

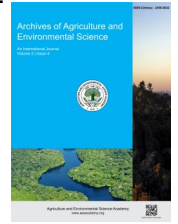


e-ISSN: 2456-6632

This content is available online at AESA

Archives of Agriculture and Environmental Science






Journal homepage: journals.aesacademy.org/index.php/aaes



REVIEW ARTICLE



A review on genetic parameters estimation, trait association, and multivariate analysis for crop improvement

Punam Roka^{1*} , Suraj Shrestha² , Shiva Prasad Adhikari¹ , Ayushma Neupane³ ,
Briksha Shreepaili²  and Mohan Kumar Bista⁴

¹Gokuleshwor Agriculture and Animal Science College, Institute of Agriculture and Animal Science, Tribhuvan University, Nepal

²Agriculture and Forestry University, Rampur, Chitwan, Nepal

³Nepal Polytechnic Institute, Purbanchal University, Nepal

⁴Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS, USA

*Corresponding author' E-mail: punamroka40@gmail.com

ARTICLE HISTORY

Received: 05 July 2024

Revised received: 25 August 2024

Accepted: 04 September 2024

Keywords

Cluster analysis

Genetic variability

Path coefficient

Principal component analysis

ABSTRACT

This review paper aims to elucidate the critical genetic parameters essential for practical crop breeding, focusing on the nature and extent of variability, its inheritance, and the complexity of traits. By evaluating genetic parameters such as Genetic Coefficient of Variation (G.C.V.), Heritability, Genetic Advance as a percentage of the Mean (G.A.M.), correlation coefficients, path coefficient analysis, cluster analysis, and principal component analysis, the review provides a comprehensive framework for optimizing breeding strategies. Emphasizing higher G.C.V. values minimizes environmental effects while highlighting Heritability and G.A.M. aids in predicting trait transmission and potential genetic improvement. The review also underscores the importance of traits with high G.C.V., Heritability, and G.A.M. for effective selection and improvement. Additionally, cluster and principal component analyses are powerful tools for identifying genetically diverse parents and reducing trait dimensionality. The findings suggest that thoroughly understanding and applying these genetic parameters can significantly enhance decision-making in plant breeding programs, ultimately leading to more efficient and targeted genetic improvements.

©2024 Agriculture and Environmental Science Academy

Citation of this article: Roka, P., Shrestha, S., Adhikari, S. P., Neupane, A., Shreepaili, B., & Bista, M. K. (2024). A review on genetic parameters estimation, trait association, and multivariate analysis for crop improvement. *Archives of Agriculture and Environmental Science*, 9(3), 618-625, <https://dx.doi.org/10.26832/24566632.2024.0903029>

INTRODUCTION

Crop improvement, a practice as ancient as agriculture itself, involves the development of new or enhanced traits in crops beyond their existing characteristics (Ahmed *et al.*, 2023). This process encompasses two primary techniques. The first technique is "selection," which harnesses the inherent genetic diversity within plant populations. While natural selection predominantly drove early crop improvement, contemporary practices primarily rely on artificial selection. Although these methods have become more targeted, they are carefully regulated to ensure a diverse gene pool is accessible for each crop species. The second technique, "breeding," involves the manipulation of crop traits through methods such as crossing and genetic

engineering (Bradshaw, 2017). Historically, early plant breeders did not understand genetic trait transmission and could not predict the outcomes of specific crosses. Despite this, advantageous traits occasionally emerged and were subsequently selected and preserved within the population.

In the current era, selection plays a vital role in crop breeding by optimizing desired traits (Plant *et al.*, 2024). Understanding the types and relative proportions of genetic components, along with the presence of allelic and non-allelic interactions for various traits, essentially guides the selection and breeding strategies for the genetic enhancement of any crop (Al-Naggar *et al.*, 2021). Breeders use genetic parameters to understand gene actions for quantitative and qualitative traits and evaluate the efficiency of different breeding methods in enhancing

genetic advancements (Anand et al., 2023). Estimating genetic parameters is crucial for determining the extent and degree of variability and Heritability of desirable traits for designing an effective genetic improvement program (You et al., 2016). Genetic parameters such as variance, coefficient of variation, Heritability, genetic advance, correlation studies, and path coefficient analysis gauge the population's alignment with breeding objectives. Although numerous studies have estimated genetic parameters, a comprehensive review must be provided. This review critically examines various genetic parameters utilized to elucidate the complexity of different qualitative and quantitative traits under investigation.

CROP VARIATION

The higher the variation for a character, the greater the extent of improvement through selection. Heritable variations in a population are the transmission of genetic differences from parents to offspring, originating from alterations in genetic material. Genetic recombination, chromosomal changes, mutations, and transposable genetic elements create these variations, which can also be induced through genetic engineering and tissue culture (Bhatia, 2018). Heritable variations are crucial for biodiversity and adaptation to changing climates (Scheffers et al., 2016). Early studies suggest that environmental factors experienced by ancestors can influence heritable variations (Yin et al., 2019). Conversely, environmental variation arises within a single generation due to environmental changes without altering genetic composition. Heterogeneous environmental factors, including disparities in nutrients, moisture, light, temperature, and exposure to disease and pests, cause non-heritable variation (Daleo et al., 2023). From a breeding perspective, heritable variations are valuable as they persist across generations, allowing for crop improvement. Genetic and environmental variations interact to affect plant traits. For instance, a rust-resistant wheat variety only outperforms a susceptible one in rust-present conditions. Breeders focus on heritable variations for long-term improvements (Cobb et al., 2019). The genetic differences existing in a breeding population determine the potentiality of its genetic improvement (Singh et al., 2023).

