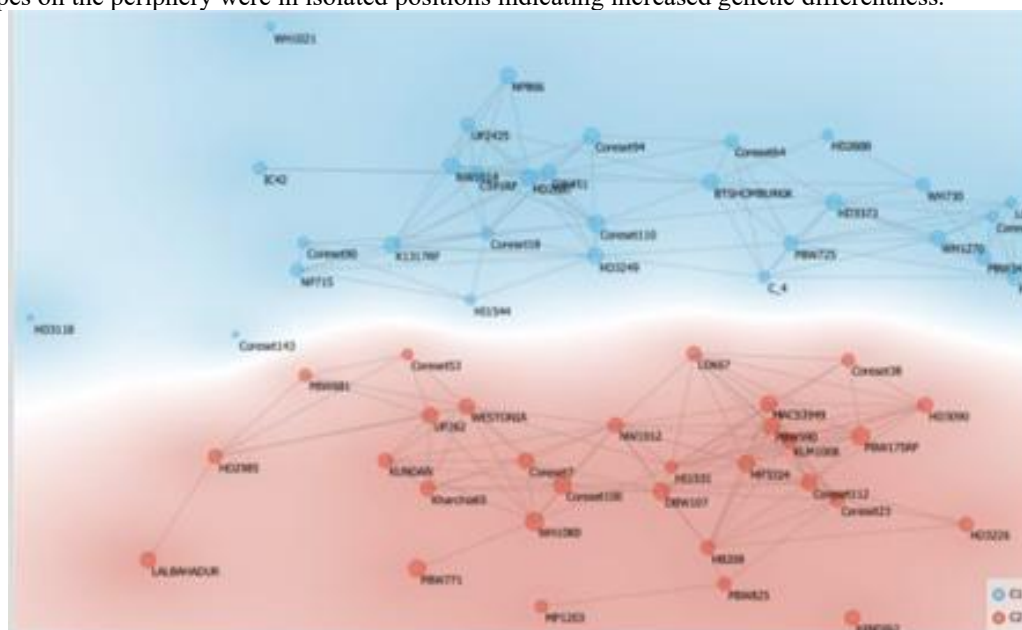


Figure 5. Two dimensional multidimensional scaling (MDS) plot of 60 wheat (*Triticum aestivum* L.) genotypes based on the Euclidean distances obtained by eight agro-morphological traits. The analysis divided the genotypes into two clusters (C1 and C2) suggesting significant phenotypic variation.

The spatial separation of the two clusters is very clear and suggests that the wheat germplasm assessed is quite diverse. The MDS plot indicates a close similarity with the hierarchical clustering, K-means and heatmap analyses, as the grouping observed was consistent across different multivariate methods. Mohammadi and Prasanna (2003) reported similar findings, stating that multivariate techniques were useful for the assessment of genetic diversity in germplasm of crop plants. Borg and Groenen (2005) also showed that MDS can be an efficient method to represent complex similarity relationships in a low-dimensional space, which makes it useful to visualize the genotype divergences. The peripheral position of genotypes like WH1021, HD3118, LALBAHADUR and HINDI62 shows more genetic uniqueness and may be useful as cross pollinators in the hybridization programme. Rasheed et al. (2018) have suggested that similar breeding strategies using genetically divergent parents will be recommended to expand the genetic base and increase genetic gain in wheat improvement programmes.

Cluster validation and trait distribution

Scatter plot and regression analysis

In both scatter plot and regression analysis, grain yield was found to be strongly and positively correlated with the weight of spike indicating a strong positive linear relationship (Figure 6). High spike weight was correlated with high yield ($r = 0.865$) and the path coefficient analysis indicated that weight per spike was a major factor affecting yield; this was confirmed by the genotypic comparisons where spike weight correlated positively with yield, and higher spike weight usually resulted in higher yield. The scatter plot showed that the genotypes in Cluster II are mostly distributed in the upper-right part of the graph, showing that these are more likely to perform well for both traits, while Cluster I are more likely to be distributed in the lower portion of the graph. The higher grain yield and spikelet weight per spike observed in high yielding genotypes, LAL BAHADUR, WH1080, HINDI62, PBW175RF, HI1531 and MP3224 indicated that they would be good donor parents for wheat improvement.

