Journal of Agricultural, Food and Environmental Sciences Vol 76 No 2 (2022) 47-54
DOI: https://doi.org/10.55302/JAFES22762047qr

Original scientific paper

PRINCIPAL COMPONENT AND CLUSTER ANALYSIS AS A TOOL IN THE ASSESSMENT OF DIFFERENT WHEAT SPECIES AND THEIR GENOTYPES

Leonora Qoku Rufati[1*], Silvana Manasievska[2]

[1]Faculty of Natural Sciences and Mathematics, University of Tetovo, Tetovo, Republic of North Macedonia
[2]Faculty of Agricultural Sciences and Food - Skopje, Ss Cyril and Methodius University in Skopje, Republic of North Macedonia
[*]corresponding author: leonora.qoku@unite.edu.mk

ABSTRACT

Determination of genetic relationships and determination of best criteria for selection of lines to be included in future breeding program is an invaluable aid in crop improvement**.** Various statistical techniques have been used to study diversity among different genotypes. Among these techniques multivariate is most frequently used one for the genetic association of genotypes. The present study was undertaken to screen genetic variation among 95 genotypes of wheat for their morphoagronomic traits. Five quantitative traits: plant height, number of fertile tillers, spike length, number of grains per spike, weight of grain per spike were evaluated by PCA and two-way cluster analysis. Three main principal components were determined explaining 85.75% of the total variation among the genotypes. 37.71% of the variation is explained by PC1 which reflects number of grains per spike, weight of grain per spike**.** PC2 and PC3 explained 28.19% and 19.85% of the total variance, mostly in relation to number of fertile tillers and the plant height, respectively. Biplot graph revealed strongest positive association between number of grains per spike and weight of grain per spike and negative association between spike length and number of fertile tillers. Two-way cluster analysis resulted with a dendrogram with one solely separated genotype, inferior for most of traits and four main clusters of genotypes defined with wide genetic diversity especially between the groups within the first cluster. From the findings of present study genotypes with high values for specific traits can be directly recommended for general cultivation or to be used in future breeding programs.

Key words: wheat, cluster, PC, morphology, traits, variability.

INTRODUCTION

Wheat (*Triticum* sp.) is one of the world’s most important crops, and constantly increasing its productivity is crucial to the livelihoods of millions of people. Narrowing the range of genetic variation, as a result of intensive breeding is however limiting the opportunity of improving wheat productivity (Sharma et al., 2021). Knowledge of genetic diversity and genetic variation based on morphological characteristics, especially those of economic interest can be an invaluable assistance in breeding programs for crop improvement (Dos Santos et al., 2009). An understanding of the genetic relationships among lines can be particularly useful in identifying diverse parental combinations to create segregating progenies with maximum genetic variability for further breeding (Bhandari et al., 2017), introgressing desirable genes from diverse germplasm into the available genetic base (Singh et al., 2021) and in precise identification with respect to plant varietal protection (Govindaraj et al., 2015).

Studies of taxonomic diversity among wheat germplasm, for morphological traits, is recognized as the first step for their classification and description although several other tools are also used extensively (Smykal et al., 2008; Sonmezoğlu et al., 2012; Babić et al., 2011; Maqbool et al., 2010). Objectives of such assessments are to classify the accessions into homogenous groups for mentioned traits, to characterize obtained groups, to analyze relationships among the groups and to identify the subset of traits that best distinguished differences among these groups. Wheat collection germplasm classification, according to morphological traits, principal component and cluster analyses are proved as useful tools (Mohammadi & Prasanna, 2003). Evaluation of phenotypic variability by multivariate analysis gives the possibility to include a large number of accessions and to identify the most suitable resources for special traits (Goel et al., 2015).

Present study was conducted to determine patterns of taxonomic variation and diversity for five morphological characteristics, wheat genotypes classification in the groups of similarity (clusters) and to identify traits, that mostly discriminate distinguished groups of genotypes.

MATERIAL AND METHODS

Plant material and experimental design

Three wheat species *Triticum durum* Desf. (durum wheat), *Triticum aestivum* L. (bread wheat) and *Triticum dicoccum* Schrank ex Schübler (emmer wheat) and the lines obtained from their hybridization (Manasievska Simik, 2006). A total of 90 wheat lines, two durum varieties, two bread varieties and one landrace of *Triticum dicoccum* were analyzed. RCBD design was applied with two replications during 2018/19 at location Opaje, Kumanovo. During the growth season, standard crop management practices were applied. Twenty plants from each plot were randomly collected for biometrical analysis of the following quantitative traits: plant height (PH), number of tillers (NT), spike length **(**SL), grain number per spike (NK) and grain weight per spike (KM).