GENETIC VARIABILITY

Analyzing variability among characteristics is critical for developing a successful breeding strategy. The genetic difference in a breeding population determines its genetic improvement potential (Sanchez et al., 2023). The higher the variation for a character, the greater the extent of improvement through selection (Aman, 2021). Knowledge of the nature and magnitude of genetic variance provides an adequate selection criterion for breeders to select a given genotype. The variation coefficients reflect the variability level of the population mean. Assessing this involves calculating both the genotypic coefficient of variation (G.C.V.) and the phenotypic coefficient of variation (P.C.V.) (Dey, 2019). Heritable and non-heritable genetic effects give rise to variability in which G.C.V. signifies the heritable part and is more critical. In contrast, PCV signifies the genetic and environ-

mental impact and their interaction with the trait (Toor & Singh, 2023). Traits with a high percentage of G.C.V. to PCV are desirable, as high G.C.V. suggests excellent potential for successful selection (Khan et al., 2020). Differences between G.C.V. and P.C.V. indicate the environmental effect on the expression of traits. The more significant difference signifies the greater influence of the environment, whereas the narrower difference signifies less influence of the environment (Regmi et al., 2021). Generally, phenotypic values surpass their corresponding genotypic values as PCV estimates the effect of genotype and environment. Genotypic variance and phenotypic variance were obtained from the analysis of variance (ANOVA) table using the formula suggested by (Gomez & Gomez, 1984), as shown in Equation (1) and Equation (2), respectively.

$$\sigma_g^2 = \frac{MSS_g - MSS_e}{r} \quad (1)$$

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2 \quad (2)$$

Where,

σ_g^2 = genotypic variance, σ_p^2 = phenotypic variance, σ_e^2 = error variance, MSS_g = genotypic mean sum of square, MSS_e = error mean sum of square, r = number of replications

The formula suggested by (Burton, 1952) was used to calculate the genotypic coefficient of variation (G.C.V.) and phenotypic coefficient of variation (P.C.V.), as shown in Equation (3) and Equation (4), respectively.

$$GCV = \frac{\sqrt{\sigma_g^2}}{\bar{X}} \times 100 \quad (3)$$

$$PCV = \frac{\sqrt{\sigma_p^2}}{\bar{X}} \times 100 \quad (4)$$

Where,

σ_g^2 = genotypic variance, σ_p^2 = phenotypic variance, \bar{X} = general population mean.

HERITABILITY

Heritability is an estimate of the proportion of variation in a phenotypic trait in a population due to genetic variation among the individuals in that population (Rasheed et al., 2023). The degree of total variance of a character is attributable to genetic factors instead of environmental factors (Pennington & Sexton, 2024). Estimates of the coefficient of variation give information on the genetic variability seen in many quantitative variables. However, it cannot identify the level of heritable variation. As a result, the coefficient of variation and heritability estimates would offer a clear picture of the degree of advancement or advancement that may be predicted from selection

(Bhadru et al., 2020). The purpose of Heritability during the genetic studies of quantitative traits is the projecting role, which suggests the consistency of a phenotypic value to direct a breeding value (Fikret, 2023). The expected phenotypic value of an individual's offspring, expressed as the progeny's divergence from the population mean, defines the breeding value. It offers the parent's estimated transmission capacity (Hallauer, 2023). Therefore, Heritability offers information on transferability traits from generation to generation. Based on the variance component, there are two definitions of Heritability: broad-sense heritability and narrow-sense heritability. A trait's broad-sense Heritability (H_{bs}) defines the proportion of phenotypic variance attributable to genetic causes, including additive gene action, dominance, and epistasis. Narrow-sense heritability (H_{ns}) of the trait is defined as the proportion of phenotypic variance attributable to additive gene action (Mishra, 2023). Heritability values differ among genotypes and vary for different traits. Knowledge of Heritability allows plant breeders to determine the selection procedure under the improvement program (Sanchez et al., 2023). Traits with high heritability estimates pass quickly from one generation to the next. Selection of such a character could be simple, hence increasing selection efficiency. High heritability traits are genetically determined, and environmental variables have little influence on their phenotypic variation. Hence, these features could be increased by conventional breeding (Shrestha et al., 2023). Traits with moderate heritability estimates may react positively to phenotypic selection and could be improved through heterosis breeding or hybridization (Adhikari et al., 2018). Low heritability estimates for a trait indicate that the environment masks the phenotypic trait, making selection for that trait challenging (Tadesse et al., 2018).

Heritability in a broad sense (H_{bs}) was computed as suggested by (Johnson et al., 1955) and expressed as a percentage as shown in Equation (5).

$$H_{bs} = \frac{\sigma_g^2}{\sigma_p^2} \times 100 \% \quad (5)$$

Where,

$$\sigma_g^2 = \text{genotypic variance, } \sigma_p^2 = \text{phenotypic variance}$$

GENETIC ADVANCE

Genetic advancement refers to the improvement in the mean genotypic value of selected plants over the parental population. It is the degree of genetic gain under particular selection pressure (Vennela, 2023). Heritability estimates are helpful as highly heritable characters quickly progress when fixed with simple selection. Estimated Heritability is unreliable because it encompasses the effects of both additive and non-additive genes (Ilska et al., 2023). Thus, combining Heritability with genetic advance provides more accurate estimates of the trait of interest in a population after selection compared to using heritability estimates alone (Adhikari et al., 2018; Kahani & Hittalmani, 2016). Assessing the effectiveness of selection in trait improvement relies more on the ultimate genetic advance as a percentage of

the mean, which derives from Heritability, phenotypic standard deviation, and selection intensity (Lipi et al., 2021). High Heritability combined with high genetic advance estimates suggests that additive genes primarily control the traits, indicating that the pedigree method of breeding can achieve improvement (Tena et al., 2023). One can manipulate these traits according to requirements and achieve improvement through direct selection. Moreover, selection at an early segregating generation will be advantageous for selecting superior varieties (Wanga et al., 2021). High Heritability coupled with low genetic advance indicates that non-additive gene action controls these traits, which can be further enhanced through hybridization or heterosis breeding (Adhikari et al., 2018), and direct selection of these traits should be avoided (Bartaula et al., 2019). High Heritability in these traits may result from favorable environmental influences rather than the genotype. Similarly, high Heritability and moderate genetic advancement suggest the involvement of both additive and non-additive gene actions (Soliman et al., 2023). Additive gene action occurs when two alleles of a gene contribute equally during the production of the phenotype, whereas non-additive or dominant gene action occurs when the expression of one allele is stronger than the other allele (Rettew et al., 2008).

The expected genetic advance for each trait under a 5% selection intensity was determined using the formula outlined by (Johnson et al., 1955), as shown in Equation (6), and G.A.M., as shown in Equation (7).

$$GA = K \times \sqrt{\sigma_p^2} \times H_{bs} \quad (6)$$

$$GAM = \frac{GA}{\bar{x}} \times 100\% \quad (7)$$

Where,

GA = Genetic advance, K = constant or selection differential; K = 2.056 at 5% selection intensity; \bar{x} = mean of the population in which selection is employed.