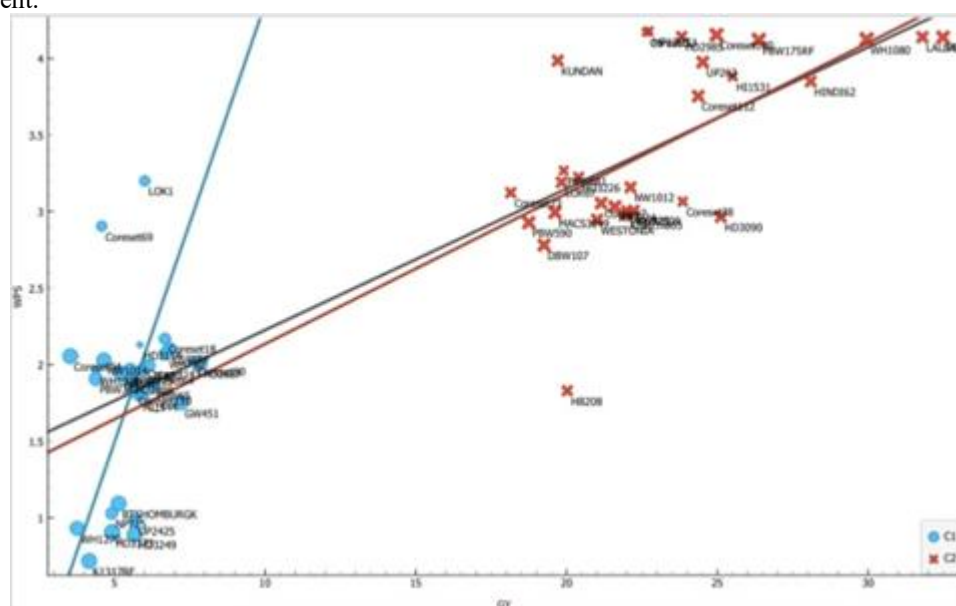


Figure 6. Scatter plot with regression lines illustrating the distribution of 60 wheat (*Triticum aestivum* L.) genotypes classified into two K-means clusters (C1 and C2). The plot depicts the relationship between grain yield (GY) and weight per spike (WPS), with fitted regression lines indicating the linear trend within each cluster.

There was a very high positive correlation between grain yield and spike weight, suggesting that spike productivity is an important part of wheat yield in the grain-yield germplasm tested. Mohammadi et al. (2012) also reported similar positive correlation between spike related traits and grain yield. The complete segregation of the Cluster I and Cluster II again confirms the validity of the multivariate analysis and the high phenotypic variability among the genotypes, indicating the usefulness of graphical analysis to detect different parental lines (Mohammadi and Prasanna, 2003). The superior performance of these genotypes (LAL BAHADUR, WH1080, HINDI62, PBW175RF, HI1531 and MP3224) indicates that these can be used as promising donor parents for wheat improvement. This is consistent with Reynolds et al. (2009), who argued that the spike-related traits are essential to improve wheat yield potential and with Acreche and Slafer (2006) who showed that the increase of spike productivity is directly linked to the increase of grain production. Similarly, Sadras and Lawson (2011) found that wheat yield improvement is closely linked with the improvement of important yield components. Overall, regression analysis along with correlation and path coefficient analysis has established that weight per spike is a good indirect selection criterion to advance wheat breeding programmes for better grain yield.

Violin plot analysis

The violin plot was used to display the distribution of PC1 scores in the two clusters identified (Figure 7). The distribution of Cluster I (C1) was relatively narrow with a low median value and hence the genotype was more homogeneous. However, Cluster II (C2) exhibited a wider distribution range with a higher median, indicating more variation among the genotypes and the possibility of higher yielding and better performing genotypes. The clear separation between the two violin plots further confirmed the wheat germplasm to be in two distinct groups.

