Genetic diversity in crop plants

Krishna Murari Prasad and Nihar Ranjan Chakraborty*
Department of Genetics and Plant Breeding
Institute of Agriculture
Visva Bharati, Sriniketan, West Bengal

*Corresponding author:
Nihar Ranjan Chakraborty (nrchakraborty@gmail.com)
ORCID: 0000-0002-5263-3366

Introduction

Crop plant diversification, whether it is anthropogenic (human) or innately, is essentially based on the population's genetic diversity. Diversity can be defined by the degree of heterogeneity present between or within species. Crop plant improvement systems are designed on the already-present intra- and inter-specific variations. It is possible that there would be limited room for improvement in the way the plants performed across a range of qualities if every member of the species had been the same. Divergence and Natural variation present among crops have been extensively identified and employed to promote crop species ever since systematic plant breeding initially began. But as the days passed on, natural variability reduced with diminishable rate as a result of (a) biased breeding techniques which were focused on merely traits like yield, (b) more often use of a small number of carefully chosen genotypes as parents in varietal development programmes, and (c) introduction of very few exceptional lines in many countries, which ultimately decreases the genetic diversity between modern crop cultivars pandemically. Diminishing genetic variation and variability among crop plant species has caused serious concern to the agricultural researchers. It will be difficult to further strengthen agricultural genotypes with less genetic diversity. Breaking yield limitations will become challenging, and plant breeders would not 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.

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.

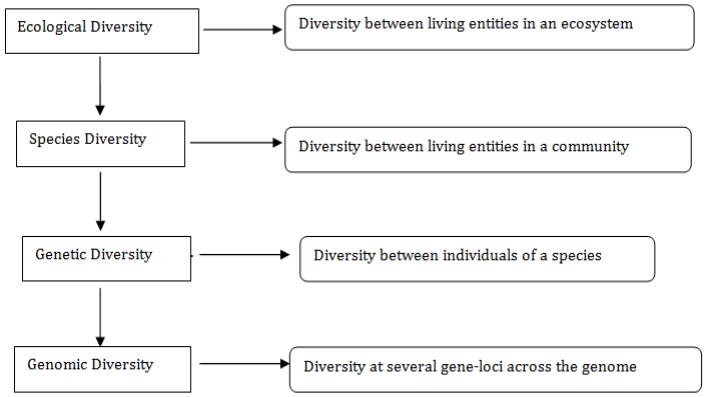

Figure 1 Hierarchy of diversity.

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 to many habitats, particularly those with shifting climatic circumstances

Genetic diversity affected by different kind of forces

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.

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.

(a)Morphology Indicators:

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.

(b) 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.

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.

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.

Measures of genetic diversity

(a) Genetic base

Genetic base of any crop expressed in terms Coefficient of Parentage (COP) or Coefficient of Correlation'. These indicate how frequently a line appears in the commercial varieties of a particular crop and is revealed by pedigree records of varieties released. COP is defined as the probability that alleles of two individuals are identical by descent. The segregating generations resulting from a cross between individuals with high COP will exhibit less variability and *vice versa*. The value of coefficient of parentage ranges from zero, where cultivars are completely unrelated, to one, where two cultivars have all alleles in common. The COP data matrix can be used to cluster genotypes and produce genealogically similar groups. Coefficient of parentage (COP) or coefficient of correlation (rxy) can be computed for all pairwise combinations of genotypes from pedigree information by formula given below (Falconer & Mackay, 1996):

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

Where, f_{xy} is a coefficient of co-ancestry, and F_x and F_y are inbreeding coefficients of X and Y, respectively. Delannay *et al*., Murphy *et al*. and Cox *et al*. developed different algorithms for calculation of coefficient of parentage. Other related measure is 'Relative Genetic Contribution (RGC)' computed by partitioning the genetic constitution of a selection into theoretical percentage attributable to different ancestors. The mean genetic contribution of a given ancestor is estimated by the mean of the relative genetic contributions of this ancestor to all varieties released. The successive summation of the mean relative genetic contributions generates cumulative relative genetic contributions over times. The assumptions underlying measure of relative genetic contribution are (i) unrelatedness of ancestors, and (ii) transmission of 50% of parental genes to the progeny with equal probability.

Studies on many crops revealed narrow genetic base in the released varieties of many crops in India. For example, the lines IR-8 and TN-1 (in rice), Spanish improved (in ground nut), Bragg (in soybean), T-1 and T-190 (in pigeonpea) and Pb-7 (in chickpea) appeared most frequently in commercial varieties of the respective crops released in India. Such frequent appearance of particular lines roughly gives the estimation of genetic base and consequently of genetic diversity.