TRAIT ASSOCIATION

Correlation coefficient

The correlation coefficient measures the extent and direction to which two variables are associated and related (Baye et al., 2020). It plays an important role in plant breeding by showing how changes in one character will bring simultaneous changes in the expression of another character. Correlation determined by a specific coefficient quantifies the degree of genetic and non-genetic association between the traits and allows the indirect selection of correlated traits (Tuliozi et al., 2023). Observing the association of traits through correlation coefficients helps select genotypes early or simultaneously when multiple traits are desired. Due to the correlated response of less complex characters with high heritability estimates and simple evaluation, indirect selection may derive higher genetic progress than

direct selection (Rehman et al., 2020). For example, suppose a primary trait X with low Heritability correlates highly with trait Y with high Heritability. In that case, indirectly improving trait Y and selecting trait X might be advantageous.

The intensity of the correlation coefficient, indicated as 'r,' ranges from -1 to +1 and is independent of the unit of measurement. The value '-1' indicates the oppositely directed relation between the characters, i.e., a high estimate of one character is related to a low estimate of another character. The value '+1' indicates the relation in the same direction and the presence of complete perfect association. And the value '0' indicates the absence of correlation between the characters and are independent of each other (Moll & Stuber, 1974). The three types of correlation in plant breeding are phenotypic, genotypic, and environmental (Resende et al., 2021). Phenotypic correlation reflects the relationship between directly observable traits, assessed by evaluating two traits across many individuals within a population. Genotypic correlation represents the association of breeding values, focusing on the additive genetic variance between two traits, and indicates the extent to which the same genes or closely linked genes cause simultaneous variation in both traits. Environmental correlations relate to environmental and non-additive genetic deviations (Isaac, 2021).

As a complex quantitative trait, grain yield is affected by many other component characters and depends on their action and interactions (Lephuthing et al., 2022). So, for an effective crop improvement program, understanding the interrelationship between grain yield and its contributing components is essential. Higher significant genetic correlation coefficients than phenotypic correlation coefficients signify the inherent relationship between various characters and suggest that the environmental effect masks the expression of the characters (Reddy & Jabeen, 2016; Tiwari et al., 2019). Traits that are positive and significantly correlated with the grain yield indicate that selection for these traits brings significant changes in the grain yield (Zewdu et al., 2024).

The phenotypic correlation coefficient and genotypic correlation coefficient among the agro-morphological traits were estimated using the standard protocols suggested by (R. et al., 1977) and from their respective variance and covariance components as in Equation (8) and Equation (9).

$$r_p = \frac{pCovxy}{\sqrt{\sigma_{px}^2 \times \sigma_{py}^2}} \tag{8}$$

$$r_g = \frac{gCovxy}{\sqrt{\sigma_{gx}^2 \times \sigma_{gy}^2}} \tag{9}$$

r_p =phenotypic correlation coefficient , r_g =genotypic correlation coefficient , $pCovxy$ = phenotypic covariance between variables x and y, $gCovxy$ = genotypic covariance between variables x and y,

σ_{px}^2 = phenotypic variance for variable x, σ_{py}^2 = phenotypic variance for variable y, σ_{gx}^2 = genotypic variance for variable

σ_{gy}^2 = genotypic variance for variable y.

Path coefficient analysis

Path coefficient analysis, a statistical tool initially developed by (Wright, 1921), facilitates understanding how variables interact within a multivariate setup. The principal use of this method is to establish and present causal relationships between predictor variables and response variables through path diagrams, which are constructed based on experimental findings or existing theoretical frameworks (Gholamin & Khayatnezhad, 2020). Although the simple correlation coefficient indicates how yield and its component traits are linked, it needs to precisely understand which component traits exert direct versus indirect effects on yield. A third variable can occasionally influence the correlation observed between two variables. Thus, path coefficient analysis allows for a more accurate depiction of variable interactions compared to correlation coefficients, highlighting the pathways through which each yield component exerts influence on overall yield (Olawamide & Fayeun, 2020). Path analysis breaks down the correlation coefficient into distinct elements: (a) the path coefficient or standardized partial regression coefficient, which quantifies the direct influence of a predictor variable on its response, and (b) the indirect influence of a predictor variable on the response through additional predictor variables (Dewey & Lu, 1959). Separating the direct influence of one variable while holding constant the indirect influences of others gives us a clear view of how each variable independently affects yield (Marinković, 1992).

Y is the yield (effect) of the causal factors x1, x2, and x3 (yield-related components'); r shows an association between variables; a, b, c, and h are path coefficients due to respective variables, and R indicates residual effect in Figure 1.

Path analysis, a distinct multivariate analysis, deals with a 'closed' system of linearly related variables. It helps to estimate the cause-effect relationship alongside effective selection recognizing the impact of multi-collinearity (Mitchell, 2020). Performing path analysis involves using the estimates of correlation coefficients.

The direct and indirect effects of the independent traits on dependent traits were estimated by the following Equation (10) as applied by (Dewey & Lu, 1959).

$$r_{ij} = p_{ij} + \sum r_{ik} \times p_{kj} \tag{10}$$

Where, r_{ij} = mutual association between the independent character (i) and dependent traits (j) as measured by the geno-

typic and phenotypic correlation coefficients, p_{ij} = components of direct effects of the independent traits (i) on the dependent traits (j) as measured by the genotypic and phenotypic correlation coefficients, $\sum r_{ik} \times p_{kj}$ = summation of components of indirect effects of a given independent character (i) on the dependent traits (j) via all other independent traits (k).

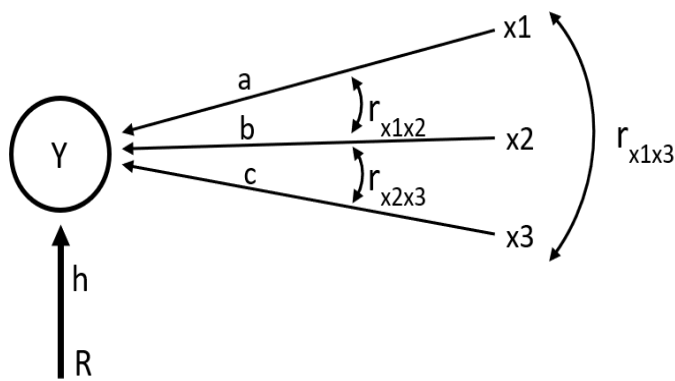


Figure 1. Path diagram showing cause and effect relationship.