Statistical analyses

The mean values of the genotypes for the quantitative traits were analyzed with different packages in the R 3.0.3 statistical software. Principal component analysis (PCA) was performed by using “psych” package (Revelle, 2014) where data matrix is standardized by default, meaning that component scores are standard scores (mean = 0, sd = 1). Components were extracted until the Eigen value > 1. Two-way cluster analysis was carried out with the quantitative traits that were highly correlated with the first three principal components. Euclidian distance between the genotypes was calculated by use of standardized values of the selected traits. The obtained values were normalized according to Roldan-Ruiz et al. (2001) and cluster dendogram was created by UPGMA method. The optimal number of clusters was estimated by multiscale bootstrap resampling for evaluation of uncertainty in hierarchical cluster analysis (Suzuki and Shimodaira, 2013). For each cluster, the p-values were calculated (a value between 0 and 1) and the clusters with approximately unbiased (AU) p-values larger than 95% after 1000 bootstrapping replications were taken to be strongly supported by data. This analysis resulted with a dendrogram which was used as a row dendrogram in the heatmap, while the column dendrogram was obtained in relation to the row dendrogram. Two-way cluster was generated by use of “heatmap.plus” package (Day, 2012).

RESULTS AND DISCUSSION

Principal component analysis

PCA represents multivariate technique which can be effectively used for transforming the number of correlated variables into a smaller number of variables called principal components. The number of components extracted is equal to the number of variables being analyzed. The first component covers the maximum variability of the data in relation to the following components. The method allows visualization of differences between genotypes and identification of potential groups and for identification of components that are positively or negatively correlated with a certain plant morphological characteristics. Characteristics with high correlation gave higher input to the variation pattern of the analyzed wheat genotypes.

By the application of principal component analysis three PCs were determined with Eigen values > 1. They accounted for 85.75% of the variation among the wheat genotypes used in this study (Table 1).

Table 1. Principal components score for the analyzed traits

Characters	Factor loadings		
	PC1	PC2	PC3
PH	-0.385	0.005	0.900
SL	0.486	-0.688	0.343
NK	0.846	0.360	0.125
KM	0.742	0.521	0.181
NT	-0.485	0.731	0.126
Eigen value	1.88	1.41	1.00
Proportion of Variance (%)	37.71	28.19	19.85
Cumulative Variance (%)	37.71	65.90	85.75

The first PC was related with grain number per spike, grain weight per spike and spike length, explaining 37.71% of the variability and largely determine the diversity of the studied genotypes. The second PC explained 28.19% of the total variance, mostly in relation to the number of tillers. The last component explains the smallest proportion of the variation (19.85%) and was strongly positively associated with plant height.

PC analysis revealed that principal discriminatory characteristics representing the clusters were grain number per spike, grain weight per spike and PH, with their highly positive contributions to the diversity of the analyzed lines. Similar results for plant height and grain yield obtained by PCA are reported from Khan et al. (2015) and Mishra et al. (2015).

Biplot has been applied to study relation among studied traits in a set of lines, grouping the wheat genotypes based on these traits as well as to identify the superior genotypes in relation to these traits.PC1 and PC2 factor loadings were plotted on Figure 1, where location of 95 genotypes displays the relation to the values of their Characteristics. Yan and Rajcan (2002) noted that cosine of the angle between trait vectors approximates the correlation coefficient between the traits, where, according to Yan and Kang (2003), correlation is positive, if angles are acute (< 90 °) and negative, if angles are obtuse (> 90 °).

