GENETIC DIVERSITY IN AGRICULTURAL PLANTS

Abstract

Genetic variety refers to the total number of genetic traits that comprise a species' genetic composition. It is distinct from genetic variability, which refers to the proclivity of hereditary characteristics to change. Crop diversity refers to genetic and phenotypic variation in agriculturally important plants. Plant breeders can develop new and improved cultivars with desirable characteristics such as many agronomical yield traits and disease resistance due to the diversity of agricultural plant genetic resources. Agricultural scientists are concerned about the loss of genetic diversity among cultivated plant species. Breaking yield constraints and crop plants will become challenging, and breeders will be unable to meet the demands caused by an increasing population. Climate change and other unforeseen events increase the importance of genetic diversity because it may be the source of several unique features that provide tolerance to various biotic and abiotic pressures. Genetic diversity leads to a variety of major agriculturally important phenomena, such as heterosis and transgressive segregation. As a result, the primary goals of any crop development programme are to identify genetically diverse lines, if they exist, or to create variations in the line and use them in subsequent research programmes. To use genetic diversity wisely in this context, it is necessary to understand all of its facts, including the variables influencing genetic variability, various methods of diversity analysis, measurement, and the software used for statistical data analysis. An effort is being made in this review to compile broadly practical ideas in the field of crop plant genetic diversity, which could be critical for advancing understanding and constructive research. This will enable agricultural researchers in their breeding efforts.

Keywords: Genetic diversity, agricultural plants, analysis, evaluation, Software,

Authors

Krishna Murari Prasad
Department of Genetics and Plant Breeding Institute of Agriculture
Visva Bharati, Sriniketan
West Bengal, India

Nihar Ranjan Chakraborty
Department of Genetics and Plant Breeding Institute of Agriculture
Visva Bharati, Sriniketan
West Bengal, India
nrchakraborty@gmail.com

I. INTRODUCTION

Crop plant diversification, whether it is anthropogenic (human) or innately, is essentially based on the population's genetic diversity. The degree of heterogeneity between or within species can be used to define diversity. All crop improvement systems are designed on the already-present intra- and inter-specific variations. If every member of the species had been the same, it's possible that there would be limited room for improvement in the way the plants performed across a range of qualities. Natural variation and divergence among crops have been extensively identified and employed to promote crop species ever since systematic plant breeding initially began. But as days progressed on, natural variability was reduced as a result of (i) biassed breeding techniques that focused on improving just a few traits (like yield and its component traits), (ii) frequent use of a small number of carefully chosen genotypes as parents in varietal development programmes, and (iii) introduction of a small number of exceptional lines to many countries, which increased genetic similarity between modern crop cultivars. Among agricultural workers, the diminishing genetic variation and variability among crop plant species has caused serious concern. It will be difficult to further strengthen agricultural types with less genetic diversity. Breaking yield limitations will become challenging, and plant breeders won't be able to keep up with the demands caused by an increasing population. Climate change and related unforeseen events make genetic diversity even more significant since it may be the source of several unique features that give tolerance to various biotic and abiotic pressures. Numerous major agriculturally significant phenomena, including as heterosis and transgressive segregation, are caused by genetic diversity. For the improvement of defects in commercial variations and the creation of new varieties, diverse lines are required. Therefore, the main objectives of any crop development programmes are the discovery of diverse lines (if available), creation of diversity (if not available or limited), and its subsequent utilisation. To make sensible use of genetic diversity in this context, it is essential to have knowledge of all its facets, including the variables influencing genetic diversity, various methods of diversity analysis, their measurement, and the software used for statistical analysis. Numerous reviews have been written with a focus on important issues like evolutionary divergence changes due to plant breeding, the genetic vulnerability of modern cultural crop cultivars, the conservation and use of genetic resources, and the assessment of genetic diversity using molecular markers and statistical tools. In the current review, an effort is being made to compile broadly applicable ideas in the field of genetic variety, which could be absolutely critical for advancing understanding and constructive research.

II. CONCEPT OF DIVERSITY

The core of the biological domain is diversity. Even maternal siblings don't exactly resemble one another in any way. Variability is the difference in one or more characteristics of the organism. Genetic variety and variability are incorrectly conflated in everyday speech, which is incorrect. Genetic variation is the diversity in DNA/RNA sequences or gene alleles within a species' or population's gene pool. This exhibits itself in several phenotypic forms. On the other hand, genetic diversification is a broad term referring to all the genetic diversity and richness within and across genotypes that are connected to a single species or between species. The amount of unique genes in a gene pool can be measured to determine genetic variety, but genetic variation cannot be counted and can only be anticipated to develop. Thus, genetic variability can be thought of as the foundation of genetic diversity. There are three