To assess the indirect effect of the path coefficient, multiply the direct effect path coefficient by the correlation coefficient in the matrix. Sum the direct effects calculated through path analysis, each multiplied by the correlation coefficient of the dependent trait across all predictor traits, to obtain the coefficient of determination (R^2). The residual factor (R.F.), which gauges the influence of unexplained factors, is computed using equations (11) and Equation (12).

$$R^2 = \sum r_{ij} \times p_{ij} \quad (11)$$

$$RF = \sqrt{1 - R^2} \quad (12)$$

R.F.'s value reflects how efficiently causative factors account for the variability observed in the dependent factor. When the R.F. value is lower, it implies that the variability of the predictor variable adequately explains the variance in the dependent trait. In contrast, a higher R.F. value suggests including other unexamined factors in the research (Yeshitila et al., 2023).

MULTIVARIATE ANALYSIS

Cluster analysis

A statistical technique used in plant breeding called cluster analysis allows the identification of unique clusters or subgroups within a population by combining genotypes that are comparable based on their genetic traits (Talefe, 2023). It enables the blending of qualitative and quantitative data for selection and reduction using a similarity coefficient. Similar genotypes are then grouped into a single cluster (Ali et al., 2020). Two key concepts in plant breeding are intra and inter-cluster analysis, which groups genotypes according to specific characteristics. While intra-cluster analysis examines the distance within a single cluster, inter-cluster analysis studies the distance between distinct clusters. The greater inter-cluster distance than the intra-cluster distance between the mean values of the two clusters indicates more significant genetic variation between the genotypes of various groups (Khan et al., 2022). The proposal of D^2 statistics offers a powerful method for evaluating genetic divergence (Mahalanobis, 1936). In order to choose genetically dissimilar parents for use in hybridization programs, the D^2 statistics technique assesses the forces of differentiation at two levels: the intra-cluster and inter-cluster levels (Thell, 2022). There are two main types of clustering methods: hierarchical

and non-hierarchical. Plant breeders commonly use hierarchical clustering methods, which they categorize into agglomerative and divisive methods. Agglomerative methods start with each individual as a separate cluster and then combine them based on their similarities. At the same time, divisive methods begin with all individuals in one cluster and then divide them into smaller clusters based on their dissimilarities (Ali et al., 2020). There is no similarity between the two things with the most significant distance. The matrix being employed determines how far apart two items are from one another (Forina et al., 2001). A dendrogram is used to illustrate the hierarchical clustering of genotypes. Relationships between various genotypes or groups of genotypes based on their genetic or phenotypic similarities or differences are shown in this tree-like diagram, which is used in cluster analysis. To produce the dendrogram, the closest pair of genotypes or genotype clusters are repeatedly merged depending on their genetic distances. The average or maximum distance between the genotypes in the two clusters is used to calculate the distance between the merged clusters, depending on the clustering algorithm used. The procedure repeats until all genotypes merge into a single cluster. Cluster analysis was used to assess the magnitude of diversity associated (Palaniyappan et al., 2020). The vertical direction (Y-axis) in the dendrogram shows the distance between clusters in some metric (Pai et al., 2021), while the horizontal direction (X-axis) shows the data points. To elucidate a dendrogram, focus on the height at which any two genotypes join. Genotypes joined at a lower height in the dendrogram are the most similar, while those joined at higher heights are the most dissimilar (Figure 2).

Principal component analysis (P.C.A.)

Plant breeders frequently accumulate extensive datasets containing numerous variables, some of which may not sufficiently differentiate germplasm during evaluation, characterization, and management. To address this challenge, plant breeders commonly employ principal component analysis (P.C.A.), a statistical technique that helps identify patterns and reduce redundant information within datasets. P.C.A. is particularly useful given the vast morphological and physiological variations typically encountered in crop species (Khatun et al., 2023). P.C.A. is an exploratory tool designed to uncover underlying patterns within complex datasets. This statistical method transforms potentially correlated variables into a smaller set of uncorrelated variables known as principal components (Beattie & Esmonde-White, 2021). This statistical method is crucial in identifying essential polygenic characteristics that hold significant importance in a plant breeding program (Singh et al., 2020). It retains the essential information from the original data by breaking down and examining the correlations between many variables into a smaller collection of components. This estimation identifies the main traits that account for a sizable portion of the observed differences by assessing each trait's contribution to the overall reported genotype variations. As a result, breeders can concentrate on desirable features that significantly contribute to variability and advance crop improvement (Das et al., 2017).

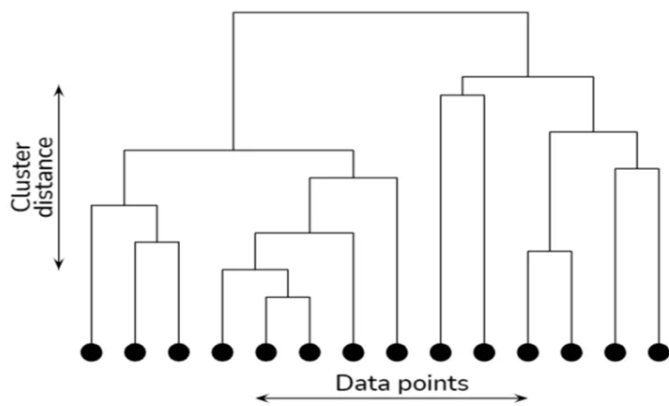


Figure 2. Visualization of dendrogram with data points (x-axis) and cluster distance (y-axis)

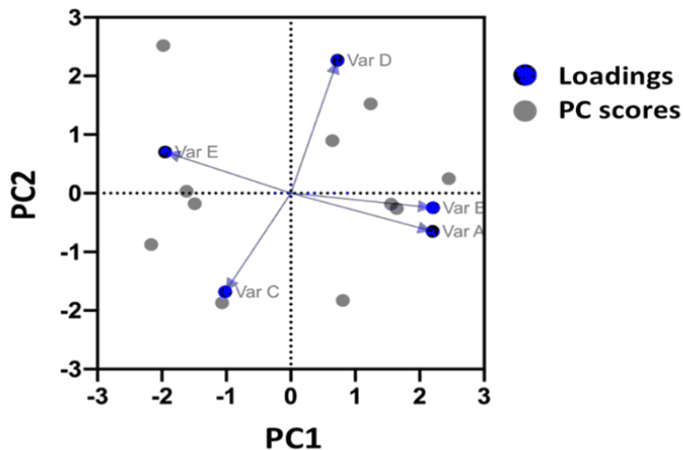


Figure 3. P.C.A. biplot showing the loadings and P.C. scores.

