International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 12; Issue: 8; August 2025
Website: www.ijrrjournal.com
E-ISSN: 2349-9788; P-ISSN: 2454-2237

Review Paper

‘D^2 Statistics’ Technique: Methodology and Applications in Plant Breeding

Dr. Vaibhav Ujjainkar

Professor (Genetics & Plant Breeding) CAS, College of Agriculture, Akola
Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola 444 104 (MS) INDIA

DOI: https://doi.org/10.52403/ijrr.20250815

ABSTRACT

Agricultural scientists deals with diverse set of characters in respect to range of crops for making decisions through various field and laboratory experiments. These varying and extreme sets of data has to be analyzed and interpreted to draw valid conclusions, deciding the faith of long experimentations. In case of plant breeders, it found that the Mahalanobis distance (D^2 statistics) is an effective multivariate distance/metric that measures the distance between a point and a distribution. It is an extremely useful metric having, excellent applications in multivariate anomaly detection, classification on highly imbalanced datasets and one-class classification. Crop Improvement consists of selection of parents, crossing those parents to create variability, selection of elite genotypes and synthesis of suitable cultivar(s) from the selection. The choice of parental germplasm with which to begin a hybridization program is the most important decision a breeder makes. However, it is only relatively recent that quantitative genetic theory has been applied to this question. Recent years it has become imperative to use the modern techniques and genetic methods to fasten the process of improvement. Mahalanobis D^2 Statistics is one of the powerful statistical tools for estimating the genetic distances between the germplasm. These may be utilized for selecting parents in crossing program. In this paper the attempt made to discuss the procedural aspect of Mahalanobis D^2 useful for conventional breeding along with its applicability over range of experiments.

Key words: ANOVA, Clustering, D^2 statistics, Genetic Divergence, Intra and Inter Cluster Distances, Mahalanobis D^2 analysis, MANOVA.

INTRODUCTION

Multivariate analysis in agriculture, involves the use statistical techniques to analyze multiple variables simultaneously e.g. number of quantitative or qualitative traits, allowing researchers to understand complex relationships and patterns within agricultural systems from given set of parameters. This approach helps in making informed decisions regarding crop selection, resource management, and overall farm productivity. Commonly techniques viz., Principal Component Analysis (PCA), D^2 statistics, Factor Analysis (FA), and Canonical Correlation Analysis (CCA) etc are employed. This paper focus on the different aspects of D^2 Statistics coupled with techniques to categorization of large number of population and their usefulness through transforming genetic distances into statistical values.

The knowledge of nature and magnitude of genetic diversity present in the germplasm is most important pre-requisite for the success of any breeding program. It is thus necessary to survey the variation present in the germplasm as the hybrids between diverse lines displays greater heterotic effects than those between the closely related. The concept of combining ability plays an important role in identification of parents and production of superior parental lines or hybrid. The general combining ability and specific combining ability effects are the foundations for any fruitful breeding program. Allard (1960) pointed out that, the common approach for selection of parents on the basis of *per se* performance is not a good indication of their superior combining ability. The choice of parents in breeding program for extreme areas has to be based on the complex genetic information and knowledge of combining ability of parents and not merely on the field performance of different genotypes. Similarly, the combining ability estimates are greatly influenced by the environment, so that studies in single environment may not provide reliable information. Hence knowledge about combining ability of parents must be judged on set of environments and use of basic biometrical techniques. The genetic diversity of selected parents is not always based on the factors such as geographical diversity or place of release or ploidy level. Hence many statistical procedures viz., D^2 statistics, non-hierarchical Eucledian Cluster, Meteroglyph etc., are developed to measure the total variability present among the genotypes.

The perception, when choosing parents is to maximize the probability of generating new lines that will perform better than the best pure lines currently in use. The opportunity for selecting superior yielding segregants in any hybridization program is dependent of choice of parents. The existing variety over which improvement is sought is a logical choice as one parent, because of its proven performance. When the objective is to improve yield *per se* rather any other bottleneck trait, the second parent can be either the second high yielding parent from the same area of adaptation or from totally different area. There is no consensus on how best to choose parents selection remains one of the critical unanswered questions in the field of breeding, despite numerous theoretical and empirical studies to identify useful parents (Qualset, 1979 and Baker, 1984). The genetic architecture of a population is the result of prolonged natural and artificial selection. The population that lives in the heterogeneous environment may be genetically strongly diversified. Such diversified populations are usually included in the hybridization program because, more diverse the parents, the better are the chances of improving a character in view. However, according to Gomez (1966), the magnitude of heterosis increased with increased divergence with restricted range of diversity and decreased with further increments of diversity. Therefore the information about the nature and magnitude of genetic divergence in a given population is essential for selection of diverse parents which upon hybridization lead to a wide spectrum of gene recombinations for quantitatively inherited traits. Among several statistical methods to quantify genetic divergence, multivariate analysis is a potent tool (Rao 1969). Mahalanobis (1928) D^2 analysis, recognized as a powerful statistical tool to assess genetically divergent genotypes was utilized by several workers in range of several agricultural crops to quantify the degree of divergence at genotypic level. Moreover, earlier worker viz., Bhatt (1973) demonstrated the usefulness of D^2 statistics for choosing parents in wheat and Arunachalam (1981) in his study on genetic distance stated that genetic distance has a definite role to play in an efficient choice of parents for hybridization program.

PROCEDURES

Statistical procedures/techniques complexity signifies the inherent challenges involved in the analysis or modeling of a dataset, which can stem from its structure, randomness, or size. In this paper the sequential procedure for using D^2 statics, dispersion techniques and subsequent their application for selection of genotypes for further improvement is discussed as below.