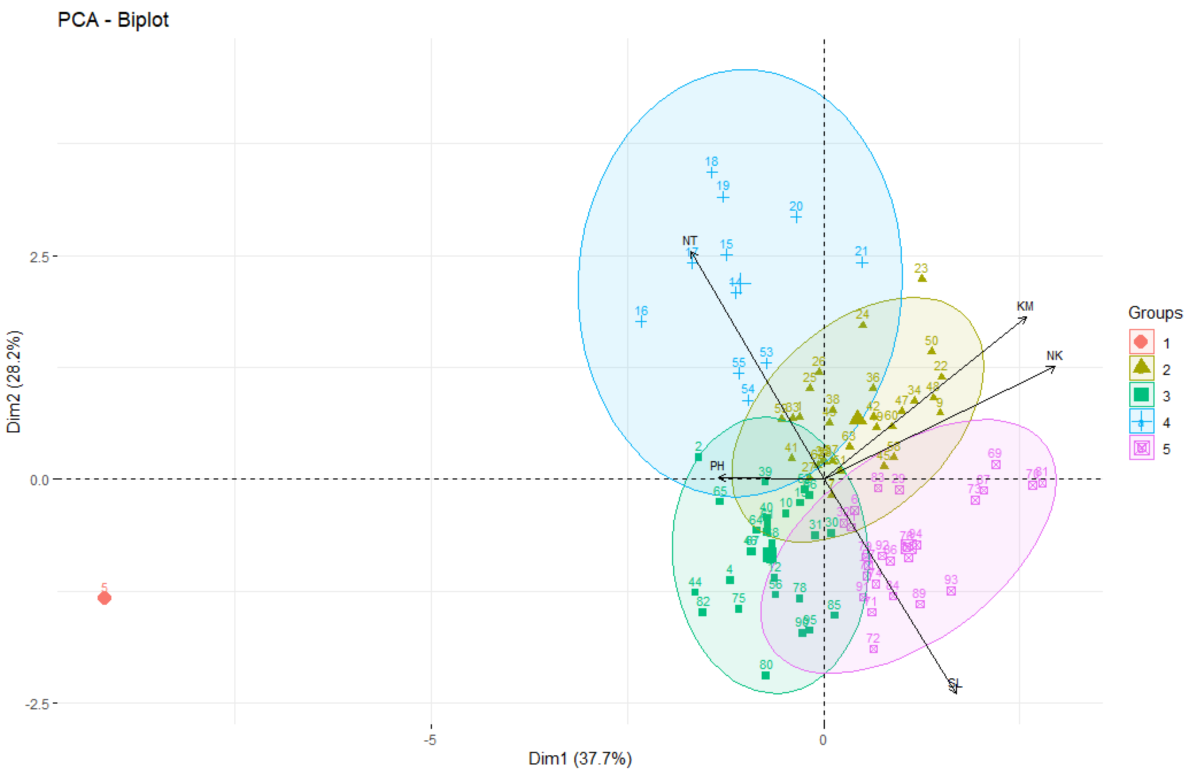

Figure 1. Biplot of PC1 and PC2 representing correlation between the genotypes and quantitative traits

The vector view of the biplot suggest a strong positive correlation among traits NK and KM also among PH and NT as indicated by the small obtuse angles between their vectors. The correlation between NT and KM, NK, SL; between KM, NK and SL was near zero as indicated by the near perpendicular vectors. The vectors indicated by the near angle of approximately 180 degrees, suggest for the existence of a strong negative correlation between PH and SL; also between PH and KM, NK. Location of the genotype, i.e. its distance from the biplot origin measures how it differs from a hypothetical “average” genotype located at the biplot origin that has an average level for all traits (Yan and Fregeau-Reid, 2008). Long vectors of the genotypes 5, 18, 19, 70 and 81 indicate that they have extreme values for one or more traits, out of which genotypes 70 and 81 are particularly superior lines for their high positive values of KM, KN and SL. Based on principal component method, the genotypes in this study were grouped into four clusters, each contained genotypes that were highly similar (Figure 1). Group 2 consisted of genotypes with high values for traits such KM and NK, group 3 consisted of genotypes with had high values for traits PH and SL. Group 4 genotypes were characterized by a large number of fertile tillers and group 5 genotypes had high values for traits NK and SL. Genotype 5 was one solely separated genotype, superior for PH, NT and inferior for NK, KM, SL.

Two-way cluster analysis

First two PCs refer to the largest amount of genotypes variability and highly correlated traits with these components were used for two-way cluster analysis. The analysis resulted with a dendrogram representing the genetic distance (degree of similarity) among 95 wheat genotypes. According to the analyzed data, one genotype 5 did not belong to any cluster. Genotype 5 is solely separated and most distanced from the others due to its low values for all traits except for PH and NT. All other genotypes were classified in four main clusters, whereas each cluster contained genotypes that were highly similar (Figure 2). Cluster 1 consisted of 41, cluster 2 of 32, cluster 3 of 13 and the fourth cluster of 28 wheat genotypes. This analysis reveals the presence of wide genetic diversity among the experimental material, according to grouping within the main clusters.