levels of diversity, according to the Convention on Biological Diversity (Figure 1). The ecosystems diversity, which represents variation among various communities of species, is situated at the top of the hierarchy. The diversity of species within a community, also known as species richness, is found at the next level of the hierarchy. The diversity exhibited within various genotypes of the same species is referred to as genetic diversity. This is brought on by various individuals developed different phenotypes caused by different alleles of the same gene. According to Swingland (2001), genetic variety is the variation in heritable traits found in a population of a single species. Heritable character variation may manifest as changed morphology, anatomy, physiological behaviour, or biochemical traits. A person's genetic variation can be classified as diversity at several gene-loci. The focus on genetic variety has been strongest among agricultural workers.
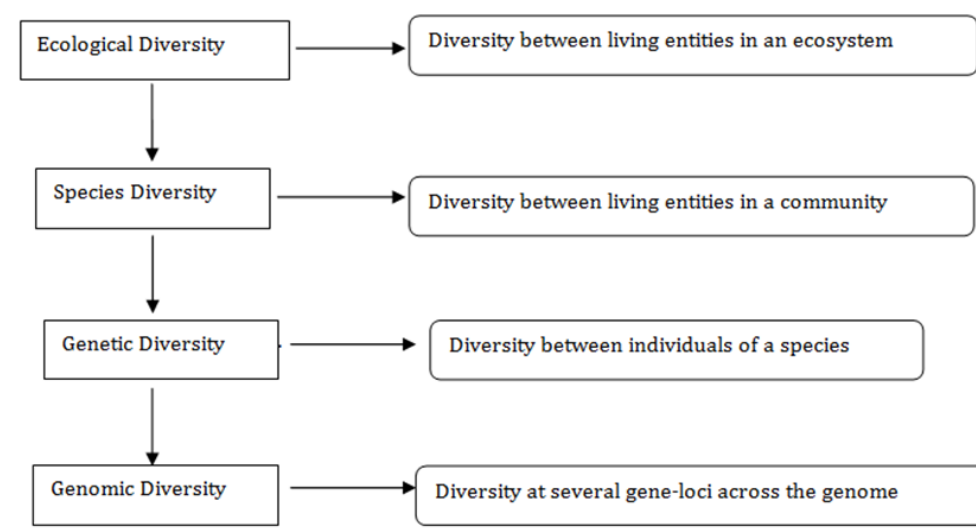

Figure 1: Hierarchy of Diversity.

III. IMPORTANCE OF GENETIC DIVERSITY

The underpinning for plant survival in nature and crop development is genetic diversification. Plant breeders can generate new and improved cultivars with desirable characteristics, especially farmer- and breeder-preferred aspects (high yield potential, large seed, etc.), attributable to the diversity of plant genetic resources (pest and disease resistance and photosensitivity, etc.). Natural genetic variability within crop species has been used since the dawn of agriculture to supply the demand for food for subsistence. Later, the emphasis turned to producing extra food to feed expanding populations. To ensure that people have a balanced diet, the yield and quality characteristics of important food crops are currently being prioritised. Breeding of climate adaptable cultivars is becoming increasingly crucial as the environmental situation changes. The presence of genetic variety, which can be found in wild species, closely related species, breeding stocks, mutant lines, etc., may act as a source of beneficial alleles and help plant breeders create varieties that are more tolerant to climatic change. Breeding plants that are climate resilient demands the development of unique features, such as resistance to potential new insect pests and diseases, high heat and cold, and a variety of air and soil contaminants. Different genes must be reserved in cultivated and cultivable crop species in the form of germplasm resources for ever-changing breeding objectives. Because there is genetic variability within and between crop plant species,

breeders can choose superior genotypes to either utilise as parents in hybridization programmes or to use directly as new varieties. To achieve heterosis and produce transgressive segregants, two parents must have genetic variety. Breeders can create varieties for specific traits like quality enhancement and resistance to biotic and abiotic challenges thanks to genetic diversity. Additionally, it makes it easier to generate new lines for non-traditional purposes such biofuel types of sorghum, maize, etc. Diversity is crucial for crop plants' ability to adapt too many habitats, particularly those with shifting climatic circumstances.

IV. CROP GENETIC DIVERSITY

Genetic diversity is essential for any crop improvement programme and the creation and management of genetic diversity is central base to crop breeding. The multivariate analysis by means of Mahalanobis's D2 statistic for estimating genetic divergence has been emphasized by many workers (Anand and Murty, 1968; Kole and Chakraborty, 2012; Gadi et al. 2020). The more diverse the parents, within overall limits of fitness, the greater are the chances of obtaining higher amount of heterotic expression in F1(Chakraborty and Bhattachraya, 2018; Sunny et al. 2022) and broad spectrum of variability in segregating generations (Anand and Murty, 1968, Singh et al, 2016a).

The distribution pattern of genotypes in different clusters indicated that genetic divergence was not related to geographical differentiation. Many genotypes of close geographic proximity fell into different cluster and vice-versa. Tendency to form such type of clustering ignoring the geographical boundaries showed the regional isolation was not the only factor contributing to diversity in natural population (Rao et al., 1980). Clustering of genotypes form different eco-geographic locations into one cluster could be attributed to the possibility of free exchange of breeding materials. However unidirectional selection, practiced for a particular trait or a group of linked traits in several places may produce similar phenotype, which can aggregate into one cluster irrespective of their geographic region (Singh and Gupta, 1968). Formation of different clusters among the genotypes of common geographic origin may be due to their parentage, developmental traits, past history of selection and different out-crossing rates (Arnold et al., 1996).

Considering genetic divergence and per se performance of genotypes as well as cluster mean, crossing between the intra-cluster genotypes are most likely to yield a considerable amount of heterosis in F1 generation and to provide a wide spectrum of recombinants in segregating generation (Sunny et al. 2022).