A] Analysis of variance

Generally, the field experiments are laid down with proper randomization and experimental units to minimize the risk of experimental errors. The various standard designs of experiments are available to provide technical assistance to researchers in their research especially in field. The analysis of variance can performed to test the significance of differences among the genotypes for all the traits under investigation. The analysis of variance is generally out as per standard method (Panse and Sukhatme, 1967; Singh and Chaudhary, 1985) for all the characters using appropriate experimental designs viz., completely randomized design (CRD), Randomized block design (RBD), Factorial designs, Split plots, Augmented designs, Lattices designs etc.,. Here the most used design in field experiments specially related to breeding or agronomical studies has been discussed. The statistical model used for completely randomized block design is

$Y_{ijk} = \mu + g_{ij} + b_k + e_{ijk}$

Where,

Y_{ijk} = Phenotypic performance of ij^{th} genotype in k^{th} block

μ =General mean

g_{ij} = Effect of ij^{th} genotype

b_k = Effect of k^{th} block

e_{ijk} = Error effect

The analysis of variance based on this model is as under:

Source of variation	d.f.	Sum of squares (SS)	Mean sum of squares (MSS)		F value
			Observed	Expected	
Replications	r-1	SSr	Mr	$\sigma^2e+g\sigma^2r$	Mr / Me
Genotypes	g-1	SSg	Mg	$\sigma^2e+r\sigma^2g$	Mg / Me
Error	(r-1)(g-1)	SSe	Me	σ^2e	
Total	(rg-1)				

Where,

r = number of replications

g = number of genotypes

Mr= mean sum of squares due to replication

Mg= mean sum of squares due to genotype

Me = mean sum of squares due to error

'F' test:

The variance due to genotypes and due to replications was tested against error variance (Me) for significance.

't' test:

The standard error of difference SE (d) and critical difference (CD) between two means were calculated for genotypes comparisons as under:

S.E. (d) = Sqrt (2Me/r)

CD = S.E. (d) x 't' value (at 5% at error df)

Analysis of variance (ANOVA) is crucial in plant breeding for partitioning total variation into different components like genotypic and environmental variations. This technique helps researchers in plant breeding trials to assess the significance of treatment differences i.e. differences among set of genotypes, estimate variance components, and evaluate genotype-environment interactions, which are vital for developing superior crop varieties through following strategic applications

(i) By understanding the sources and magnitude of variation/differences, breeders can make more informed decisions about which traits to select for, which parents to cross, and which breeding methods to apply.

(ii) In case, if a character/trait is highly influenced by environmental factors, plant breeders might focus on developing varieties that are more stable across different sets of environment.

(iii) If a trait is primarily determined by genetic factors, breeders might focus on selection and breeding methods that can efficiently exploit this genetic variation governed by genes.

B] Analysis of dispersion and Wilk's criterion

The Wilks' lambda (^) is a test statistic that's reported in results from Multivariate Analysis of Variance (MANOVA), discriminant analysis, and other multivariate procedures. MANOVA is a statistical test used to compare means of two or more groups when there are multiple continuous dependent variables. It expands on the concept of ANOVA by handling situations with several response variables simultaneously. The features of these techniques are

(i) In MANOVA, '^ - tests' if there are differences between group means for a particular combination of dependent variables. It is similar to the F-test statistic in ANOVA. Lambda is a measure of the percent variance in dependent variables not explained by differences in levels of the independent variable. A value of zero means that there isn't any variance not explained by the independent variable (which is ideal). In other words, the closer to zero the statistic is, the more the variable in question contributes to the model. You would reject the null hypothesis when Wilk's lambda is close to zero, although this should be done in combination with a small p-value.

(ii) In discriminant analysis, Wilk's lambda tests how well each level of independent variable contributes to the model. The scale ranges from 0 to 1, where 0 means total discrimination and 1 means no discrimination. Each independent variable is tested by putting it into the model and then taking it out — generating a Λ statistic. The significance of the change in Λ is measured with an F-test; if the F-value is greater than the critical value, the variable is kept in the model

The analysis of dispersion was used to test the significance differences in the mean values between the genotype for the aggregate of fourteen characters ($X_1 - X_{14}$) based on Wilk's criterion as described by Rao, (1952).

$$\wedge = \frac{\text{Determinants of error variance and covariance matrix}}{\text{Determinants of (genotype+error) variance and covariance matrix}}$$

The determinants were obtained by following the pivotal condensation method and multiplying each of the pivotal elements of the corresponding matrices. Further 'V' statistic was calculated below.

$$\text{'V' (Stat)} = - m \log_e \wedge$$

Where,

$$\text{'}m = n - \frac{\text{'}p+q+1}{2}$$

p = Number of characters
q = Number of genotypes
n = d.f. for error + d. f for genotypes
e = 2.7183

The calculated 'V' (Stat) was tested at pq d.f at 5% level of significance.

C] Estimation of Mahalanobis distance i.e. D^2 values

Mahalanobis distance is the distance between a point and a distribution; and not between two distinct points. It is effectively a multivariate equivalent of the Euclidean distance. It was introduced by Prof. P. C. Mahalanobis in 1936 and has been used in various statistical applications ever since. Mahalanobis distance helps identify outliers in multivariate datasets by considering the relationships between variables. It is particularly useful when dealing with correlated variables, as it adjusts for these correlations when calculating the distance.

The inverse matrix of original genotypic variance covariance matrix was computed to derive the relationship by which the original character mean ($X_1 - X_{14}$) were transformed to an uncorrelated set of variables $Y_1 - Y_{14}$. In term of variance and covariance, the D^2 values were obtained as follows (Mahalanobis, 1936)

$$D^2_P = \sum_{i=1}^{P} \sum_{j=1}^{P} W_{ij} \left[\bar{X}^1_i - \bar{X}^2_i\right] \left[\bar{X}^1_j - \bar{X}^2_j\right]$$

Where

$[\overline{X^1_i - X^2_i}]$ = the mean of i^{th} character in the first and second genotype, respectively

W_{ij} = Inverse of variance and covariance matrix.

The D^2 value obtained for a pair of population is taken as the calculated value of χ^2 and is tested against the tabulated value of χ^2 for 'p' degrees of freedom, where 'p' is the number of characters.