Use various approaches to determine the number of components to consider in Principal Component Analysis (P.C.A.). One initial approach is to create a plot of the eigenvalues based on their magnitudes and observe if there is a distinct point (often referred to as the 'elbow') where the slope of the graph transitions from steep to flat. Subsequently, we select and retain only those components before the elbow point. This method is known as the scree or elbow test (Kumar et al., 2022). Eigenvalues are coefficients applied to eigenvectors that give the vectors their length or magnitude (Rekha, 2019). The first principal component has the highest eigenvalue, the second principal component has the second highest eigenvalue, and so on (Kherif & Latypova, 2020). Another conventional approach is to retain only the components with eigenvalues more significant than the average. In correlation-based P.C.A., experts commonly advise keeping only the eigenvalues that exceed 1 (Graffelman & De Leeuw, 2023). However, following this method may result in the omission of crucial information. Another technique involves considering the total variance explained by the principal components, such as aiming for a threshold >80% (Kherif & Latypova, 2020). A PCA biplot is a graphical representation combining information about samples and variables from a data matrix. It displays samples as points and variables as either vectors, linear axes, or nonlinear trajectories (Nishisato et al., 2021). Merging a standard P.C.A. plot with a plot of loadings shows clusters of samples based on their similarities and illustrates how strongly each characteristic influences a principal component. In the P.C.A. biplot, the left and

bottom axes represent the P.C.A. scores of the samples (dots), while the top and right axes correspond to the loading plot, indicating the influence strength of each variable (vector) on the principal components. This arrangement is illustrated in Figure 3 (Escobar-Flores et al., 2019). The greater the distance of these vectors from the origin of a principal component (P.C.), the stronger their influence on that P.C. In the P.C.A. biplot, variables A, B, and E exhibit strong correlations (values near 1 or -1) with PC1, while their correlations with PC2 are comparatively weaker; conversely, variables C and D show strong correlations with PC2, but their correlations with PC1 are relatively weaker. The loadings also indicate correlations among variables: a slight angle between vectors suggests positive correlation, a large angle suggests a negative correlation and a right angle indicates no correlation between variables (Greenacre et al., 2022)

Conclusion

There is the central role of genetic parameters and multivariate analysis in crop breeding. Direct selection can target traits with higher values of G.C.V., Heritability, and G.A.M. High Heritability and genetic advancement favor conventional breeding, while lower values suggest hybridization strategies. Trait association analysis, such as correlation and path coefficients, illuminated the relationships between yield and its components, guiding indirect selection. Multivariate techniques like Principal Component Analysis (P.C.A.) simplified complex datasets and identified crucial traits, while cluster analysis grouped the observation into similar groups. Integrating advanced molecular tools with traditional methods was essential to optimize breeding programs. Collaborative efforts for robust data sharing can lead to innovative breeding strategies. These approaches collectively offer a robust framework for advancing genetic enhancement in crops and ensuring sustainable agricultural productivity.

DECLARATIONS

Authors' contribution

Conceptualization and design: P.R. and S.S.; Critical analysis and synthesis: P.R., S.S., and M.K.B.; Writing and drafting: P.R., S.S., S.P.A., A.N., and B.S.; Literature search and collection: P.R., S.S., S.P.A., A.N., and B.S.; Editing and revising: P.R., S.S., A.N., S.P.A. and M.K.B.

Conflicts of interest: The authors declare no conflict of interest.

Ethics approval: This study did not involve animal or human participants, so ethical approval did not apply to participants.

Consent for publication: All co-authors gave their consent to publish this paper in AAES.

Data availability: The data supporting this study's findings are available upon request from the corresponding author.

Supplementary data: Not.

ACKNOWLEDGEMENTS

We want to express our heartfelt gratitude to our parents for their unwavering support and encouragement throughout the preparation of this review paper. Their constant beliefs in our abilities and their patience have been invaluable in completing this work.

Open Access: This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) or sources are credited.