PCA is a statistical technique used to identify and eliminate duplicate genotypes with similar characteristics (Singh et al., 2016b). It allows for the natural classification of genotypes and gives an accurate indication of genotypic differences. The primary benefit of PCA is that each genotype may be assigned to only one group (Singh et al., 2016b; Debnath et al. 2022). In addition, this test is used to categorize a large number of variables into important components and assess their contribution to the total variance. So, PCA was done to determine how the various attributes were connected and to identify the traits that co-segregated.

Induced mutagenesis, particularly using gamma (γ) irradiation, plays an important role in generating genetic variability in crops including cereals, pulses and oilseeds for desirable traits of economic importance and these variants were used further in cross breeding programme of several seed propagated crops (Chakraborty and Paul, 2012) and also meets up challenges including world food and nutritional security (Kant et al. 2020).

Hybridization is important tools for generation of genetic variability. Thus, for greater success in transferring the desirable traits through hybridization, it is pertinent to have information on both genotypic and phenotypic variation together with heritability and genetic advance for the possibility of direct selection as well as prediction of the inheritable nature of the traits. Sunny et al (2022) worked on aromatic rice and finally selected some F1 which possessed high grain yield with semi-dwarf to medium stature, high tillering behavior, resistance to lodging and pleasant aroma.

V. FORCES AFFECTING GENETIC DIVERSITY

Natural recombination is the primary mechanism through which genetic variety is perpetuated. During meiosis, recombinations occur when homologous chromosomes cross throughout and establish new pairs. Plants' genetic diversity is driven by a range of environmental factors. Changes in allelic frequency in a community and their impact on genetic diversity are ongoing effects of evolutionary processes like selection, mutation, migration, and genetic drift. Some alleles become more common as a result of genetic manipulation during domestication, while others become less common. Therefore, domestication lessens genetic diversity relative to wild populations. Significantly influencing genetic diversity is natural selection too though. Selection that is both directional and stabilising tends to reduce genetic variety, while selection that is both asymmetric and disruptive tends to increase it. It has been suggested that mutation can significantly boost genetic diversity. Symptoms of qualitative mutation include dramatic alterations in structure and function at the molecular and cellular level. Unlike their more spectacular siblings, the qualitative mutations, quantitative or micro-mutations have modest and gradual impacts that build over time and produce alterations. It's possible that mutation also causes many chromosomal abnormalities. Genetic variety is manifested in the form of phenotypic variation, which is brought about by smaller sub-lethal or non-lethal abnormalities. The method by which agricultural plants reproduce also has an impact on the species' overall genetic diversity. While outbreeding enhances genetic variety, inbreeding decreases it. Reduced genetic diversity occurs when uncommon alleles are lost caused by genetic drift. A species' genetic variation is influenced not only by the number of individuals but also by their geographic dispersal. The greater the geographical separation between people, the less likely it is that they have the same DNA. Increases in genetic diversity can be achieved through the use of methods that lead to the production of unique phenotypes, such as wide-hybridization, hybridization across incompatible types, or introgression from previously isolated populations. Conversely, hybridization within the same species has a deleterious effect on genetic diversity. The introduction of novel alleles through gene flow within a population promotes genetic diversity.

VI. DIVERSITY ANALYSIS TECHNIQUES

Morphological, cytological, biochemical, and molecular characterization can all be used in a biodiversity analysis. As with their counterparts, morphological markers have continued to be used in biodiversity studies. These were all-natural variations of a single plant species. Ultimately, genetic diversity was measured in part by looking at cytological and biochemical variations within a species' genotypes. Since the emergence of genomic information, molecular markers have dominated efforts to quantify genetic differences.

1. **Indicators of morphology:** Germplasm lines, purelines, enhanced varieties, etc., are cultivated according to a predetermined experimental design for the intention of these analyses. Because morphological traits are so significant in deciding a plant's agronomic value and taxonomy classification, this process requires morphological characterization of numerous entries established in the field. Direct, low-cost, simple, and requiring no fancy equipment, morphological evaluations are a fantastic choice. However, it's costly because of the enormous amounts of land and human labourers needed over time. Their limitations in comparison to other methods include environmental sensitivity and subjective characterisation. These traits are typically dominant or recessive in nature, have a measurable biological impact, and render some physical variants impotent.

2. **Cytological indicators:** Research of cytological characteristics, such as chromosome size, secondary constriction in chromosomes, centromere position, arm ratio, constitutive heterochromatic patterns, banding characteristics (G, Q, R and N banding), DNA content, total genomic chromosome length, chromosome volume, etc., are embroiled. In maize, potato, lentil, radish, and other plants, several cytological traits have been used to evaluate genetic diversity within and between species. However, due to their scarcity and low resolution, these have only a few applications in genetic diversity analysis.

3. **Biochemical indicators:** It comprises the segmentation of proteins or isozyme polymorphisms into specified banding patterns. Not the genes, but distinct alleles' products are reflected in the isozymes. These isozymes can be utilised as genetic markers to map other genes and can be transferred onto chromosomes. This methodology of evaluating variety is quick and only needs a small sample of plant tissue. However, they are few in number, vulnerable to changes in the climate, and unable to be used to create a full genetic map.