(b) Genetic Distance

Genetic distance was first defined by Nei as the difference between two entities that can be described by allelic variation. This definition was later (1987) modified to "extent of gene differences among populations that are measured using numerical values". Beumont *et al.* provided a more comprehensive definition of genetic distance as any quantitative measure of genetic difference at either sequence or allele frequency level calculated between genotype individuals or populations. In simple terms, genotypes with many similar genes have smaller genetic distance between them. Euclidean or straight-line measure of distance is the most commonly used statistic for estimating genetic distance between individuals (genotypes or populations) by morphological data. Mohammadi & Prasanna have described in different measures of genetic distance in detail. Euclidean distance between two genotypes can be defined mathematically as below:

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

Where, d (a, b) is the Euclidean distance between genotype a and b; X_i is the observation on i^{th} phenotypic character, and Y_i is the observation on i^{th} 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 i^{th} trait for inbred lines a and b and *Var* $X_{(i)}$ is the variance for i^{th} trait over all inbred.

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

$$d\ (a,b) = constant\sqrt{(\sum \left([Xai - Xaj]^2\right)}$$

Where, d (a, b) is the Euclidean distance between genotype a and b; X_{ai} is the frequency of the allele *a* for individual *i*; X_{aj} is the frequency of the allele *a* for individual *j* and *r* is the constant based on coefficient used.

(c) Allelic diversity

Allelic diversity is used when genetic marker data or molecular marker data can be interpreted by locus/allele model. In such cases, data is used to generate binary matrix for further analysis. Allelic diversity can be described by (i) the percentage of

polymorphic loci (p), (ii) mean number of alleles per locus (n), (iii) total gene diversity or average expected heterozygosity (H), and (iv) polymorphism information content (PIC). Percentage of polymorphic loci (p) gives an estimate of number of polymorphic loci with respect to total loci including polymorphic and monomorphic loci and can be expressed as:

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

Where, N_p is the number of polymorphic loci and N_t is the number of total loci (polymorphic and monomorphic).

Mean number of alleles per locus (n) is calculated by dividing total number of alleles by the number of loci and can be expressed as:

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

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, P_i is the frequency of i^{th} allele at any particular locus.

Estimation of genetic diversity using statistical tools

Multivariate statistics are used to assess genetic diversity among different strains/varieties/entries of a species. These techniques have a very sound theoretical basis to provide most reliable information regarding the real genetic distances between genotypes and thus can be used for assessment of genetic diversity. These techniques can be used in assessment of genetic divergence, classification of germplasm into different groups and in selection of diverse parents to develop transgressive segregants. Some of the multivariate techniques being used are detailed below:

(i) Metroglyph analysis

Anderson developed a semi-graphical approach for displaying genetic diversity among a number of lines referred to as 'Metroglyph analysis'. This method represents each genotype by a circle of fixed radius (called glyph) with rays emanating from its periphery. Each variable is assigned a position on the glyph. The length of the ray represents index score of the variate. This method uses a range of variations arising from trait such that extent of trait variation is determined by the length of rays on the glyph. The performance of a genotype is adjudged by the value of the index score of that genotype. The score value determines the length of ray which may be small, medium or long.

(ii) D^2 Statistics

This technique also called Mahalanobis' generalized distance was developed by Mahalanobis. This technique reduces the number of comparisons among genotypes by classifying them into different clusters. D^2 values are estimated by transforming correlated variables into uncorrelated variables using pivotal condensation method. In general, the Mahalanobis distance is a measure of distance between two points in the space defined by two or more correlated variables. For example, if there are two variables that are uncorrelated, then we could plot points in a standard two-dimensional scatterplot; the Mahalanobis distances between the points would then be identical to the Euclidean distance. If there are three uncorrelated variables, we could also simply use a ruler (in a 3-D plot) to determine the distances between points. If there are more than 3 variables, we cannot represent the distances in a plot any more. In those cases, the simple Euclidean distance is not an appropriate measure, while the Mahalanobis distance will adequately account for the correlations.

(iii) Cluster analysis

This analysis assumes discontinuities within the data. It depicts the pattern of relatedness between genotypes based on evolutionary relationships or phenotypic performance. It is used to group similar lines/germplasm in one group and differentiate other groups. It is based on methods namely (i) Unweighted paired group method using arithmetic mean (UPGMA), (ii) Unweighted paired group method using centroid (UPGMC), (iii) Weighted paired group method using arithmetic mean (WPGMA), (iv) single linkages (SLCA), (v) complete linkage (CLCA) and (vi) Median linkage (MLCA). UPGMA and UPGMC provide more accurate grouping information on breeding materials used in accordance with pedigrees and calculated results found most consistent with known heterotic groups than the other clusters.