From interpretation point of view, a larger D^2 value indicates a greater distance from the mean of the distribution, suggesting the data point is more likely to be an outlier. Whereas, the D^2 value can be transformed into a p-value using a chi-squared distribution with degrees of freedom equal to the number of variables, to assess the statistical significance of the distance

D] Grouping of genotypes into different clusters

The grouping of genotypes was done by Tocher's method as described by Rao (1952). The criterion used in clustering was that any two varieties belonging to the same cluster should at least on an average show smaller D^2 values than those belonging to different clusters.

E] Average intra and inter – cluster distances and cluster means

The average intra – cluster distance was obtained by using following formula

$$\text{Average intra} - \text{cluster distance } D^2 = \frac{\sum_{i=1}^{n} D_i^2}{n}$$

Where ΣD_i^2 = the sum of distance between all possible combinations (n) of genotypes involved in a cluster.

The average inter – cluster distance was calculated using all possible D^2 values between genotypes included in any two clusters. The cluster means were calculated for individual character on the basis of mean performance of the genotypes included in that cluster.

F] Character contribution towards divergence:

For trait based selection program, it very important to understand the contribution of individual trait in diversity among candidate genotypes under study. All the possible combination of genotypes [n (n-1)/2], each trait was ranked based on di values ($y_i^1 - y_i^2$). The highest mean difference was rank first and lowest mean difference was rank "p", where "p" is total number of traits were studied.

$$\text{Contribution of trait (\%) towards divergence} = \frac{\text{No of times appearing first in ranking by } X_i}{100\ [n(n-1)/2]}$$

Where, X_i = i th character and n = number of genotypes

G] Selection of parents for hybridization program

The concept of genetic diversity is very important in differentiating a population. It is a prerequisite for initiating a hybridization program because the choice of potent and diverse parents determines the success of such programs. Selection of diverse parents in crossing program serve the purpose of combining desirable genes to obtain superior recombinations.

Selection of parents for hybridization from different clusters was done on the basis of mean statistical distance suggested by Bhatt (1970).

$$\overline{D} = \frac{\Sigma \, di}{n\,(n+1)/2}$$

Where,
D = mean statistical distance.
Σ di = sum of all D^2 values between all intra and inter – clusters.
n = number of clusters.

The crosses between parents belonging to different clusters having same or higher inter – cluster distance than the mean statistical distance may be attempted.

H] Applications of D^2 statistic in plant breeding

In essence, the Mahalanobis distance i.e. D^2 values provides a powerful statistical tool to understand and establish the relationships between data points and distributions, especially when dealing with multivariate data. The D^2 used in plant breeding through various aspects viz.,

As a tool to measure Genetic Divergence/distances among the large population/genpool

Basis for selecting Parents for Hybridization

Exploitation of phenomenon of Heterosis through crossing genetically diverse parents

Helpful as an effective technique for grouping genotypes into clusters based on their genetic similarity, which helps in selecting diverse parents from different clusters.

Overall, the D^2 statistic is a reliable Estimates of Genetic Divergence compared to other methods.

The trait-based improvement strategies can be plan by improving the selection based on diversity

This techniques helpful for identifying and utilizing the diverse genetic resources in planned manner to enhance to efficiency of germplasm utilization

Although, there are reports of applicability and usefulness of D^2 statistics in breeding program has long list. Few important reports in leading crops are summarized in Table 1

Table 1: Application of D^2 statistics techniques in important crops by research workers

Crop	Workers
Black gram	Jayamani and Sathya (2013), , Reni et al (2013), Panigrahi et al (2014), Kuralasaran et al (2018), Chowdhury (2020), Suchitra et al (2023) and Chauhan et al (2025)
Chickpea	Manivannan (2002), Dwivedi and Lal (2009), Gaikwad et al (2014), Nimbalkar et al (2017), Santosh et al (2017), Tamvar et al (2019) and Gamit et al (2024)
Cotton	El-Moghny Ahmed et al (2015), Anjani et al (2018), Gnanasekaran et al (2018), Bhimate et al (2019), Ujjainkar and Patil (2020), Ishaq et al (2021), Jyothi et al (2021), Satish (2021), Suman Parre and Patil (2021), Dashmukhe and Ujjainkar (2022), Meena et al (2022), Ujjainkar (2022), Rani Chapra et al (2023) and Surekha Salve et al (2023)
Fingermillet	Shinde et al (2013), Negi et al (2017), Subramaya and Ravikumar (2020), Meera (2021), Reddy and Vengadessan (2022), Bharathi et al (2023) and Patel et al (2024),
Green gram	Sridhar et al (2022), Udayasri et al (2022), Bindu et al (2023), Jadhav et al (2023) and Laxmi Kanwar et al (2024)
Groundnut	Dolma et al (2010), Waghmode et al (2017), Hampannavar and Khan (2018), Shruti et al (2019), Abadya et al (2021), Dhutatra et al (2022) and Rathore et al (2022), Yadav et al (2022)
Maize	Udaykumar et al (2013), Suryanarayana et al (2017), Jakhar et al (2018), Punya sharma et al (2020), Hussain et al (2021), Patil et al (2021) and Rumit Patel et al (2024)
Pigeonpea	Upadhaya et al (2007), Patel et al (2018), Verma et al (2018), Zavinon et al (2019), Kandarkar et al (2020), Zavinon et al (2020), Naing et al (2022) and Baria et al (2024)

Sesame	Kadir et al (2001), Ujjainkar et al. (2002), Jadhav and Mohrir (2013), Swathy et al (2018), Arpitha et al (2019), Hukumchand and Parmeshwarappa (2019), Gokulakrishnan et al (2019), Ramprasad et al (2019), Sirisha et al (2020), Swapna et al (2023) and Chanchal Shakyawal et al (2024)
Sorghum	Doijad et al (2016), Tesfaye (2017), Prasad et al (2018), Kavithamani et al (2019) Prasad and Sridhar (2019), Prashanth et al (2021), Verma and Biradar (2022) and Bhadouriya and Bhadouriya (2024)
Soybean	Sareo et al (2018), Joshi et al (2018), Darai et al (2020), Nag et al (2020), Upadhay et al (2020), Tigga and Nag (2021), Beyene and Jalata (2022) and Banerjee et al (2023),
Wheat	Ali et al (2017), Fouad (2020), Pandey et al (2021), Chaudhary et al (2022), Neeru et al (2022) and Lohitha et al (2024)

Although, its cleared from review of literature of workers using Mahalanobis D^2 statistics techniques to assess the genetic diversity among germplasm for their further utilization in improvement programs. The ulility of this about century old statiscal tool is always highlighted by crop breeders for screeing large material.