4. **Molecular indicators:** It encompasses investigating genotype variation at the DNA/RNA level. Different molecular markers can be used for multiple reasons because of their unique features. They are usually categorized as PCR- and hybridization-based. New generations of markers based on platforms for sequences or arrays have recently been created. Based on their activity and expression, they can also be categorised as neutral markers, genetic markers, and functional markers. These indicators may also be based on differences in the genome, ribosomal RNA, or organelle genome sequences. In wheat, barley, apples, rice, pearl-millet, and other species, chloroplast microsatellites have been produced and used to measure genetic diversity at the intra-specific level. Contrarily, due to its quantitative scarcity, mitochondrial DNA from plants has been shown to be an inadequate tool for assessing genetic variation. Due to their great repeatability, superior genomic coverage, automation-friendliness, neutrality, and lack of environmental

sensitivity, molecular markers are the method of choice for assessing genetic variation. It has been observed that many genetic diversity investigations use both morphological and molecular markers at the same time.

VII. MEASURES OF GENETIC DIVERSITY

1. **Genetic base:** Coefficient of Parentage (COP) or Coefficient of Correlation are terminology used to describe the genetic foundation of every crop. These are disclosed by pedigree records of varieties that have been released and show how frequently a line appears in the commercial variations of a specific crop. COP is defined as the chance that alleles of two individuals are similar by descent. Less variability will be present in the segregating generations that arise from a cross between individuals with high COP, and vice versa. Coefficient of parentage values range from zero, where two cultivars have no genetic relationship at all, to one, where two cultivars share all of their alleles. Genotypes can be grouped into groups with a similar genealogy using the COP data matrix. The formula shown below (Falconer & Mackay, 1996) can be used to compute the coefficient of parentage (COP) or coefficient of correlation (rxy) for all pairwise combinations of genotypes.

$$r_{xy} = 2f_{xy}/\sqrt{(1 + F_x)(1 + F_y)}$$

Where Fx and Fy are the inbreeding coefficients of X and Y, respectively, and fxy is a co-ancestry coefficient. For the purpose of calculating the coefficient of parentage, multiple algorithms were devised by Delannay et al., Murphy et al., and Cox et al. Another comparable metric is called "Relative Genetic Contribution (RGC)," which is calculated by dividing a selection's genetic make-up into notional percentages attributable to several ancestors. The average of the relative genetic contributions made by a certain ancestor to all kinds that have been published is used to assess the ancestor's mean genetic contribution. Cumulative relative genetic contributions are produced over time by adding up the mean relative genetic contributions throughout time. The assumptions behind the relative genetic contribution measurement are (i) the unrelatedness of forebears and (ii) the equally likely transfer of 50% of parental genes to the children.

Numerous crop studies have shown that the released varieties of various crops in India have a limited genetic foundation. For instance, the lines IR-8 and TN-1 in rice, Spanish improved in groundnuts, Bragg in soybeans, T-1 and T-190 in pigeonpeas, and Pb-7 in chickpeas all commonly showed up in commercial varieties of the aforementioned commodities that were made available in India. The genetic diversity and genetic basis are roughly estimated by the frequency of emergence of specific lines.

2. **Genetic separation:** Nei was the first to define genetic distance as the distinction between two things that can be explained by allelic variation. The term "extent of gene differences among populations that are quantified using numerical values" was later added to this definition (1987). A more thorough definition of genetic distance was offered by Beumont et al. as any quantitative measure of genetic difference calculated between genotype individuals or groups, either at the sequence or allele frequency level. Simply said, there is less genetic distance between genotypes that share a lot of genes. The most popular statistic for determining the genetic separation between individuals (genotypes or populations) from morphological data is the euclidean or straight-line

measure of distance. In-depth descriptions of several genetic distance measurements have been provided by Mohammadi and Prasanna. The following definition of the Euclidean distance between two genotypes is mathematical:

$$d\,(a,b) = \sum_{i=0}^{n} \sqrt{[Xi - Yi]}x^2$$

Where, d (a, b) is the Euclidean distance between genotype a and b; Xi is the observation on ith phenotypic character, and Yi is the observation on ith phenotypic character.

Smith et al. developed another measure of genetic diversity in inbred lines which can be expressed as below:

$$d\,(a,b) = \sum \sqrt{[(X1_{(i)} - X2_{(i)})^2 / Var\,X_{(i)}]}$$

Where, d (a, b) is the Euclidean distance between genotype a and b; X1 and X2 are the values for ith trait for inbred lines a and b and Var X(i) is the variance for ith trait over all inbred.

Genetic distances can be measured in molecular marker data where PCR amplification follows allele/locus model in following ways:

Where genotypes a and b's Euclidean distance, d (a, b), is the; Xai is the allele frequency for individual i; r is a constant determined by the coefficient used, and Xaj is the frequency of the allele a for individual j.

3. **Molecular variety:** When a locus/allele model may be utilised to evaluate genetic marker data or molecular marker data, allelic diversity is used. In these circumstances, data is utilised to create a binary matrix for additional research. The percentage of polymorphic loci (p), the average number of alleles per locus (n), the average anticipated heterozygosity (H), and the content of polymorphism information can all be used to define allelic diversity (PIC). In terms of all loci, including polymorphic and monomorphic loci, the percentage of polymorphic loci (p) provides an approximation of the number of polymorphic loci.

$$P = \frac{N_p}{N_t} \times 100$$

Where Nt is the total number of loci and Np is the number of polymorphic loci (polymorphic and monomorphic).