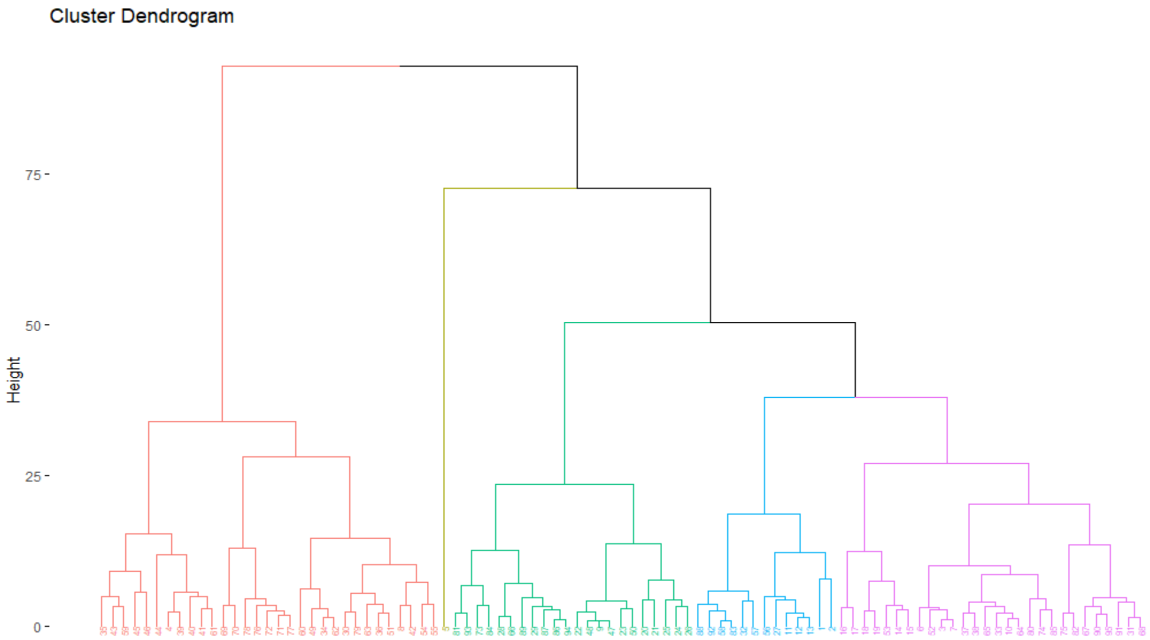

Figure 2. Dendrogram representing clustering of the analyzed wheat genotypes

The traits were classified in two groups where KM, NT along with SL belonged to the first group. The second group of traits consisted of NK and PH. (Figure 3).

First group of genotypes is characterized with low values for plant height and moderate values for the other analyzed traits. Second group consisted of genotypes with medium-high to high values for plant height, number of grains per class and mass of grains. The second group could be divided in two main subgroups (Figure 3). Most of the genotypes from the first subgroup had low values for spike length and in the second subgroup genotypes had medium-high to high values for the same trait. In the third group are located genotypes with medium-high values of plant height and spike length and also genotypes 1 and 2 except for plant height and grain weight per spike, had low values for all other traits. Three subgroups were detected in the fourth group. The first subgroups contain genotypes with high values for the number of tillers and low values for spike length, with medium-high to high plant height. The second subgroup included genotypes with medium-high values of plant height, medium-low to low values for the number of grains and low values for spike length (with the exception of genotypes 80, 74 and 85, within high values for spike length). The third subgroup contained genotypes with high values of plant height and spike length and low to medium-low values for the other traits.

Superior genotypes in some clusters might be usefully involved in hybridization programs for obtaining transgressive segregating lines with high genetic yield potential. All genotypes not belonging to any cluster (2, 14, 6, 1 and 3) could be used in the breeding program for improvement of certain characteristics for which they show high values.