I] Merits of Mahalanobis D^2 statistics

D^2 statistics helps in quantification of Genetic Divergence: The D^2 analysis provides a numerical measure (the D^2 value/estimates) that quantifies the genetic distance between different groups.

It helps to identify the promising parents: By identifying groups with high D^2 values, breeders can easily select the parents for hybridization that are likely to produce offspring with a wide range of desirable traits with the hypothesis of better performance of heterozygous individuals

D^2 statistics facilitates the handling of multiple quantifiable characters in single analysis: D^2 analysis can incorporate multiple traits/quantitative characters simultaneously in single analysis, providing a more comprehensive assessment of genetic diversity than methods that consider traits individually. It also represent the overall diversity estimates for set of traits under study.

Effective technique for screening of Large Germplasms: D^2 statistics allows for the evaluation of a large number of germplasm lines at once based on important traits.

Reliable Estimates: Breeding decisions mainly based on estimates and interpretations of data obtained by analysis. D^2 provides reliable estimates of genetic distances, which can be immensely helpful for making breeding decisions regarding genotypes.

Helpful for understanding the Genetic Relationships within and among the genepool: The D^2 statistics analysis helps understand the relationships between different groups, revealing which are more or less related based on statistical distances (D^2 values/distances).

J] Demerits of D^2 statistics

Complexities in calculations: It's observed that D^2 statistics involves long calculations, matrix, inverses of matrices etc., making it difficult. The calculations involved in D^2 analysis can be complex, when dealing with a large number of traits and genotypes so the expertise is essential.

Sufficient Data: D^2 analysis requires a substantial amount of data, including measurements for all traits on all genotypes being compared. The more data is better to predict the exact quantification of diversity among the members of experimental material. Moreover, the unreplicated data is not suitable for this analysis.

Non realistic assumptions: In field experiments, many conditions are not in control of research workers. The analysis relies on certain assumptions about the data, such as multivariate normality, which may not always be met in real-world scenarios or in other words it's impossible in reality for experimentations under the vagaries of environmental conditions.

Problems of Statistical Robustness: The estimates obtained from D^2 analysis may not always be statistically robust, especially if the data is not well-behaved and irrelevant data is analyzed.

Interpretation Challenges: Breeders mostly keep distance with D^2 statistics due to exaggerated data tables/data values and subsequent interpreting the results of D^2 analysis can sometimes be challenging, especially when dealing with complex datasets or when the genetic relationships are not straightforward.

Computational Resources: Analyzing large datasets with D^2 can require significant computational resources which makes it challenging.

CONCLUSION

Selection of parents based on *per se* performance, eco–geographical diversity had a limited success in past. However, several methods of multivariate analysis and metroglyph analysis have been found to be useful in selecting parents in crossing program. However, to quantify genetic divergence and identify parents with desirable traits for breeding programs is efficiently achieved through Mahalanobis D^2 statistics, which uses a numerical approach based on second-order statistics to measure the force of differentiation among genepool. It's also observed that this approach is widely used in plant breeding to select genetically diverse parents for crosses as it efficiently express and quantify the degree of divergence between groups, and the contribution of each trait to the overall divergence can be determined. It may be concluded that the D^2 statistic is a significant instrument for plant breeders to measure genetic diversity, select parent plants for hybridization, and ultimately produce enhanced crop varieties with superior traits.

Declaration by Author

Acknowledgement: None
Source of Funding: None
Conflict of Interest: No conflicts of interest declared.

REFERENCES

1. Aditya K.B. and Ashwini K.N. (2021) P.C. Mahalanobis : Mathematical Contributions to Statistics., International Advance Journal of Engineering, Science and Management (IAJESM) Vol (16) No. I ., 111-114
2. Ali, T., Singh M.K., Bharadwaj D.N. and Singh L. (2017). Analysis of genetic divergence in wheat (*Triticum aestivum* L.). Environ. Ecol., 35(3B), 2081-2083
3. Allard, R.W. (1960). Principles of plant breeding., John Willey and Sons., Inc., New York.
4. Anjani A, Padma, V Ramana J V and Satish (2018) Evaluation of genetic diversity in American cotton (*Gossypium hirsutum* L.). *Int J Cur Microbiol Appl Sci.*; 7: 3905-3911.
5. Arpitha, G. D., S. Manonmani, P. L. Viswanathan and M. Raveendran (2019) Genetic diversity study in germplasm lines of sesamum (*Sesamum indicum* L.), Electronic Journal of Plant Breeding, 10(2): 772-777.
6. Arunachalam, V. (1981). Genetic distance in plant breeding. *Indian J. Genet*., 41: 226 – 236.
7. Baker, R. J. (1984). Quantitative genetics principles in plant breeding. In Gustafsan J.P. (Ed), Gene manipulation in plant improvement I. Plenum press, New York: 147 – 176.
8. Baria K., S.S. Patil and Hemali Pandya (2024) Genetic divergence studies in pigeon pea [*Cajanus cajan* (L.) Millsp.] International Journal of Statistics and Applied Mathematics 2024; SP-9(1): 334-337
9. Beyene, D. G. and Jalata, Z. (2022). Diversity of soybean [*Glycine max* (L.) Merrill] genotypes based on agromorphological parameters. Journal of Pure and Applied Agriculture, 7(2), 30-37
10. Bhadouriya G.P.S and N.S. Bhadouriya (2024) D^2analysis and principal component analysis to assess genetic divergence in *Sorghum bicolor* (L.) moench for yield and its related attributes International Journal of Research in Agronomy 2024; 7(1): 322-326
11. Bharathi G., K. Radhika, N. Anuradha, K. Jayalalitha and V. Srinivasa Rao (2023) Genetic divergence studies in finger millet *Eluesine coracana* L. Gaertn. genotypes. Electronic Journal of Plant Breeding, 14 (1), 354-359.
12. Bhatt, G.M. (1970). Multivariate analysis approach to selection of parents for

hybridization aimed at yield improvement in self-pollinated crops., *Aust. J. Agric. Res.,* 21 : 1 – 7.