Calculated by dividing the total number of alleles by the number of loci, the mean number of alleles per locus (n) is represented as follows:

$$n = (1/k) \sum_{i=1}^{k} n_i$$

Where, k is the number of loci, and n¬¬i is the number of alleles at i^{th} locus Polymorphism information content (PIC) is an indirect estimate of number of alleles per locus. This can be expressed as below:

$$PIC = 1 - \sum_{i=1}^{n} (P_i)^2$$

Where, Pi is the frequency of ith allele at any particular locus.

VIII. GENETIC DIVERSITY ESTIMATION WITH STATISTICAL METHODS

When evaluating the genetic diversity among various strains, variations, or entries of a species, multivariate statistics are used. These methods can be used to evaluate genetic diversity since they have a very solid theoretical foundation and can provide the most accurate information about the actual genetic distances between genotypes. These methods can be used to evaluate genetic divergence, divide germplasm into groups, and choose diverse parents to produce transgressive segregants. Here are some examples of multivariate approaches in use:

1. **Analysis of metroglyphs:** Anderson created a semi-graphical method known as "Metroglyph analysis" for demonstrating genetic variability among several lines. With this approach, each genotype is represented by a circle (referred to as a glyph) with a defined radius and rays extending from its edge. Every variable has a specific place on the glyph. The index score of the variate is represented by the ray's length. This method makes use of a variety of trait variants such that the length of the rays on the glyph determines how much trait variation there is. The value of a genotype's index score is used to evaluate that genotype's performance. The length of the ray—which might be short, medium-length, or long—is determined by the score value.

2. **Statistics D2:** Mahalanobis created this method, which is also known as the generalised distance. By grouping genotypes into various clusters, this approach minimises the amount of comparisons between genotypes. The crucial condensation method is used to convert correlated data into uncorrelated variables in order to determine D2 values. The Mahalanobis distance, in its most basic form, is a measurement of the separation between two locations in a space defined by two or more correlated variables. The Mahalanobis distances between the points in a normal two-dimensional scatterplot, for instance, would be the same as the Euclidean distance if there are two uncorrelated variables. In a 3-D display, we could also just use a ruler to measure the separations between points if there are three uncorrelated variables. We are no longer able to plot the distances for more than three variables. The Mahalanobis distance will sufficiently account for the correlations in those situations, but the simple Euclidean distance is not a suitable measurement.

3. **Cluster studies:** The study makes the assumption that the data are discontinuous. It shows the pattern of genetic similarity between genotypes based on phenotypic characteristics or evolutionary links. It is used to distinguish between distinct groups and put related lines or genomes in one group. It is based on the following techniques: I single linkages (SLCA), (ii) complete linkage (CLCA), (iii) weighted paired group method using arithmetic mean (WPGMA), (iv) single linkages (WPGMA), (v) weighted paired group method using centroid (UPGMC), and (vi) median linkage (MLCA). Comparing breeding materials employed in accordance with pedigrees and calculated findings found

to be most consistent with recognised heterotic groupings, UPGMA and UPGMC provide more accurate grouping information than the other clusters.

4. **The principal component analysis (PCA):** One way to describe principal components analysis (PCA) is as a data-reduction method that works with quantitative types of data. Multi-correlated data are converted by PCA into another set of uncorrelated variables for additional research. Linear combinations of the original variables make up this new collection of variables. The generation of eigen-values and mutually independent eigen-vectors (principal components), sorted in descending order of variance size, is the basis of this method. These elements produce scatter plots of data that have the best characteristics for examining underlying variability and correlation. If X1, X2,..., Xn are the study's original data, then principal components can be described as:

$$z_1 = a_{11}x_1 + a_{12}x_2 + \ldots\ldots + a_{1n}x_n$$

With the condition such that a112 + a122+………..+ a1n2 = 1 similarly other principal components can be defined as:

With the condition, ap12 + ap22+………..+ apn2 = 1

This method is merely a means to an end for additional investigation. No statistical model or original variate distribution assumptions are needed for this procedure. It is important to note that this analysis is not necessary when the initial variables are uncorrelated. When distinct variables have the same unit, this is most appropriate. Standardizing all the variables allows for the avoidance of the challenges presented by various scales. Each variable is standardised by dividing it by the estimated standard deviation. It has been noted that PCA use in genetic diversity investigations has increased recently.

5. **Principal coordinate analysis (PCOA):** Schoenberg created another ordination technique that is somewhat akin to PCA. The Gower or Euclidean distances used to estimate the distances between all of the data points are matrices that the PCoA frequently discovers the eigen-values and eigen-vectors of. It creates a 2 or 3 dimensional scatter map of the samples with the least amount of distortion possible, with the distances between the samples in the plot reflecting their genetic distances. This has the drawbacks of being complex functions of the original variables and I not offering a direct connection between the components and the original variables.

6. **Canonical research:** It was Bartlett who initially proposed the concept of canonical analysis. By removing linear correlations between characters, it improves prediction by assuming additivity in all characters. The method to explain the relationships between two sets of variants was put out by Hotelling. Using the same variance-covariance matrix, Seal characterised it as "a process of differentiating as clearly as feasible between two or more multivariate normal universes." The benefit of this approach is that it is scale-neutral. Furthermore, when compared to PCA, group variable comparison is simpler.