(iv) Principal component analysis (PCA)

Principal components analysis (PCA) can be defined as a data reduction technique applicable to quantitative type of data. PCA transforms multi-correlated variables into another set of uncorrelated variables for further study. These new set of variables are linear combinations of original variables. It is based on the development of eigen-values and mutually independent eigen-vectors (principal components) ranked in descending order of variance size. Such components give scatter plots of observations with optimal properties to study the underlying variability and correlation. Suppose X_1, X_2,……,X_n be the original data in a study, then principal components may be defined as:

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

With the condition such that $a_{112} + a_{122}$+……..+ $a_1n_2 = 1$
Similarly other principal components can be defined as:

$$z_p = a_{p1}x_1 + a_{p2}x_2 + \ldots\ldots + a_{pn}x_n$$

With the condition, $ap_{12} + ap_{22}$+……..+ $apn_2 = 1$

This technique is not an end rather a mean for further analysis. This technique does not require any statistical model or assumption about distribution of original variate. It is worth mentioning that when original variables are uncorrelated then there is no need to carry out this analysis. This is most suitable when different variables have same unit. The difficulty of different scales can be avoided by standardizing all the variables. Standardization is done by dividing each variable by its estimated standard deviation. Recently, a spurt has been reported in the use of PCA in genetic diversity studies.

(v) Principal coordinate analysis (PCoA)

It is another ordination method, somewhat similar to PCA, was developed by Schoenberg. The PCoA routinely finds the eigen-values and eigen-vectors of a matrix containing the distances between all data points, measured with the Gower distance or the Euclidean distance. It produces a 2 or 3 dimensional scatter plot of the samples such that the distances among the samples in the plot reflect the genetic distances among them with a minimum of distortion. This suffers from the disadvantages of (i) not providing a direct link between the components and the original variables and (ii) being complex functions of the original variables.

(vi) Canonical analysis

Bartlett was the first to give the idea of canonical analysis. It assumes additivity in all characters and improves prediction by eliminating linear correlations between characters. Hotelling proposed the technique to describe the dependencies between two sets of variants. Seal defined it as 'a procedure of discriminating as clearly as possible between two or more multivariate normal universes with the same variance-covariance matrix'. This method has the advantage of being neutral to scale. Further, comparison of group of variables is easier when compared to that in PCA.

(vii) Factor analysis

This technique reduces data into smaller meaningful groups based on their inter-correlations or shared variance. It is based on the assumption that correlated variables measure a similar factor or trait. It is used to describe the covariance relationships among many variables in terms of few underlying random quantities called factors. The main goal of factor analysis is to explain as much variance as possible in a data set by using the smallest number of factors and the smallest amount of items or variables within each factor. For interpretation of analysis, the factors with Eigen values greater than 1.0 are considered.

(viii) Correspondence Analysis

Correspondence analysis (CA) is an ordination method, somewhat similar to PCA, but for counted or discrete data. It uses Chi-square distance between the objects under study. Correspondence analysis can compare associations containing counts of taxa or counted taxa across associations. Different methods of genetic diversity analysis have been found to give similar results and hence can be used interchangeably. Chandra compared two methods (Mahalanobis D^2 distance and Metroglyph analysis) and found strikingly similarity in grouping pattern of flax genotypes. On this basis, he suggested that metroglyph analysis can be used for preliminary grouping in large number of germplasms. Ariyo compared the extent of genetic diversity in okra using factor, principal component and canonical analysis and found similar results between factor and principal component analysis.

Softwares for genetic diversity analysis

Many types of software have been developed for analysing genetic diversity. Most of these softwares are based on multivariate statistics. Most of the softwares are freely available on internet and suitable for PCs. Tanavar *et al*. have described different programs available. Some of the softwares are briefed below:

(i) SAS

SAS offers the package for different multivariate techniques. It involves canonical correlation, correspondence analysis, cluster analysis, factor analysis, principal component analysis etc. Principal component analysis can be performed using PROC PRINCOMP or PROC PRINQUAL. PROC CORRESP, PROC CANCORR and PROC FACTOR can be used for performing correspondence analysis, canonical correlation analysis and factorial analysis, respectively.

(ii) SPAR 3.0

IASRI, New Delhi have designed Statistical Package for Agricultural Research (SPAR). Apart from other modules, it is also capable of carrying out multivariate statistics.

(iii) Past

Paleontological Statistics software was developed by Hammer *et al*. It is a free, user friendly and comprehensive package. Functions found in PAST include parsimony analysis with cladogram plotting, detrended correspondence analysis, principal component analysis, principal coordinates analysis, time-series analysis, geometrical analysis etc.

(iv) NTSYSpc: (Numerical Taxonomy System for personal computer)

It is a popular program used to analyse genetic diversity from molecular marker data and has been used in different areas of science. It is based on similarity indices and works on 0, 1 matrix of genotypic data. It is used for several applications namely cluster analysis, principal component analysis, principal coordinate analysis, *etc*.