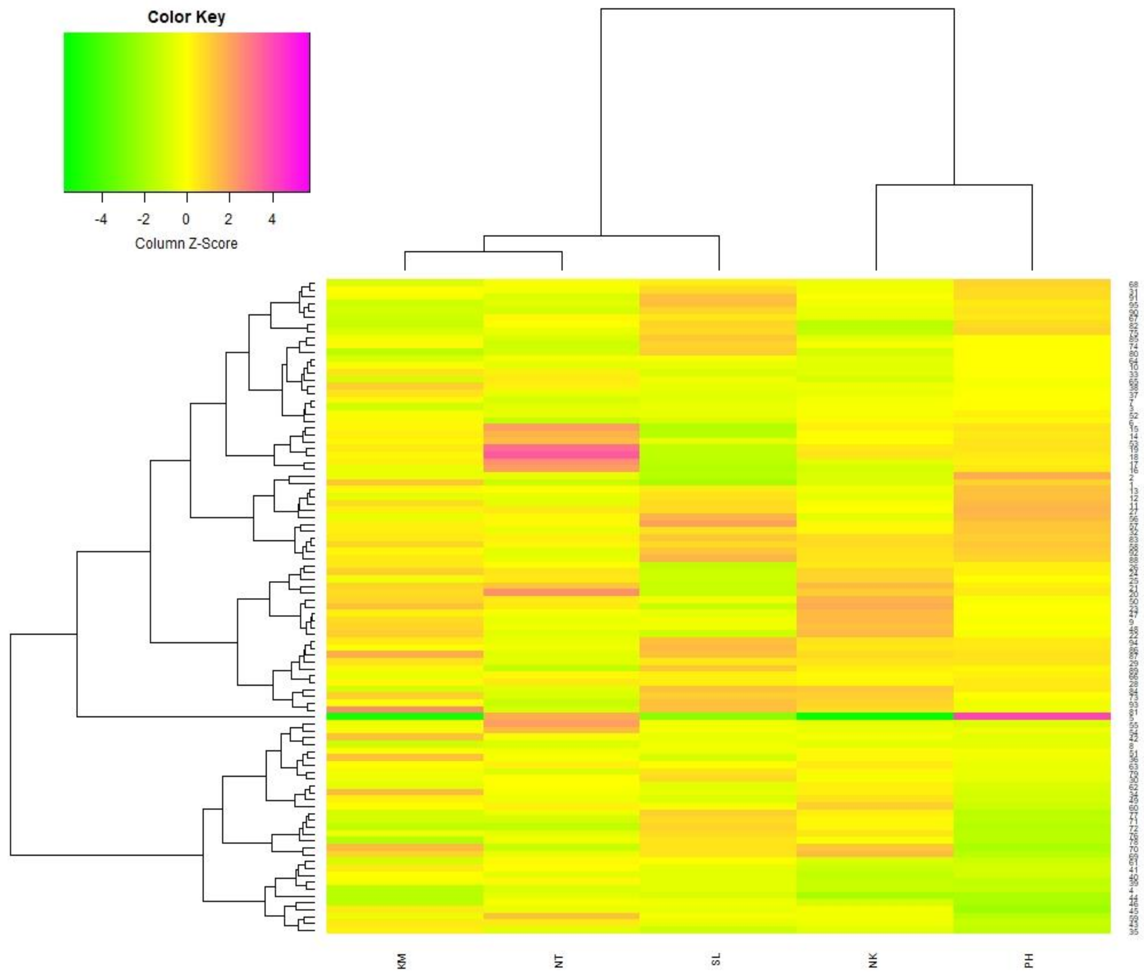

Figure 3. Two-way cluster with classification of genotypes according to the effect of the analyzed traits

CONCLUSIONS

The existing variability between the analyzed wheat genotypes in this study was effectively evaluated by PCA and two-way cluster analysis. The significant differences found in the present study show the existence of a high genetic variability among the 95 bread wheat genotypes and analyzed quantitative traits, adequate for selection of desirable traits, and creation of new favorable gene combinations. Three main PCs explained over 85.75% of the variation among the genotypes, as grain number per spike and grain weight of grain per spike and spike length were principal discriminatory traits. Obtained results by biplot analysis are positively associated with wheat yield, especially genotypes 5, 18, 19, 70 and 81.

The use of principal component analysis (showing the largest contributor to the total variance) and cluster analysis in the wheat germplasm, simplify wheat classification, the identification of the superior genotypes (considering the evaluation of mean values) and their relation with morphological traits with possibility expansion in breeding programs. Identification of the most important wheat quantitative agronomical traits can facilitate selection of species, varieties and lines with desirable traits, increasing the information of the wheat germplasm. The traits with more significant weighting on respective PC variance can be utilized successfully, as quantitative markers for evaluation and characterization of the wheat germplasm.

REFERENCES

Babić, V., Babić, M., Filipović, N., Delić, B & Purar, M. (2011). Vizuelna ocena fenotipa po UPOV deskriptoru i srodnost linija kukuruza [Visual evaluation of the phenotype according to the UPOV descriptor and relatedness of maize lines]. Ratar. Povrt, 48, 321-326.

Bhandari, H.R., Bhanu. A. N,. & Srivastava, K. (2017) Assessment of genetic diversity in crop plants - an overview. Adv Plants Agric Res,7(3), 279-286.