7. **Factor evaluation:** Based on their inter-correlations or shared variance, this method breaks down large amounts of data into smaller, more manageable groups. It is predicated on the idea that correlated variables capture the same quality or trait. It is used to explain

the covariance relationships between numerous variables in terms of a small number of fundamental random variables known as factors. The basic objective of factor analysis is to use the fewest available factors and the fewest possible items or variables within each factor to explain as much variance in a data set as possible. The elements with Eigen values larger than 1.0 are taken into consideration for analytical interpretation.

8. **Communication analysis:** Similar to PCA in certain ways, correspondence analysis (CA) is an ordination technique that uses numbered or discrete data. It makes advantage of the Chi-square distance between the research objects. Correspondence analysis allows for comparison of relationships with counts of taxa or associations with counts of taxa. It has been discovered that several techniques for genetic diversity analysis produce comparable outcomes and may therefore be used interchangeably. Chandra analysed the results of the Metroglyph analysis and the Mahalanobis D2 distance and discovered a startling consistency in the grouping of flax genotypes. On the basis of this, he proposed using metroglyph analysis for preliminary grouping in several germplasms. Ariyo used factor, principal component, and canonical analysis to examine the degree of genetic variation in okra. She discovered that the results from factor and principle component analysis were comparable.

IX. SOFTWARE FOR ANALYSING GENETIC DIVERSITY

Software of many different kinds has been created to analyse genetic diversity. Multivariate statistics is the foundation of the majority of these programmes. The majority of PC-compatible software is freely accessible online. Various programmes are offered, according to Tanavar et al. Below is a brief description of some of the software:

1. **SAS:** SAS provides the package for several multivariate methodologies. It includes principal component analysis, cluster analysis, factor analysis, correspondence analysis, and canonical correlation, among other things. PROC PRINCOMP or PROC PRINQUAL can be used to do principal component analysis. You may do correspondence analysis, canonical correlation analysis, and factorial analysis, respectively, using PROC CORRESP, PROC CANCORR, and PROC FACTOR.

2. **SPAR 3.0:** The Statistical Package for Agricultural Research was created by IASRI, New Delhi (SPAR). In addition to other modules, it has the ability to do multivariate statistics.

3. **Past:** Hammer et al. created software for paleontological statistics. It is a no-cost, user-friendly, all-inclusive package. Parsimony analysis with cladogram plotting, detrended correspondence analysis, principal component analysis, principal coordinate's analysis, time-series analysis, geometrical analysis, etc. are some of the functions present in PAST.

4. **NTSYSPC (Numerical Taxonomy System for personal computer):** It is a well-known algorithm that has been applied in numerous scientific fields for analysing genetic diversity from molecular marker data. It uses a 0, 1 genotypic data matrix and is based on similarity indexes. It is employed in a number of applications, including principal coordinate analysis, principle component analysis, and cluster analysis.

5. **GenAlEx (Genetic Analysis in Excel):** It is easy-to-use software that is based on Excel. It was created for use in diversity genetics investigations using SSR, SNP, AFLP, allozyme, multilocus markers, and DNA sequencing data. It accepts three different types of data: codominant, dominant, and geographic. Frequency by locus, observed and expected heterozygosity, marker index, fixation index, allelic patterns, allele list, private alleles list, haploid diversity by population, haploid diversity by locus, haploid disequilibrium and pairwise Fst, Nei's genetic distance, principal component analysis, Shannon index, etc. are all included in the GenALEx analysis.

6. **Popgene:** It is another another approachable tool for studying genetic variation in and within natural populations. It makes it possible to carry out intricate analyses, generate statistical data that is supported by science, and analyse the population genetic structure using the desired markers or features. It accepts codominant, dominant, and quantitative qualities as three different types of data. Gene frequency, effective allele number, polymorphic loci, gene diversity, Shannon index, homozygosity test, F-statistics, gene flow, genetic distance (based on Nei coefficient), dendrogram (based on UPGMA and neighbour-joining method), neutrality, and neutrality are all included in the analysis

7. **Power marker:** It is a brand-new programme, with the initial authorised distribution occurring in January 2004. It was created especially for the analysis of population genetics using SSR/SNP data. Data setup is relatively simple and may be done by importing data from Excel or other formats. Options include population structure, phylogenetic analysis, association analysis, and tools. Summary statistics (allele number, gene diversity, inbreeding coefficient, estimation of allelic, genotypic, and haplotypic frequency; Hardy-Weinberg disequilibrium and linkage disequilibrium) are also available (Utility tools such as SNP simulation and identification, Mantel test and exact p-values for contingency tables)

X. CROP GENETIC DIVERSITY

Every crop development programme needs genetic diversity, and crop breeding is based on the creation and maintenance of genetic diversity. Many researchers have emphasised the use of Mahalanobis' D2 statistic in multivariate analyses to estimate genetic divergence (Anand and Murty, 1968; Kole and Chakraborty, 2012; Gadi et al. 2020). Within general limitations of fitness, the more different the parents, the higher the likelihood of obtaining larger levels of heterotic expression in F1 and a wide range of variability in segregating generations (Chakraborty and Bhattachraya, 2018; Sunny et al., 2022). (Anand and Murty, 1968, Singh et al, 2016a).