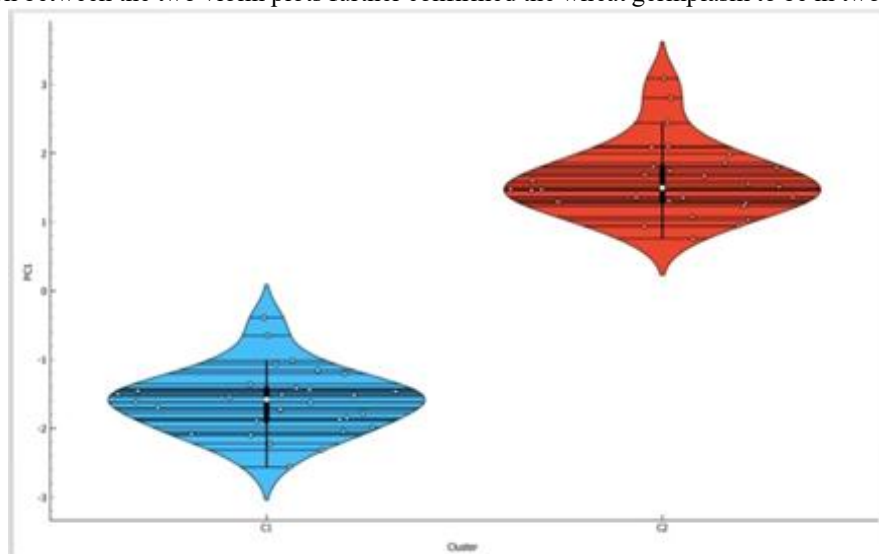


Figure 7. Violin plot showing the distribution of principal coordinate (PC1) scores among wheat (*Triticum aestivum* L.) genotypes grouped into two K-means clusters (C1 and C2). The width of each violin represents the density of genotypes, while the embedded boxplot indicates the median and interquartile range.

The PC1 score differences between the two clusters suggest significant phenotypic variation among the wheat genotypes evaluated. There is more variation in Cluster II than in Cluster I, which would suggest a greater genetic diversity and possibly better breeding material are present in Cluster II than in Cluster I. Mohammadi and Prasanna (2003) also found similar trends of variation among the genetically distinct groups in wheat germplasm, and advocated the use of multivariate techniques for diversity assessment. Govindaraj et al (2015) also pointed out that graphical visualization techniques can enhance understanding of genetic variability and help in the selection of superior breeding materials. Moreover, the agreement of the violin plot with the hierarchical, K-means, heatmap and MDS showed the consistency of the genetic structure revealed, and reinforced confidence in the classification of the germplasm considered. These complementary multivariate methods are widely recommended to efficiently characterize wheat genetic resources (Rasheed et al., 2018; Fu, 2015).

Cluster mean profile analysis

The cluster mean profile plot showed significant difference between the two genetic clusters for agro-morphological traits. The mean values of the GY, WPS and TGW for Cluster II were always higher than those in Cluster I while the difference between the two for DFF and PH were comparatively small. The differences between cluster profiles suggest considerable variation in phenotype among wheat genotypes evaluated. The better performance of Cluster II for yield and spike-related traits indicate that there are higher yielding genotypes in this cluster. The higher mean value for weight per spike is consistent with the correlation and path coefficient analysis as it was considered as one of the major contributing characters to grain yield. Wheat spike productivity in relation to wheat productivity has been reported by Mohammadi et al. (2012) and spike productivity was stressed to improve yield potential by Reynolds et al. (2009). The mean values of the yield related traits were relatively lower in Cluster I, but it showed higher genetic divergence which is important to obtain complementary alleles for breeding. Mohammadi and Prasanna (2003) and Rasheed et al. (2018) also noted that genetically divergent parents are essential for the diversifying the gene pool of wheat. Hence, the use of high yielding genotypes from Cluster II and genetically different genotypes of Cluster I may lead to the development of better recombinants with a higher productivity and greater genetic variability.