13. Bhatt, G.M. (1973). Comparison of various methods of selecting parents for hybridization in common bread wheat. *Aust.J.Agric.Res.,* 24: 457 – 464.
14. Bhimate, M., Deshmukh, S., Rathod, D., Jadhav, P. and Ujjainkar, V. (2019). Assessment of genetic diversity in cotton (*Gossypium hirsutum* L.) genotypes using morphological traits. Multilogic in science, 8(A): 142- 147.
15. Bindu, B. H., Satyanarayana, N. H., Babu, J. D. P. and Ramesh, D. (2023). Genetic divergence study in green gram [*Vigna radiata* (L.) Wilczek]. Electronic Journal of Plant Breeding, 14(1), 336-342.
16. Chanchal Shakyawal, Manju Singh, Rishabh Shukla.(2024) Determination of Genetic Diversity in Sesame (*Sesamum indicum* L.). Journal of Experimental Agriculture International, 46 (10), pp.136-141.
17. Chaudhary, H., Jaiswal J.P., Kumar A. and Joshi S. (2022). Determination of genetic variability and diversity in bread wheat for yield and Yield contributing traits. Int. J. Plant Soil Sci., 34(19), 16-23.
18. Chauhan M.P., Mishra A.C., Singh Kumar Ashok (2025). Genetic divergence studies in urd bean (*Vigna mungo* l.). Legume Research. 31(1): 63 - 67.
19. Chowdhury T., Arpita Das, G. S. Mandal, Sudip Bhattacharya and Shayree Chatterjee. (2020) Genetic Variability, Character Association and Divergence Study in Urdbean [*Vigna mungo* (L). Hepper]. Int.J.Curr.Microbiol.App.Sci. 9(02): 1726-1734.
20. Darai, R., Dhakal, K. H. and Sah, R. P. (2020). Genetic variability of soybean accessions for yield and yield attributing traits through using multivariate analysis. International journal of Horticulture, Agriculture and Food science, 4(3), 108-125.
21. Dashmukhe Akshay R. and Vaibhav V.Ujjainkar (2022) Dispersion Analysis studies in Desi Cotton (*Gossypium arboreum* L.) based on Wilk's Criteria and Canonical Vectors International Journal of Scientific Research in Engineering and Management (IJSREM) Vol. 06 (12) 1-9
22. Doijad S.B., A. B. Bagade and A. W. More (2016). Evaluation of sorghum germplasm for genetic diversity using D^2 statistics. Elect. J. Plant Breed., 7: 934-938.
23. Dwivedi K.K. and Lal G.M.(2009) Assessment of genetic diversity of cultivated chickpea (*Cicer arietinum* L.). Asian Journal of Agricultural Sciences;1(1):7-8.
24. Elliot, A. (2006). Statistical Analysis Quick Reference Guidebook: With SPSS Examples 1st Edition. SAGE.
25. El-Moghny Ahmed, S. M. Martiz and H. A. G. Reham. (2015). Nature of genetic divergence among some cotton genotypes., Journal of Cotton Science., 19 (3) : 368 – 374
26. Fisher, R. A. and Yates, F., (1953). Statistical tables for agricultural, biological and medical research. Edinborough, Oliver & Boyd.
27. Fouad, H. (2020). Principal component and cluster analyses to estimate genetic diversity in bread wheat (*Triticum aestivum* L.) genotypes. J. Plant Prod., 11(4), 325-331
28. Gaikwad A.R., Desai N.C., Pawar G.H., Langhi A.M.(2014) Genetic divergence in chickpea. Int J Agric Sci ;10(2):812-815
29. Gamit R.R., Savaliya J.J., Hadiya A.M., Vyas U.M. and Pargi K.L.(2024) Genetic divergence in Chick pea (*Cicer arietinum* L.) International Journal of Statistics and Applied Mathematics 2024; SP-9(1): 184-187
30. Gnanasekaran M, K Thiyagu and M Gunasekaran.(2018) Genetic variability heritability and genetic advance studies in cotton (*Gossypium hirsutum* L.) Elect J Plant Breed.; 9(1) : 377 – 382.
31. Gokulakrishnan, J., B. Priya and S. Saraswathy (2019) D^2analysis for estimating genetic divergence in sesame genotypes (*Sesamum indicum* L.) Journal of Emerging Technologies and Innovative Research, 6(2): 931-939.
32. Gomez, A.A. (1966) Relationship of heterosis and genetic diversity in varietal crosses of maize. Proc. Third Inter - Asian Corn Improvement workshop, India.
33. Hukumchand and S. G. Parameshwarappa (2019) Genetic diversity analysis for quantitative traits in advanced breeding lines of sesame (*Sesamum indicum* L.), International Journal of Current Microbiology and Applied Sciences, 8(12): 2970-2979.
34. Hussain K, RS Sudan, Praveen Singh, Manmohan Sharma and Mahrukh (2021) Genetic Divergence Study in Maize (*Zea*

mays L.) Inbred Lines under North Western Himalayan Conditions., Indian J. Plant Genet. Resour. 34(2): 274-278