Genetic divergence was not associated to geographic differentiation, according to the genotype distribution pattern in the various clusters. Numerous genotypes that are geographically adjacent to one another fall into separate clusters, and vice versa. The propensity for such clustering to emerge notwithstanding borders of geography demonstrated that regional isolation was not the only factor causing variation in natural population (Rao et al., 1980). One explanation for the grouping of genotypes from various eco-geographical places into one cluster is the potential for free exchange of breeding materials. However, unidirectional selection used for a specific trait or a set of related traits in many locations may yield a comparable phenotype that can clump together into one cluster regardless of their

location (Singh and Gupta, 1968). Different genotypes of common geographic origin may form distinct clusters due to their parentage, developmental features, history of selection, and various out-crossing rates (Arnold et al., 1996).

Crossing between the intra-cluster genotypes is most likely to result in a significant amount of heterosis in the F1 generation and to produce a wide range of recombinants in the segregating generation, taking genetic divergence, per se performance of genotypes, as well as cluster mean, into consideration (Sunny et al. 2022).

PCA is a statistical method for locating and getting rid of duplicate genotypes with comparable traits (Singh et al., 2016b). It provides a precise indicator of genotypic differences and permits the natural classification of genotypes. One group may be allocated to each genotype in PCA, which is its main advantage (Singh et al., 2016b; Debnath et al. 2022). This test is also used to divide a large number of variables into significant components and determine how much each contributes to overall variation. Therefore, PCA was used to examine the relationships between the different attributes and to pinpoint the features that co-segregated.

Using gamma (γ) irradiation in particular, induced mutagenesis plays a significant role in producing genetic variability in crops such as cereals, pulses, and oilseeds for desirable traits of economic importance. These variants were then used in cross breeding programmes of several seed propagated crops (Chakraborty and Paul, 2012) to address challenges such as global food and nutritional security (Kant et al. 2020).

A key strategy for creating genetic diversity is hybridization. As a result, information on genotypic and phenotypic variation, heritability, and genetic advance are necessary for the possibility of direct selection as well as the prediction of the inheritable nature of the traits. This will increase the success of transferring the desirable traits through hybridization. The F1 rice varieties with high grain yield, semi-dwarf to medium stature, high tillering behaviour, resistance to lodging, and pleasant aroma were chosen by Sunny et al. (2022) after extensive research on fragrant rice.

XI. GENETIC DIVERSITY-RELATED THREATS

A sizable collection of valuable crop plant germplasm—roughly 6 million—is kept in gene banks around the world. Less than 1% of them have been used by breeders. This is a result of a biassed approach to plant breeding that prioritises a small number of crucial features that affect yield at the expense of other qualities. Numerous other germplasm accessions with a variety of characteristics are still untapped. This causes agricultural types to have a limited genetic foundation, which increases their genetic fragility and can be disastrous in the face of shifting climatic conditions. A sizable area of land has been converted to monoculture as a result of increased mechanisation in agriculture. This has taken the place of many local types and landraces from farmers' fields, which are the genetic sources of many beneficial features. In addition, the potential for producing natural variety in the form of wild forms and wild relatives of crop plants has been diminished due to the degradation of natural habitats in the name of urbanisation and modernization. A few lines have been employed exclusively in the breeding of new varieties and hybrids due to the commercialization of agriculture, virtually to the exclusion of other lines. As a result, many

cultivars now have plateaued yields and are more vulnerable to various biotic and abiotic stressors. The source of crucial genes, such as those for biotic and abiotic challenges, is genetic variety, which takes the shape of various landraces and germplasm

XII. CONCLUSION

To feed the growing population on dwindling arable land, plant breeding faces challenges. In this aspect, modern plant breeding has had some success. However, because of the limited genetic diversity of cultivated types in many crops, it has led to genetic fragility. Therefore, a paradigm shift in plant breeding that centres on a variety of genetic resources is required. Genetic variety is increasingly recognised as a distinct area that can contribute to the security of food and nutrition. Determining what to protect and where to conserve it will be easier with a better grasp of genetic variety. Crop plant genetic diversity is the cornerstone for the long-term creation of new types. Therefore, it is necessary to use various statistical approaches to characterise the distinct genetic resources and use them in the breeding programme. For accurate characterisation of germplasm resources, morphological and molecular data are used. High throughput molecular marker technologies have made it possible to characterise more germplasms in a shorter amount of time and with less resources. To enable better interpretation, the analysis is based on statistical tools. D2 statistics and PCA are the most popular statistical methods for morphological data because of how simple they are to interpret. For the analysis of molecular diversity, PCoA is widely used. The main software programmes used are POWERMARKER and GenAlEX because to their great informativeness. The diversity shown by various analyses can also be used in interogression of alien genes for particular phenotypes, heterosis breeding, and transgressive breeding.

REFERENCES

[1] Acharjee S., Chakraborty N. R., Das S. P. (2021). Marker based genetic variability analysis of rice (Oryza sativa L.) landraces for drought tolerance. African Journal of Biological Sciences 17 (1): 117-136.

[2] Anand, I.J. and Murty, B.R. 1968. Genetic divergence and hybrid performance in linseed. Indian J. Genet., 28: 178-185.

[3] Anderson E. A semigraphical method for the analysis of complex problems. Proc Natl Acad Sci U S A. 1957;43(10):923–927.