Breeding implications

Results of the present study showed that there was considerable phenotypic variation among the wheat genotypes evaluated indicating their potentialities in the breeding programme. Correlation and path coefficient analysis showed tiller number and weight per spike were the important yield contributing characters and could be considered as reliable selection criteria for enhancing grain yield. The hierarchical clustering, K-means clustering, heatmap and MDS analyses also clearly clustered the genotypes into genetically distinct groups, further supporting the existence of genetically distinct groups in the germplasm. Genotypes with desirable agronomic performance combined with genetic divergence (genotypes HI1531, MP3224, PBW590, Kharchia65, WH1021,

IC42, NP866 and GW451) were valuable parents for wheat improvement. Inter-crosses of genetically distinct and superior genotypes such as WH1021 × HI1531, IC42 × PBW590, NP866 × HD3090 and GW451 × Kharchia65 could produce wide genetic diversity and enhance the chances of selecting superior recombinants in future generations. The results are helpful in wheat parent selection and in the development of high yielding wheat varieties from a wider genetic base.

CONCLUSION

The results of this study showed that there is high phenotypic variation among the wheat germplasm used for evaluation, which should facilitate wheat improvement. The grain yield was strongly positively correlated with tiller number and weight per spike and had high direct effects of these traits among the studied traits, respectively, indicating their importance as the most important determinants of grain yield. These characteristics can therefore be used as adequate screening parameters to improve grain productivity. Hierarchical clustering, K-means clustering, heatmap and multidimensional scaling showed the genotype classification into two main groups, which indicated the strength of the genetic structure found and the significant variation among germplasm. Agro-morphological characterization coupled with multivariate statistical analyses helped to identify genetically different and high-yielding genotypes for parent selection. The study confirms that the use of the agro-morphological evaluation with multivariate analyses is a useful tool for evaluating genetic diversity and selecting breeding materials. The superior and divergent genotypes identified are valuable genetic resources to be used in developing high yielding wheat cultivars with wider genetic base and adaptation to various agro-climates.

REFERENCES

1. Acreche M M and Slafer G A. 2006. Grain weight response to increases in number of grains in wheat in a Mediterranean area. *Field Crops Research*. 98(1) pp52–59.
2. Ali N, Hussain I, Ali S, Khan N U et al. 2021. Multivariate analysis for various quantitative traits in wheat advanced lines. *Saudi J Biol Sci*. 28(1) pp347-352.
3. Blum A. 2011. *Plant Breeding for Water-Limited Environments*. Springer, New York.
4. Borg I and Groenen P J F. 2005. *Modern Multidimensional Scaling: Theory and Applications*. 2nd ed. Springer, New York, USA.
5. Broccanello C, Bellin D, DalCorso G and Furini A et al. 2023. Genetic approaches to exploit landraces for improvement of *Triticum turgidum* ssp. *durum* in the age of climate change. *Front Plant Sci*. 14.
6. Dewey D R and Lu K H. 1959. A Correlation and Path-Coefficient Analysis of Components of Crested Wheatgrass Seed Production I. *Agronomy Journal*. 51(9) pp515-518.
7. Dubcovsky J and Dvorak J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*. 316(5833) pp1862-6.
8. FAO. 2026. FAOSTAT Statistical Database. Food and Agriculture Organization of the United Nations. Rome. Italy.
9. Federer W T. 1956. Augmented (or hoonuiaku) designs. *Hawaiian Plant Rec LV*. (2) pp191–208.
10. Franco J, Crossa J, Taba S and Shands H. 2003. A Multivariate Method for Classifying Cultivars and Studying Group × Environment × Trait Interaction. 43(4).
11. Fu Y B. 2015. Understanding crop genetic diversity under modern plant breeding. *Theor Appl Genet*. 128 pp2131–2142.
12. Govindaraj M, Vetriventhan M and Srinivasan M. 2015. Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. *Genetics Research International*. pp 14.
13. Hill W G and Mackay T F C. 2004. D. S. Falconer and Introduction to Quantitative Genetics. *Genetics*. 167(4) pp1529–1536.
14. Jaradat A A. 2013. *Wheat Landraces: Genetic Resources for Sustenance and Sustainability*. USDA-ARS Monograph.
15. Kannaujiya T, Kumar K, Yadav M, Gaur S et al. 2020. Studies on Correlation and Path Coefficient for Yield and its Component Traits in Bread Wheat (*Triticum aestivum* L. em. Thell). *International Journal of Current Microbiology and Applied Sciences* 11 pp688-696.
16. Khan N and Naqvi F N. 2012. Correlation and Path Coefficient Analysis in Wheat Genotypes under Irrigated and Non-Irrigated Conditions. *Asian Journal of Agriculture Sciences*. 4(5) pp346-351.
17. Khodadadi M, Fotokian M H and Miransari M. 2011. Genetic diversity of wheat (*Triticum aestivum* L.) genotypes based on cluster and principal component analyses for breeding strategies. *Australian Journal of Crop Science*. 5(1) pp17-24.
18. Kumar P K, Singh S V, Yadav R K, Singh L et al. 2024. Correlation and Path Coefficient Analyses for Grain Yield and Its Contributing Traits in Bread Wheat (*Triticum Aestivum* L. Em. Thell). *Journal of Advances in Biology and Biotechnology*. 27 (3) pp208-18.
19. Kumar R, Singh S K, Bagare V, Singh A et al. 2023. “Path Coefficient Analysis for Yield and Its Attributing Traits in Wheat (*Triticum Aestivum* L.)”. *International Journal of Plant and Soil Science*. 35 (20) pp651-58.
20. Lopes M S, El-Basyoni I, Baenziger P S, Singh S et al. 2015. Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of Experimental Botany*. 66(12) pp3477–3486.