REFERENCES

- Adhikari, B. N., Joshi, B. P., Shrestha, J., & Bhatta, N. R. (2018). Genetic variability, Heritability, genetic advance and correlation among yield and yield components of rice (*Oryza sativa* L.). *Journal of Agriculture and Natural Resources*, 1(1), 149–160. <https://doi.org/10.3126/janr.v1i1.22230>
- Ahmed, B., Dhasarathan, B., Singh, S., & Gaur, H. (2023). *Edition 1 st Genetic Improvement of Crops Editors* (Issue August).
- Al-Naggar, A. M. M., Al-Azab, K. F., Younis, A. S. M., Hassan, I. O., Basyouny, M. A. E., & Ayaad, M. (2021). Genetic parameters controlling the inheritance of glaucousness and yield traits in bread wheat. *Brazilian Journal of Biology*, 82, e253864.
- Ali, N., Hussain, I., Ali, S., Khan, D. N., & Hussain, I. (2020). Multivariate analysis for various quantitative traits in wheat advanced lines. *Saudi Journal of Biological Sciences*, 28. <https://doi.org/10.1016/j.sjbs.2020.10.011>
- Aman, M. (2021). Genetic Variability, Heritability and Association of Quantitative Traits in Maize (*Zea mays* L) Genotypes: Review Paper. *E.A.S. Journal of Biotechnology and Genetics*, 3(2), 38–46. <https://doi.org/10.36349/easjbg.2021.v03i02.002>
- Anand, A., Subramanian, M., & Kar, D. (2023). *Breeding techniques to dispense higher genetic gains*. January, 1–6. <https://doi.org/10.3389/fpls.2022.1076094>
- Bartaula, S., Panthi, U., Timilsena, K., Acharya, S. S., & Shrestha, J. (2019). Variability, Heritability and genetic advance of maize (*Zea mays* L.) genotypes. *Research in Agriculture Livestock and Fisheries*, 6(2), 163–169. <https://doi.org/10.3329/ralf.v6i2.42962>
- 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 & Agriculture*, 6(1), 1752603.
- Beattie, J. R., & Esmonde-White, F. W. L. (2021). Exploration of principal component analysis: deriving principal component analysis visually using spectra. *Applied Spectroscopy*, 75(4), 361–375.
- Bhadru, D., Swarnalatha, V., Mallaiiah, B., Sreelatha, D., Kumar, M. V. N., & Reddy, M. L. (2020). Study of Genetic Variability and Diversity in Maize (*Zea mays* L.) Inbred Lines. *Current Journal of Applied Science and Technology*, 31–39. <https://doi.org/10.9734/cjast/2020/v39i3831093>
- Bhatia, S. (2018). Introduction to genetic engineering. In *Introduction to Pharmaceutical Biotechnology, Volume 1* (pp. 3–64). I.O.P. Publishing. <https://doi.org/10.1088/978-0-7503-1299-8ch3>
- Bradshaw, J. E. (2017). Plant breeding: past, present and future. *Euphytica*, 213(3), 60. <https://doi.org/10.1007/s10681-016-1815-y>
- Burton, G. W. (1952). Quantitative inheritance in grasses. *Pro VI Int Grassl Cong*, 277–283.
- Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., & Ng, E. H. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's Equation. *Theoretical and Applied Genetics*, 132(3), 627–645. <https://doi.org/10.1007/s00122-019-03317-0>
- Daleo, P., Alberti, J., Chaneton, E. J., Iribarne, O., Tognetti, P. M., Bakker, J. D., Borer, E. T., Bruschetti, M., MacDougall, A. S., Pascual, J., Sankaran, M., Seabloom, E. W., Wang, S., Bagchi, S., Brudvig, L. A., Catford, J. A., Dickman, C. R., Dickson, T. L., Donohue, I., & Hautier, Y. (2023). Environmental heterogeneity modulates the effect of plant diversity on the spatial variability of grassland biomass. *Nature Communications*, 14(1), 1809. <https://doi.org/10.1038/s41467-023-37395-y>
- Das, S., Sarma, D., Chakraborty, I., Roy, N., & Das, S. S. (2017). *Principal Component Analysis in Plant Breeding*.
- Dewey, D. R., & Lu, K. H. (1959). A Correlation and Path-Coefficient Analysis of Components of Crested Wheatgrass Seed Production 1. *Agronomy Journal*, 51(9), 515–518. <https://doi.org/10.2134/agronj1959.00021962005100090002x>
- Dey, P. (2019). Estimation of Phenotypic Coefficients of Variation (PCV), Genotypic Coefficients of Variation (G.C.V.), Heritability and Genetic Gain for Yield and its Components in Rice Landraces of Odisha. *International Journal of Agriculture Environment and Biotechnology*, 12(3). <https://doi.org/10.30954/0974-1712.08.2019.1>
- Escobar-Flores, P. C., Ladroit, Y., & O'Driscoll, R. L. (2019). Acoustic Assessment of the Micronekton Community on the Chatham Rise, New Zealand, Using a Semi-Automated Approach. *Frontiers in Marine Science*, 6. <https://doi.org/10.3389/fmars.2019.00507>
- Fikret, I. (2023). *Unravelling nature's blueprint: Understanding Heritability and genetic influence in plant and animal breeding*. V.S.N. International Ltd. <https://vsni.co.uk/blogs/Understanding-Heritability>
- Forina, M., Lanteri, S., & Rosso, S. (2001). Confidence intervals of the prediction ability and performance scores of classifications methods. *Chemometrics and Intelligent Laboratory Systems*, 57(2), 121–132. [https://doi.org/10.1016/S0169-7439\(01\)00129-0](https://doi.org/10.1016/S0169-7439(01)00129-0)
- Gholamin, R., & Khayatnezhad, M. (2020). The study of path analysis for durum wheat (*Triticum durum* Desf.) yield components. *Bioscience Biotechnology Research Communications*, 13(4), 2139–2144.
- Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural research*. John Wiley & Sons.
- Graffelman, J., & De Leeuw, J. (2023). Improved approximation and visualization of the correlation matrix. *The American Statistician*, 77(4), 432–442.
- Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 100. <https://doi.org/10.1038/s43586-022-00184-w>
- Hallauer, A. R. (2023). Breeding systems. In C.R.C. *Handbook of Plant Science in Agriculture* (pp. 61–87). C.R.C. press.
- Hoffmann, A. A., & Sgró, C. M. (2011). Climate change and evolutionary adaptation. *Nature*, 470(7335), 479–485. <https://doi.org/10.1038/nature09670>
- Ilska, J. J., Tolhurst, D. J., Tumas, H., Maclean, J. P., Cottrell, J., Lee, S. J., Mackay, J., & Woolliams, J. A. (2023). Additive and non-additive genetic variance in juvenile Sitka spruce (*Picea sitchensis* Bong. Carr). *Tree Genetics & Genomes*, 19(6), 53.
- Isaac, U. C. (2021). Phenotypic, genetic and environmental correlations between body weight and linear body traits of chicken genotypes. *Agrobiological Records*. <https://api.semanticscholar.org/CorpusID:230532831>
- Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955). Estimates of Genetic and Environmental Variability in Soybeans 1. *Agronomy Journal*, 47(7), 314–318. <https://doi.org/10.2134/agronj1955.00021962004700070009x>
- Kahani, F., & Hittalmani, S. (2016). Identification of F2 and F3 segregants of fifteen rice crosses suitable for cultivation under aerobic situation. *SABRAO Journal of Breeding and Genetics*, 48, 219–229.
- Khan, M. M. H., Rafiq, M. Y., Ramlee, S. I., Jusoh, M., & Mamun, A. (2020). Genetic Variability, Heritability, and Clustering Pattern Exploration of Bambara Groundnut (*Vigna subterranea* L. Verdc) Accessions for the Perfection of Yield and Yield-Related Traits. *BioMed Research International*, 2020, 1–31. <https://doi.org/10.1155/2020/2195797>
- Khan, S., Mahmud, F., & Ahmed, T. (2022). Genetic Diversity with Cluster Analysis of Maize Genotypes *Zea mays* L.). *Advances in Bioscience and Biotechnology*, 13(07), 273–283. <https://doi.org/10.4236/abb.2022.137017>
- Khatun, M., Islam, M. R., Khan, M. A. R., Hossain, M. K., & Islam, A. (2023). Principal component analysis and estimated breeding values for selecting suitable parental genotypes in rice (*Oryza sativa* L.). *Sabrao Journal of Breeding & Genetics*, 55(6).
- Kherif, F., & Latypova, A. (2020). *Chapter 12 - Principal component analysis* (A. Mechelli & S. B. T.-M. L. Vieira (eds.); pp. 209–225). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-815739-8.00012-2>
- Kumar, M., Sandhya, P. K., Gaurav, A., Soni, R. K., & Meena, D. (2022). *Principal component analysis for assessment of genetic diversity in rice under Humid South-Eastern plain of Rajasthan*.
- Lephuthing, M. C., Khumalo, T. P., Tolmay, V. L., Dube, E., & Tsilo, T. J. (2022). Genetic Mapping of Quantitative Trait Loci Associated with Plant Height and