[4] Ariyo OJ. Genetic diversity in West African okra (Abelmoschus caillei) (A . Chev.) Stevels– Multivariate analysis of morphological and agronomic characteristics. Genetic Resources and Crop Evolution. 1993;40(1):25–32.

[5] Bartlett MS. Further aspects of the theory of multiple regression. Proc Camb Phil Soc. 1938;34(1):33–40.

[6] Beumont MA, Ibrahim KM, Boursot P, Bruqord MW. Measuring genetic distance. In A Karp editor. Molecular tools for screening biodiversity. London: Chapman and Hall; 1998. 325 p.

[7] Chakraborty N.R. and Bhattachraya S. (2018). Genetic diversity analysis in non-basmati Aromatic rice. International journal of Agricultural Science and Research 8(2): 191-198.

[8] Chakraborty, N. R. and Paul, A. (2012). Role of induced mutations in enhancing nutrition quality and production of food. International Journal of Bio-Resource and Stress management 4(1): 014-019.

[9] Chandra S. Comparison of Mahalanobis's method and Metroglyph technique in the study of genetic divergence in LinumusitatissimumL. germplasm collections. Euphytica. 1977;26(1):141–148.

[10] Cox TS, Murphy JP, Rodgers DM. Changes in genetic diversity in the red winter wheat regions of the United States. Proc Natl Acad Sci U S A. 1986;83(15):5583–5586.

[11] Debnath S, Sarkar A, Perveen K, Bukhari NA., Kesari K K, Verma A, Chakraborty NR, Tesema M. (2022) Principal component and path analysis for trait selection based on the assessment of diverse lentil populations developed by gamma-irradiated physical mutation. BioMed Research International, Volume 2022, Article ID 9679181, 14 pages. https://doi.org/10.1155/2022/9679181.

[12] Delannay X, Rodgers DM, Palmer RG. Relative genetic contributions among ancestral lines in North American soybean cultivars. Crop Sci. 1983;23:944–949.

[13] Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. 4th ed. Longman, Essex, UK; 1996. 153 p.

[14] Gadi Jarman, Chakraborty Nihar Ranjan and Imam Zafar (2020). Genetic diversity analysis in Indian mustard (Brassica juncea). Journal of Pharmacognosy and Phytochemistry 2020; 9(1): 952-955.

[15] Hammer K, Arrowsmith N, Gldis T. Agrobiodiversity with emphasis on plant genetic resources. Naturwissenschaften. 2003;90(6):241–250.

[16] Hotelling H. Simplified calculation of Principal Components. Psychometrical. 1936;1(1):27–35.

[17] Kant A; Chakraborty N. R.; Das BK (2020). Immediate radiation effects and determination of optimal dose of gamma rays on non-basmati aromatic rice (oryza sativa l.) Of eastern india. Journal of Experimental Biology and Agricultural Sciences, 8(5): 586-604.

[18] Kole, P.C. and Chakraborty, N. R. (2012). Assesment of genetic divergence in induced mutants of short grain aromatic non-basmati rice (Oryza sativa L.). Tropical Agriculture (Trinidad). 89(4):211-215.

[19] Mahalanobis PC. On the generalized distance in statistics. Proc Nat Inst Sci India B. 1936;2(1):49–55.

[20] Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants–salient statistical tools and considerations. Crop Science. 2003;43(4):1235–1248.

[21] Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA. 1973;70(12):3321–3323.

[22] Schoenberg IJ. Remarks to Maurice Frchet's article "Sur la dfinition axiomatique d'une elasse d'espaces distancibs vectoriellement applicable sur l'espace de Hilbert." Ann Math. 1935;38(3):724–732.

[23] Seal HL. Mutivariate Statistical Analysis for Biologists. London: Methuen and Co. Ltd; 1964;148(3676):1455.

[24] Singh Sarnam , Prakash Aruna, Chakraborty, N.R., Wheeler Candace, Agarwal P.K., Ghosh Arup (2016b). Trait selection by path and principal component analysis in Jatropha curcas for enhanced oil yield. Industrial Crops and Products. 86: 173–179.

[25] Singh Sarnam, Prakash Aruna, Chakraborty N. R., Wheeler Candac, Agarwal P. K, Ghosh A. (2016a). Genetic variability, character association and divergence studies in Jatropha curcas for improvement in oil yield. Trees structure and function. 30:1136-1180 DOI 10.1007/s00468-016-1354-0

[26] Smith JSC, Smith OS, Boven SL, et al. The description and assessment of distances between inbred lines of maize. III:A revised scheme for the testing of distinctiveness between inbred lines utilizing DNA RFLPs. Maydica. 1991;36:213–226.

[27] Sunny A, Chakraborty N R, Kumar A, Singh BK, Paul A, Maman S, Sebastian A and Darko D A. (2022). Understanding Gene Action, Combining Ability, and Heterosis to Identify Superior Aromatic Rice Hybrids Using Artificial Neural Network. Journal of Food Quality Volume 2022, Article ID 9282733, 16 pages https://doi.org/10.1155/2022/9282733.

[28] Swingland IR. Biodiversity. Definition of Encyclopedia of Biodiversity. 2001;1:377–390.

[29] Tanavar M, Kelestanie ARA, Hoseni SA. Software Programs for analyzing genetic diversity. Int J Farming and Allied Sci. 2014;3(5):462–466.