21. Mohammadi M, Sharifi P, Karimizadeh R and Shefazadeh M K. 2012. Relationships between Grain Yield and Yield Components in Bread Wheat under Different Water Availability (Dryland and Supplemental Irrigation Conditions). *Not Bot Horti Agrobo*. 40(1) pp195-200.
22. Mohammadi S A and Prasanna B M. 2003. Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Science*. 43(4) pp1235-1248.
23. Rasheed A, Mujeeb-Kazi A, Ogbonnaya F C, H Zhonghu et al. 2018. Wheat genetic resources in the post-genomics era: promise and challenges. *Annals of Botany*. 121(4) pp603–616.
24. Reif J C, Zhang P, Dreisigacker S, Warburton M L et al. 2005. Wheat genetic diversity trends during domestication and breeding. *Theor Appl Genet*. 110(5) pp859-64.
25. Reynolds M P and Braun H J. 2022. Wheat Improvement. In: Reynolds M P, Braun H J. (eds) *Wheat Improvement*. Springer, Cham. pp3-15.
26. Reynolds M, Foulkes M J, Slafer G A, Berry P et al. 2009. Raising yield potential in wheat. *Journal of Experimental Botany*. 60(7) pp1899–1918.
27. Sadras V O and Lawson C. 2011. Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. *Crop and Pasture Science*. 62(7) pp533–549.
28. Shewry P R and Hey S J. 2015. The contribution of wheat to human diet and health. *Food and Energy Security*. 4(3) pp178–202.
29. Singh R K and Chaudhary B D. 1985. *Biometrical Method in Quantitative Genetics Analysis*. Kalyani Publishers. New Delhi.
30. Tanksley S D and McCouch S R. 1997. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science*. 277 pp1063-1066.
31. Tester M and Langridge P. 2010. Breeding Technologies to Increase Crop Production in a Changing World. *Science*. 327(5967) pp818-822.
32. United Nations. 2022. *World Population Prospects*.
33. Wilkinson L and Friendly M. 2009. The History of the Cluster Heat Map. *Am Stat*. 63(2) pp179-184.
34. Xiong W and Reynolds M P. 2024. Wheat breeding strategies for increased climate resilience. *Nature Climate Change*. 14(8) pp791–792.
35. Zhang H, Mittal N, Leamy L J, Barazani O et al. 2017. Back into the wild Apply untapped genetic diversity of wild relatives for crop improvement. *Evol Appl*. 10 pp5-24.