- Yield Component Traits in a Wheat (*Triticum aestivum* L.) Doubled Haploid Population Derived from Tugela-DN × Elands. In *Agronomy* (Vol. 12, Issue 10). <https://doi.org/10.3390/agronomy12102283>
- Lipi, L., Hasan, M., Akter, A., Quddus, M., Biswas, P., Ansari, A., & Akter, S. (2021). Genetic Variation, Heritability and Genetic Advance in Some Promising Rice Hybrids. *SAARC Journal of Agriculture*, 18(2), 39–49. <https://doi.org/10.3329/sja.v18i2.51107>
- Marinković, R. (1992). Path-coefficient analysis of some yield components of sunflower (*Helianthus annuus* L.). *I. Euphytica*, 60(3), 201–205. <https://doi.org/10.1007/BF00039399>
- Mishra, A. (2023). Characterization, heterosis and gene action studies for biomass related quantitative traits in rice. Department of Plant Breeding and Genetics, OUAT, Bhubaneswar.
- Mitchell, R. J. (2020). Path analysis: pollination. In *Design and analysis of ecological experiments* (pp. 211–231). Chapman and Hall/C.R.C.
- Moll, R. H., & Stuber, C. W. (1974). *Quantitative Genetics—Empirical Results Relevant to Plant Breeding* (pp. 277–313). [https://doi.org/10.1016/S0065-2113\(08\)60874-3](https://doi.org/10.1016/S0065-2113(08)60874-3)
- Nishisato, S., Beh, E. J., Lombardo, R., Clavel, J. G., Nishisato, S., Beh, E. J., Lombardo, R., & Clavel, J. G. (2021). History of the biplot. *Modern Quantification Theory: Joint Graphical Display, Biplots, and Alternatives*, 167–179.
- Olawamide, D. O., & Fayeun, L. S. (2020). Correlation and Path Coefficient Analysis for Yield and Yield Components in Late Maturing Pro-vitamin A Synthetic Maize (*Zea mays* L.) Breeding Lines. *Journal of Experimental Agriculture International*, 64–72. <https://doi.org/10.9734/jeai/2020/v42i130452>
- P.C. Mahalanobis. (1936). *1936_Mahalanobis.pdf* (pp. 541–588).
- Pai, S., Emmerling, J., Drouet, L., Zerriffi, H., & Jewell, J. (2021). Meeting well-below 2°C target would increase energy sector jobs globally. *One Earth*, 4(7), 1026–1036. <https://doi.org/10.1016/j.oneear.2021.06.005>
- Palaniyappan, S., Arunachalam, P., Banumathy, S., Mini, M. L., & Muthuramu, S. (2020). Genetic divergence and clustering studies in advanced breeding lines of rice (*Oryza sativa* L.). *Electronic Journal of Plant Breeding*, 11(02), 499–504.
- Pennington, L. K., & Sexton, J. P. (2024). Species Range Limits. In S. M. Scheiner (Ed.), *Encyclopedia of Biodiversity (Third Edition)* (Third Edit, pp. 219–236). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-822562-2.00357-1>
- Plant, M., Montesinos-lo, O. A., Isidro, J., Alemu, A., Johanna, A., Tadesse, W., Vetukuri, R. R., Carlsson, A. S., Ortiz, R., Chawade, A., & Ceplitis, A. (2024). *Genomic selection in plant breeding: Key factors shaping two decades of progress*. 2023, 552–578. <https://doi.org/10.1016/j.molp.2024.03.007>
- Rasheed, A., Ilyas, M., Khan, T. N., Mahmood, A., Riaz, U., Chattha, M. B., Al Kashgry, N. A. T., Binothman, N., Hassan, M. U., & Wu, Z. (2023). Study of genetic variability, Heritability, and genetic advance for yield-related traits in tomato (*Solanum lycopersicon* MILL.). *Frontiers in Genetics*, 13, 1030309.
- Reddy, V. R., & Jabeen, F. (2016). Narrow sense heritability, correlation and path analysis in maize (*Zea mays* L.). <https://api.semanticscholar.org/CorpusID:199423598>
- Regmi, S., Poudel, B., Ojha, B. R., Kharel, R., Joshi, P., Khanal, S., & Kandel, B. P. (2021). Estimation of Genetic Parameters of Different Wheat Genotype Traits in Chitwan, Nepal. *International Journal of Agronomy*, 2021, 6651325. <https://doi.org/10.1155/2021/6651325>
- Rehman, A., Mustafa, N., Du, X., & Azhar, M. T. (2020). Heritability and correlation analysis of morphological and yield traits in genetically modified cotton. *Journal of Cotton Research*, 3(1), 23. <https://doi.org/10.1186/s42397-020-00067-z>
- Rekha, M. (2019). *Eigen Decomposition and P.C.A.* Clairvoyant Blog, Medium. <https://blog.clairvoyantsoft.com/eigen-decomposition-and-pca-c50f4ca15501>
- Resende, R. T., Piepho, H.-P., Rosa, G. J. M., Silva-Junior, O. B., e Silva, F. F., de Resende, M. D. V., & Grattapaglia, D. (2021). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics*, 134(1), 95–112. <https://doi.org/10.1007/s00122-020-03684-z>
- Rettew, D. C., Rebollo-Mesa, I., Hudziak, J. J., Willemsen, G., & Boomsma, D. I. (2008). Non-additive and Additive Genetic Effects on Extraversion in 3314 Dutch Adolescent Twins and Their Parents. *Behavior Genetics*, 38(3), 223–233. <https://doi.org/10.1007/s10519-008-9192-5>
- Sanchez, D., Sadoun, S. Ben, Mary-Huard, T., Allier, A., Moreau, L., & Charcosset, A. (2023). Improving the use of plant genetic resources to sustain breeding programs' efficiency. *Proceedings of the National Academy of Sciences*, 120(14), e2205780119.