(v) GenAlEx: (Genetic Analysis in Excel)

It is an Excel-based and user-friendly program. It was designed for the use of SSR, SNP, AFLP, allozyme, multi locus markers and sequencing DNA data in diversity genetics analyses. It accepts three types data *viz.*, codominant data, dominant, and geographic data. GenALEx analysis include 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.

(vi) Popgene

It is another user-friendly package for the analysis of genetic diversity among and within natural populations. It enables to perform complex analysis and produce scientifically sound statistics and analyse population genetic structure using the target markers/traits. It accepts three types of data *viz*., codominant data, dominant and quantitative traits. The analysis includes gene frequency, allele number, effective allele number, polymorphic loci, gene diversity, Shannon index, homozygosity test, F-statstics, gene flow, genetic distance (based on Nei cofficent) and dendrogram (based on UPGMA and neighbour-joining method) and neutrality.

(vii) Power marker

It is a new program, with the first official version released in January 2004. It was designed specifically for the use of SSR/SNP data in population genetics analyses. Data can be imported from Excel or other formats, making data set-up very easy. Available options include summary statistics (allele number, gene diversity, inbreeding coefficient; estimation of allelic, genotypic and haplotypic frequency; Hardy-Weinberg disequilibrium and linkage disequilibrium), population structure, phylogenetic analysis, association analysis and tools (Utility tools such as SNP simulation and identification, Mantel test and exact p-values for contingency tables).

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 D^2 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 F_1(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 F_1 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 F_1 which possessed high grain yield with semi-dwarf to medium stature, high tillering behavior, resistance to lodging and pleasant aroma.

Threats to genetic diversity

Gene banks across the world maintain a large number of germplasm (about 6 million) of important crop plants. Of them, less than 1% has been utilized by breeders. This is because of lopsided approach of plant breeding aiming at only few important traits contributing towards yield at the cost of other traits. Many other germplasm accessions possessing diverse traits remain unutilized. This leads to narrow genetic base of crop varieties leading to genetic vulnerability which may be devastating in context of changing climatic conditions. Increased mechanization in agriculture has paved the way for monoculture over a large tract of land. This has replaced many landraces and local varieties from the farmers' field which are the genetic reservoirs of many useful traits. Apart, destruction of natural habitats in the name of urbanization and modernization has reduced the scope of generating natural variation in the form of wild forms and wild relatives of crop plants. With the commercialization in agriculture, few lines have been used exhaustively in breeding new varieties/hybrids almost to the exclusion of others. This has resulted in yield plateau and susceptibility of these varieties to different biotic and abiotic stresses. Genetic diversity in form of different landraces and germplasm serve as the source of important genes like for biotic and abiotic stresses.

Conclusion

Plant breeding confronts a difficulty in feeding an ever-increasing population on dwindling cultivable land. In this regard, modern plant breeding has had some success. However, because of the narrow genetic base of cultivated varieties in many crops, it has resulted in genetic susceptibility. As a result, there is a need for a paradigm shift in plant breeding that focuses on diverse genetic resources. Genetic diversity is now recognized as a separate area that can help with food and nutritional security. A better understanding of genetic diversity will assist in deciding what and where to conserve. Crop plant genetic diversity is the foundation for the long-term development of new varieties. As a result, there is a need to characterise the diverse genetic resources using various

statistical tools and incorporate them into the breeding programme. Morphological and molecular data are used together to precisely characterise germplasm resources. With the advancement of high throughput molecular marker technologies, it is now possible to characterise a greater number of germplasms in less time and with limited resources. For better interpretation, the analysis is based on statistical tools. Because of their ease of interpretation, D2 statistics and PCA are the most commonly used statistical tools for morphological data. PCoA is widely used in molecular diversity analysis. Because of their high informativeness, POWERMARKER and GenAlEX are the most commonly used software. The diversity revealed by various analyses can then be used in heterosis breeding and transgressive breeding.

References

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

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

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.

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

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.

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.

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.

Delannay X, Rodgers DM, Palmer RG. Relative genetic contributions among ancestral lines in North American soybean cultivars. *Crop Sci*. 1983;23:944–949.
Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*. 4th ed. Longman, Essex, UK; 1996. 153 p.

Hammer K, Arrowsmith N, Gldis T. Agrobiodiversity with emphasis on plant genetic resources. *Naturwissenschaften*. 2003;90(6):241–250.
Hotelling H. Simplified calculation of Principal Components. *Psychometrical*. 1936;1(1):27–35.

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

Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants–salient statistical tools and considerations. *Crop Science*. 2003;43(4):1235–1248.
Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*. 1973;70(12):3321–3323.

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.

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

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.

Swingland IR. Biodiversity. *Definition of Encyclopedia of Biodiversity*. 2001;1:377–390.

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

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

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.

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.

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.

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.

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.
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.

Chakraborty, NR. 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.

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.