Day, A. (2012): Package "heatmap.plus": Heatmap with more sensible behavior. http://cran.r-project.org/web/packages/heatmap.plus/heatmap.plus.pdf

Dos Santos, T. M., Ganança, F., Slaski, J. J., & Pinheiro de Carvalho Miguel, A. A. (2009). Morphological characterization of wheat genetic resources from the Island of Madeira, Portugal. *Genetic Resources Crop Evolution*, 56, 363-375.

Goel, S., Humaraswamy, H.H., Grewal, S., Singh, K., Jaat, S.R., & Singh, K.N. (2015). Morphological and agronomical characterization of common wheat landraces (Triticum aestivum L) collected from different regions of India. *International Journal of Current Research and Academic Review*, 11 (3), 14 - 23.

Govindaraj, M., Vetriventhan, M., & Srinivasan M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet Res Int*, 97, 431-487.

Khan, M.A., A. Anjum, M.A. Bhat, B.A. Padder, Z.A. Mir, Kamaludin (2015): Multivariate analysis for morphological diversity of bread wheat (Triticum aestivum L.) germplasm lines in Kashmir valley. *Journal of Science* 5 (6),372-376.

Manasievska S.S., C. (2006). Polimorfizam na glutenini, glijadini i komponenti na prinos kaj hibridi pcenica od vidovite *Triticum aestivum, Triticum durum* i *Triticum dicoccum* [Polymorphism on glutenin, gliadin and yield components in wheat hybrids of Triticum aestivum, Triticum durum and Triticum dicoccum], Doktorska disertacija,Univerzitet Sv. Kiril i Metodij vo Skopje,Fakultet za Zemjodelski nauki i hrana – Skopje.

Maqbool, R., Sajjad, M., Khaliq, I., Rehman, A., Khan, S. A., & Khan, H. S. (2010). Morphological diversity and trait association in bread wheat (Triticum aestivum L.). *American-Eurasian Journal of Agriculture and Environment Sciences*, 8(2), 216-224.

Mishra, C.N., Tiwari V., Satish, V., Gupta, A., Kumar, I. & Sharma D (2015): Genetic diversity and genotypes by trait analysis for agromorphological traits of wheat (Triticum aestivum L.). *SABRAO J. Breed. Genet*., 47(1): 21-28.

Mohammadi S.A., & Prasanna B.M. (2003). Analysis of genetic diversity in crop plants: salient statical tools and considerations. *Crop Sci* 43, 1235-1248.

Revelle W. (2014): Package "psych": Procedures for Psychological, Psychometric, and Personality Research. http://cran.r-project.org/web/packages/psych/psych.pdf

Roldan-Ruiz, I., Van Euwijk, T.J., Gilliland, P., Dubreuil, C., Dillman, J., Lallemand, M. & De Loose, C.P. (2001): A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (Lolium perenne L.) varieties. *Theor. Appl. Genet.*, 103,1138-1150.

Sharma, S., Schulthess, A. W., Bassi, F. M., Badaeva, E. D., Neumann, K., Graner, A., Özkan, H., Werner, P., Knüpffer, H., & Kilian, B. (2021). Introducing Beneficial Alleles from Plant Genetic Resources into the Wheat Germplasm. *Biology*, *10*(10), 982.

Singh, S., Jighly, A., & Sehgal, D. (2021). Direct introgression of untapped diversity into elite wheat lines. *Nat Food* 2**,** 819–827 .

Smýkal, P., Horáèek, R. Dostálová, M. & Hýbl, J. (2008). Variety discrimination in pea (Pisum sativum L.) by molecular, biochemical and morphological markers. *J. Appl. Genet*. 49(2),155-166.

Sonmezoğlu, O.A., Bozmaz, B., Yildirim, A., Kandemir, N., & Aydin, N. (2012). Genetic characterization of Turkish bread wheat landraces based on microsatellite markers and morphological characters. *Turk J Biol*, 36, 589-597.

Suzuki, R., & Shimodaria H. (2013): Package “pvclust”: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resambling http://www.is.titech.ac.jp/shimo/prog/pvclust.pdf.

Yan, W., & Kang, M.S. (2003). GGE biplot analysis: A graphical tool for breeders, geneticists and agronomists. CRC Press, Boca Raton, FL, USA. Yan, W., & Rajcan I.R. (2002). Biplot analysis of test sites and trait relations of soybean in Ontario. *Can. J. Plant Sci*, 42,11–20.