35. Ishaq M, Hassan A, Munir S, Shahzad A, Anjam M, Bhutta M and Qureshi M K. (2021) Effect of heritability, genetic advance and correlation on yield contributing traits in upland cotton. J Agric Sci.; 27(3):353-359
36. Jadhav R. S. and M. N. Mohrir (2013) Genetic divergence analysis in sesame (*Sesamum indicum* L.) Electronic Journal of Plant Breeding, 4(1): 1090- 1092.
37. Jadhav, R. A., Mehtre, S. P., Patil, D. K. and Gite, V. K. (2023). Multivariate analysis using D2 and principal component analysis in mung bean [*Vigna radiata* (L.) Wilczek] for study of genetic diversity. Legume Research-An International Journal, 46(1), 10-17
38. Jakhar D.H., Ray A, and Singh R.(2018) Genetic Diversity Analysis in Maize (*Zea mays* L.) by D^2 Clustering of Yield and Yield Accredited Traits. International Journal of Agriculture, Environment and Biotechnology; 833-836.
39. Jayamani P. and Sathya M. (2013) Genetic diversity in pod characters in blackgram. Legume Research.;36(3):220-223
40. Joshi, D., Pushpendra, S. K., and Adhikari, S. (2018). Study of genetic divergence in soybean germplasm. Chemical Science Review and Letters, 7(26), 533-539.
41. Jyoti Gauswami, Valu M G, Odedara Geeta N.(2021) Correlation, path coefficient and D^2 analysis study of seed cotton yield and fibre quality traits in American cotton (*Gossypium hirsutum* L.). J. Pharmacognosy and Phytochemistry.; 10(3): 222-230.
42. Kadir, M. M., M. M. Ali, M. N. Islam and A. S. M. Golam Masum Akond (2001) Genetic diversity in sesame, Bangladesh Journal of Agricultural Research, 26(1): 131-135.
43. Kandarkar K.G., Kute N.S., and Shinde A.I.(2020) Genetic divergence studies in pigeon pea [*Cajanus cajan* (L.) Millsp.]. Int. J Chem. Studies.;8(4):2067-2070.
44. Kavithamani D., Yuvaraja A., and Selvi B.(2019) Principal component analysis and grouping of sorghum (*Sorghum bicolor* L. Moench) gene pool for genetic diversity. Electronic Journal of Plant Breeding.;10(4):1426-1434.
45. Kuralarasan V., Vanniarajan C., Kanchana S., Veni K., Anandhi Lavanya S. (2018) Genetic divergence, heritability and genetic advance in mutant lines of urdbean [*Vigna mungo* (L.) Hepper]. Legume Res. ;41:833-836
46. Laxmi Kanwar, Khajan Singh, P.K.P. Meena, C.B. Meena, M.K. Sharma and Prachi Singh (2024). D^2 Statistic Analysis among Green gram [*Vigna radiata* (L.) Wilczek] Genotypes. Biological Forum – An International Journal, 16(6): 72-74.
47. Lohitha B, A. Dutta and A. Maji (2024) Genetic divergence of wheat (*Triticum aestivum* l.) Genotypes Plant Archives Vol. 24 (2), 295-298
48. Mahalanobis, P. C. (1936). On the generalized distance in statistics. Proceedings of the National Institute of Sciences of India, 2(1), 49-55
49. Mahalanobis, P.C. (1928). A statistical study of Chinese head measurement *Man in India.,* 8 : 32 – 64.
50. Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proc.National Institute. Sci. India* 2 : 49 – 55.
51. Manivannan N. (2002)Genetic Diversity in Chickpea (*Cicer arietinum* L.). Int. J Curr. Microbiol. App. Sci. ;6:904-910.
52. Meera, G. (2021) Genetic diversity and studies on nutraceutical importance in white and brown finger millet (*Eleusine coracana* (L).). M. Sc (Ag) Thesis. Acharya N.G. Ranga Agricultural University, Guntur, Andhra Pradesh, India.
53. Nag S.K., Painkra C., and Tiwari A. (2020) Genetic variability, heritability and genetic advance in soybean [*Glycine max* (L.) Merrill]. Int J Curr Sci.;8(1):3124-3126
54. Naing, T., Verma, S.K., Yadav, H., Chauhan, C., Gautam, A., Karn, A. and Gaur, A.K. (2022). Assessment of morphological and molecular genetic diversity in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. Legume Research. 45(9): 1088- 1093.
55. Nath, R. and Pavur, R. (1985) A new statistic in the one way multivariate analysis of variance, Computational Statistics and Data Analysis, 2, 297–315.
56. Neeru ., Vikram Singh, I. S. Panwar, . Anu, Satender Yadav, D. K. Janghel, and . Kiran. (2022). "Analysis of Genetic Diversity by D2-Statistics Using Morpho-Physiological Traits of Late Sown Wheat (*Triticum aestivum* L.)". *International Journal of Plant & Soil Science* 34 (12):66-72