- Scheffers, B. R., De Meester, L., Bridge, T. C. L., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. T., Butchart, S. H. M., Pearce-Kelly, P., Kovacs, K. M., Dudgeon, D., Pacifici, M., Rondinini, C., Foden, W. B., Martin, T. G., Mora, C., Bickford, D., & Watson, J. E. M. (2016). The broad footprint of climate change from genes to biomes to people. *Science*, 354(6313). <https://doi.org/10.1126/science.aaf7671>
- Shrestha, S., Niraula, D., Regmi, S., Basnet, S., Chhetri, S. T., & Kandel, B. P. (2023). Performance evaluation and genetic parameters estimation of multi-companies maize hybrids in Lamahi Dang, Nepal. *Heliyon*, 9(3), e14552. <https://doi.org/10.1016/j.heliyon.2023.e14552>
- Singh, A., Singh, A., Barb, J., & Mahama, A. A. (2023). Genetic variation and germplasm usage. *Crop Improvement*.
- Singh, P., Jain, P. K., & Tiwari, A. (2020). Principal component analysis approach for yield attributing traits in chilli (*Capsicum annum* L.) genotypes. *Chemical Science Review and Letters*, 9(33), 87–91.
- Singh, R. K., & Chaudhary, B. D. (1977). Biometrical methods in quantitative genetic analysis. Kalyani.
- Soliman, A. A., Mousa, M. I., Mosalam, A. M., Ghareeb, Z. E., Ibrahim, S. D., Rehan, M., Yu, H., & He, Y. (2023). The Potential Genetic Effect for Yield and Foliar Disease Resistance in Faba Bean (*Vicia faba* L.) Assessed via Morphological and SCOT Markers. *Plants*, 12(20), 3645. <https://doi.org/10.3390/plants12203645>
- Tadesse, J., Leta, T., Techale, B., & Lemi, B. (2018). Genetic variability, heritability and genetic advance of maize (*Zea mays* L.) inbred lines for yield and yield related traits in southwestern Ethiopia. *Journal of Plant Breeding and Crop Science*, 10(10), 281–289. <https://doi.org/10.5897/JPBCS2018.0742>
- Talefe, W. (2023). Cluster and principal component analysis of maize inbred lines in low moisture stress areas in Ethiopia. *International Journal of Agricultural Science and Food Technology*, 9, 59–63. <https://doi.org/10.17352/2455-815X.000194>
- Tena, E., Tadesse, F., Million, F., & Tesfaye, D. (2023). Phenotypic diversity, Heritability, and association of characters in sugarcane genotypes at Metehara Sugar Estate, Ethiopia. *Journal of Crop Improvement*, 37(6), 874–897.
- Thell, L. (2022). *Analysis of Genetic Divergence in some Elite Cultivars of Bread Wheat*. 14(3), 573–576.
- Tiwari, D. N., Tripathi, S. R., Tripathi, M. P., Khatri, N., & Bastola, B. R. (2019). Genetic Variability and Correlation Coefficients of Major Traits in Early Maturing Rice under Rainfed Lowland Environments of Nepal. *Advances in Agriculture*, 2019, 1–9. <https://doi.org/10.1155/2019/5975901>
- Toor, A. K., & Singh, A. (2023). Evaluation of variability, Heritability and genetic advance for yield and yield components in linseed (*Linum usitatissimum* L.) genotypes. *Journal of Eco-Friendly Agriculture*, 18(1), 80–85.
- Tuliozi, B., Mantovani, R., Schoepf, I., Tsuruta, S., Mancini, E., & Sartori, C. (2023). Genetic correlations of direct and indirect genetic components of social dominance with fitness and morphology traits in cattle. *Genetics Selection Evolution*, 55(1), 84. <https://doi.org/10.1186/s12711-023-00845-8>
- Vennela, M. (2023). Genetic Variability, Heritability and Genetic Advance Studies in Field Pea (*Pisum Sativum* L.) for Yield and Its Attributing Traits. *Int. J. Environ. Clim. Change*, 13(8), 2190–2197.
- Wanga, M. A., Shimelis, H., Mashilo, J., & Laing, M. D. (2021). Opportunities and challenges of speed breeding: A review. *Plant Breeding*, 140(2), 185–194.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20(7), 557.
- Yeshitila, M., Gedebo, A., Degu, H. D., Olango, T. M., & Tesfaye, B. (2023). Study on characters associations and path coefficient analysis for quantitative traits of amaranth genotypes from Ethiopia. *Scientific Reports*, 13(1), 20981. <https://doi.org/10.1038/s41598-023-47869-0>
- Yin, J., Zhou, M., Lin, Z., Li, Q. Q., & Zhang, Y. Y. (2019). Transgenerational effects benefit offspring across diverse environments: a meta-analysis in plants and animals. *Ecology Letters*, 22(11), 1976–1986. <https://doi.org/10.1111/ele.13373>
- You, F. M., Song, Q., Jia, G., Cheng, Y., Duguid, S., Booker, H., & Cloutier, S. (2016). ScienceDirect Estimation of genetic parameters and their sampling variances for quantitative traits in the type 2 modified augmented design. *C.J.*, 4(2), 107–118. <https://doi.org/10.1016/j.cj.2016.01.003>
- Zewdu, D., Mekonnen, F., & Geleta, N. (2024). Correlation and Path Analysis for Yield and Yield Component Traits on Bread Wheat (*Triticum aestivum* L.) Genotypes. *International Journal of Bio-Resource and Stress Management*, 15(1), 1–13. <https://doi.org/10.23910/1.2024.5040>