57. Negi S, Kumar V. and Bhatt A. (2017) Genetic diversity among finger millet [*Eleusine coracana* (L.) Gaertn] genotypes for yield and its contributing traits. Int J Curr Microbiol Appl Sci.; 6 (8):3332-3337.
58. Nimbalkar R.D., Katre Y.Y.,and Phad D.S.(2017) Genetic diversity in chickpea (*Cicer arietinum* L.). Bioinfolet..
59. Pandey, J., Tripathi R.M., Singh A., Singh V. and KU U. (2021). Genetic diversity and principal component analyses for yield and yield components of advanced lines of wheat (*Triticum aestivum* L.). The Pharma Innov. J., 10(7), 998-1003.
60. Panigrahi, K. K., Mohanty, A. and Baisakh, B., (2014). Genetic divergence, variability and character association in landraces of blackgram (*Vigna mungo* [L.] Hepper) from Odisha. J. Crop Weed. 10 (2): 155- 165.
61. Panse, V.G. and P.V.Sukhatme. (1967) Statistical method for agricultural workers, New Delhi, ICAR Publication.
62. Patel D.A., Mahesh B Vaja, Rumit Patel and Priyanka Solanki (2024) Phenotypic diversity using D2 statistics in finger millet [*Eleusine coracana* (L.) Gaertn.] International Journal of Advanced Biochemistry Research ; 8(3): 947-951
63. Patel H.N., Patel A.M., Patel J.M. and Patel N.I.(2018). Genetic divergence study through D2 analysis in pigeon pea [*Cajanus cajan* (L.) Millsp.]. Int. J Chem. Studies.;6 (5):1947-1950.
64. Patil N.L., Shanthakumar G and Biradar B.D.(2021). Genetic diversity analysis of Maize (*Zea mays* L.) inbred lines under drought stress using grain yield and yield related traits. Journal of Crop and Weed.;17(3):192-198.
65. Prasad B. H. V., B. D. Biradar and L. K. Verma (2018). Genetic diversity studies among maintainers and restorers on milo and maldandi cytoplasm from minicore collection of sorghum using D2 statistics. Elect. J. Plant Breed., 9: 233-243.
66. Prasad B.V. and V Sridhar.(2019) Diversity Studies in Yellow Pericarp Sorghum [*Sorghum bicolor* (L.) Moench] Genotypes for Yield Attributes. Int J Curr Microbiol App Sci. ;8(12):361-366
67. Prasanth A, Premnath A and Muthurajan R.(2021) Genetic divergence study for duration and biomass traits in sorghum [*Sorghum bicolor* (L.) Moench]. Electronic Journal of Plant Breeding.;12(1):22-27.
68. Punya Sharma V.K., Pankaj Kumar and Anjani Kumar Singh (2020) Genetic divergence studies and heterotic grouping in maize inbred lines using microsatellite markers. Ind. J. Plant Gene. Resor. 33(2): 209-216
69. Qualset, C.O. (1979). Mendelian genetics of quantitative characters with reference to adoptation and breeding of wheat. In: Ramanjan S. (Ed) *Proc. 5th International Wheat Genetic Symposium., Indian Soc. Genetics and Plant Breeding.,* Vol. 2: 577 – 590., New Delhi India.
70. Ramprasad E., S. Senthilvel, J. Lal Jatothu, K. N. Yamini and K. Singh Dangi (2019) Analysis of genetic diversity in sesame (*Sesamum indicum* L.) advanced breeding lines and varieties collected from major breeding centers in India, Electronic Journal of Plant Breeding. 10(3): 1269-1274.
71. Rani Chapara, K V Siva Reddy, M Sudha Rani, B Sree Lakshmi, V Roja and J Pranaya.(2023) Variability studies and genetic divergence in cotton (*Gossypium hirsutum* L.) germplasm using multivariate analysis. Electronic Journal of Plant Breeding.;13(4): 1305-1311.
72. Rao, C.R. (1952). Advanced statistical methods in Biometrical Research., Edn. John Willey and Sons. Inc., New York.
73. Rao, C.R. (1969) Multivariate analysis on indispensible statistical and applied research, *Sankhya.,* 22 : 317 – 338.
74. Rathore, M.S., Tiwari S., Tripathi M.K., Gupta Neha, Yadav S. and Singh S. (2022). Genetic diversity analysis of groundnut germplasm lines in respect to early and late leaf spot diseases and biochemical traits. Legume Res.-An Int. J., 1. 10.18805/ LR-4833
75. Ravindra Kumar Meena, Ram Kishor Meena, Prakashkumar Ranchhodbhai Patel, Laxmansingh Dolatsingh Parmar, Jaishree Tanwar, Ganesh Kumar Koli, Deepak Kumar, Om Prakash Yadav and Sandeep Kumar Pancholi (2022). Insights of Genetic Divergence through Mahalanobis' D^2 Statistics among American Cotton (*Gossypium hirsutum* L.) Genotypes. Biological Forum – An International Journal, 14(2): 790-795.
76. Reddy, B.T. and Vengadessan, V. (2022). Genetic divergence in finger millet (*Eleusine coracana* L.). Electronic Journal of Plant Breeding, 13 (1): 279-285.

77. Reni, Y. P., Rao, Y. K., Satish, Y. and Babu, J. S., (2013). Estimates of genetic parameters and path analysis in blackgram (*Vigna mungo* (L.) Hepper). Int. J. Plant Animal and Env. Sci. 3 (4): 231-234.
78. Rumit Patel, Dipak A Patel, Juned Memon, Arna Das and Kalyanrao Patil (2024) D^2 clustering of yield and yield accredited attributes for genetic diversity analysis in maize (*Zea mays* L.) The Pharma Innovation Journal 2023; 12(3): 2210-2213
79. Santosh H.B., Bhardwaj C., Santosh S., Raghavendra K.P., and Kumar J. (2017) Genetic diversity analysis for productivity enhancement through deshi kabuli introgression breeding in Chickpea. Elect. J Pl. Breed.;8(1):125-133.
80. Sareo, H., Devi, H. N., Devi, T. R., Devi, T. S., Karam, N. and Devi, L. S. (2018). Genetic diversity analysis among soybean [*Glycine max* (L.) Merrill] genotypes based on agro morphological characters. Indian Journal of Agricultural Sciences, 88(12), 1839-1842
81. Satish, Y. (2021). Genetic diversity analysis in American cotton (*Gossypium hirsutum* L.) genotypes. Electronic Journal of Plant Breeding, 12(4): 1448-1453
82. Shinde, S.R., Desai, S.V. and Pawar, R.M., (2013). Genetic diversity pattern in finger millet (*Eleusine coracana* (L.) Gaertn). Electronic Journal of Plant Breeding, 4 (3): 1242-1245
83. Singh, R.K. and B.D.Chaudhary. (1985). Biometrical methods in Quantitative Genetics Analysis. Ludhiana, Kalyani Publication.
84. Sirisha A. B. M., S. K. H. Banu and R. Saritha (2020) Genetic divergence studies using Mahalanobis D square analysis in sesame (*Sesamum indicum* L.) germplasm, International Journal of Current Microbiology and Applied Sciences, 9(8): 2224-2229
85. Sridhar, V., Sriram, A., Rao, S. S. and Gopal, G. V. (2022). Evaluation of green gram (*Vigna radiata* L.) germplasm for seed yield and yield related traits through cluster analysis. International Journal of Environment and Climate Change, 12(11), 3563-3574.
86. Subramanya A.S. and Ravikumar R.L.(2020) Genetic Divergence Studies in Cultivated Tetraploid Finger Millet [*Eleusine coracana* (L.) Gaertn] genotypes using D 2 analysis. Int. J. Curr. Microbiol. App. Sci. ;9(1):109-118.
87. Suchitra , Chitrada, and Gaibriyal M. Lal. (2023). Study on Genetic Diversity in Black Gram (*Vigna Mungo* L. Hepper). *International Journal of Plant & Soil Science* 35 (20):302-11
88. Suman Parre and Rajesh S Patil (2021). Assessment of genetic diversity using D^2 analysis among F_3 progenies derived from F_2 populations of intra-hirsutum cotton hybrids (*Gossypium hirsutum* L). Pharma Innovation;10(3):960-963.
89. Surekha Salve B., Minnu Sasi and Manisha Phaugat.(2023) Genetic diversity analysis in cotton using morphological traits (*Gossypium hirsutum* L.). Pharma Innovation;12(7):1089-1096
90. Suryanarayana, L., M. Reddi Sekhar, D. Ratna Babu, A.V. Ramana and Srinivasa Rao, V. (2017). Genetic Divergence Studies in Maize (*Zea mays* L.). Int.J.Curr.Microbiol.App.Sci. 6(7): 360-365
91. Swapna, B., Thippani, S., Donda, P. and Nayak, S. S. (2023). Genetic diversity for yield and yield contributing characters in sesame (*Sesamum indicum* L.). International JournalofEnvironment and Climate Change, 13(8):1717-1725.
92. Swathy, V., Premnath, A., Viswanathan, P. L., Raveendran, M. and Manonmani,S. (2018). Determination of genetic divergence based on morphological traits in sesame (*Sesamum indicum* L.). Electronic Journal of Plant Breeding. 9(2): 747 - 752.
93. Tamvar M.R., Patel S.R., Patil S.S., Naik A.J. and Patel H.N. (2019) Diversity study through D^2 analysis in chickpea. The Pharma. Inno. J.;8(9):140-143
94. Tesfaye K. (2017). Genetic diversity study of sorghum (*Sorghum bicolor* (L.) Moench) genotypes, Ethiopia .Acta Univ. Sapientiae, Agric. Environ., 9: 44-54.
95. Tigga A. and Nag S.K. (2021) Determination of significant characters for improving seed yield in soybean (*Glycine max* L. Merrill) via correlation and path coefficient analysis. Int J Agric Environ Biotechnol.;14(3):299-305.
96. Todorov, V. and Filzmoser, P. (2007) Robust statistic for the one-way MANOVA, submitted to the Journal of Environmetrics.
97. Udaykumar K, D Madalageri, L Malakannavar and P Ganagashetty (2013) Genetic diversity studies in newly derived

inbred lines of maize (*Zea mays* L.) Mol. Plant Breed. 4 (9): 77-83

98. Ujjainkar V. V. and M. W. Marawar (2020). Genetic Diversity: A tool for assessing potential of genepool., Remarking An Anlisation., 4 (11) 38 – 44
99. Ujjainkar V. V. and V. D. Patil (2020) Marker assisted selection in American cotton genotypes using biochemical and molecular profiling techniques, Int. J. Adv. Res., Ideas and Innovations Tech, Vol. 6 (3) 989 – 993
100. Ujjainkar V. V., S. B. Datke and T. H. Rathod. (2002) Genetic divergence studies in sesame (*Sesamum indicum* L.)., Annals of Plant Physiology., 16 (1) : 18 – 22
101. Ujjainkar Vaibhav V. (2022) Study of Environmental Influence on Genetic Diversity Estimates (D2) of Parental Lines in Upland Cotton (*Gossypium hirsutum* L.). Int. J. Curr. Microbiol. App. Sci. 11(07): 212-221
102. Upadhyay P, Shrivastava M.K., Amrate P.K. and Yadav R.B.(2020) Yield determining traits, genetic variability and character association in exotic lines of soybean [*Glycine max* (L.) Merrill]. Soybean Res.;18(2):95-102.
103. Upadhyaya, H.D., Reddy, K.N., Gowda, C.L.L., Sube, Singh. (2007). Phenotypic diversity in the pigeonpea (*Cajanus cajan*) core collection. Genet Resour Crop Evol. 54: 1167-1184
104. Vavilov, N. I. (1926). The origin, variation, immunity and breeding of cultivated plants (translation by K.S. Chester). Chronical Botany, 13, 1-364.
105. Verma L. K. and B. D. Biradar (2022) Genetic divergence study through D2 statistics in rabi sorghum [*Sorghum bicolor* (L.) Moench] Emer Life Sci Res 8(2): 178-185
106. Verma, S.K., Gaur, A.K., Bisht, C. and Chandra, D. (2018). Estimation of genetic diversity for yield and its component traits in pigeonpea using D^2 statistics. Journal of Hill Agriculture. 9(4): 383-386.
107. Weinig C. and J. Schmitt.(2004) Environmental effects on the expression of quantitative trait loci and implications for phenotypic evolution., BioScience., 54 (7) 627 – 635.
108. Wilks, S.S. (1932) Certain gernalizations in the analysis of variance. Biometrika 24 :471-494
109. Zavinon, F., Adoukonou-Sagbadja, H., Bossikponnon, A., Dossa, H. and Ahanhanzo, C. (2019). Phenotypic diversity for agro-morphological traits in pigeon pea landraces [*Cajanus cajan* L.) Millsp.] cultivated in southern Benin.Open Agriculture. 4(1): 487-499
110. Zavinon, F., Adoukonou-Sagbadja, H., Keilwagen, J., Lehnert, H., Ordon, F. and Perovic, D. (2020). Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. Genet Resour Crop Evol. 67: 191-208.

How to cite this article: Vaibhav Ujjainkar. 'D^2 statistics' technique: methodology and applications in plant breeding. *International Journal of Research and Review*. 2025; 12(8): 127-139. DOI: *10.52403/ijrr.20250815*

******