Assessment of genetic diversity in crop plants - an overview

Introduction

Crop plant evolution either natural or human-directed, is primarily based on existing genetic diversity in the population. Diversity can be described as the degree of differentiation between or within species. Existing intra- and inter-specific differences are at the base of all crop improvement programmes. If all the individuals within the species would have been similar, then possibly there could not have been any scope for improvement in plant performances for different traits. Since the beginning of systematic plant breeding, natural variability and divergence between crops have been extensively identified and used in improvement of crop species. However, with the progress of time, natural variability got depleted due to (i) lopsided breeding practices focusing on improvement of only few traits (like yield and its component traits), (ii) frequent use of few selected genotypes as parents in varietal development programme and (iii) introduction of few outstanding lines to many countries thereby leading to increased genetic similarity between modern crop cultivars. Reduced genetic variability and diversity among crop plant species has raised serious concern among agricultural workers. With reduced genetic diversity, further improvement in crop varieties will be an arduous task. Breaking yield barriers will become difficult and plant breeders will be unable to meet the requirements arising out of ever-increasing demand on account of exploding population. Genetic diversity becomes more important in context of climatic change and associated unforeseen events as it may serve as the reservoir of many novel traits conferring tolerance to different biotic and abiotic stresses. Genetic diversity is the underlying cause of many important agriculturally important phenomena like heterosis and transgressive segregation. Diverse lines are needed for defect correction of commercial varieties and development of novel varieties. Hence, identification of diverse lines (if available), creation of diversity (if not available or limited) and its subsequent utilization are the major goals of any crop improvement programmes. In this context, knowledge on all aspects of genetic diversity viz., factors affecting genetic diversity, different methods of diversity analysis, their measurement and the softwares for carrying statistical analysis becomes imperative in order to utilize them prudently. Many reviews have been written focusing on vital issues like changes in genetic diversity under plant breeding,[1] genetic vulnerability of modern crop cultivars,[2] conservation and utilization of genetic resources[3,4] assessment of genetic diversity using molecular markers[5] and measurement of genetic diversity using statistical tools.[6–8] In the present review, an attempt is being made for comprehensive compilation of overall concepts in the area of genetic diversity, which could be of immense significance for extending knowledge and meaningful research.

Concept of diversity

Diversity is the essence of biological world. No two living things (even maternal twins) are exactly similar to each other. The difference in one or a few traits of the organism is referred to as variability. In common parlance, genetic variability and genetic diversity are considered synonym to each other which is erroneous. Genetic variability is the variation in alleles of genes or variation in DNA/RNA sequences in the gene pool of a species or population. This

HR Bhandari,[1,3] A Nishant Bhanu,[1] K Srivastava,[1] MN Singh,[1] Shreya,[1] A Hemantaranjan[2]

[1]Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, India
[2]Department of Plant Physiology, Institute of Agricultural Sciences, Banaras Hindu University, India
[3]Central Seed Research Station for Jute and Allied Fibres, India

Correspondence: A Nishant Bhanu, Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, India, Email nishant.bhanu@gmail.com

expresses itself in terms of alternate forms in phenotype. Genetic diversity, on the other hand, is a broad term encompassing all the variability occurring among different genotypes with respect to total genetic make-up of genotypes related to single species or between species. Genetic diversity can be measured by counting the number of different genes in a gene pool, but genetic variation can only be expected to occur and cannot be measured. Genetic variability thus, can be considered as the building blocks of genetic diversity. As recognized by Convention on Biological Diversity, there are three levels of diversity (Figure 1). At the highest hierarchy, lies the ecosystem diversity representing variability among different communities of species. In the next level of hierarchy, lies the species diversity representing different species within a community, also referred to as species richness. Genetic diversity is referred to the diversity present within different genotypes of same species. This is due to contrasting alleles of a gene in different individuals producing contrasting phenotypes. Swingland,[9–14] defined genetic diversity as the variation of heritable characteristics present in a population of the same species. The variation in heritable characters may express itself in the form of altered morphology, anatomy, physiological behaviour or biochemical features. The genomic diversity can be defined as diversity at several gene-loci within an individual. Genetic diversity has received the greatest attention among agricultural workers.
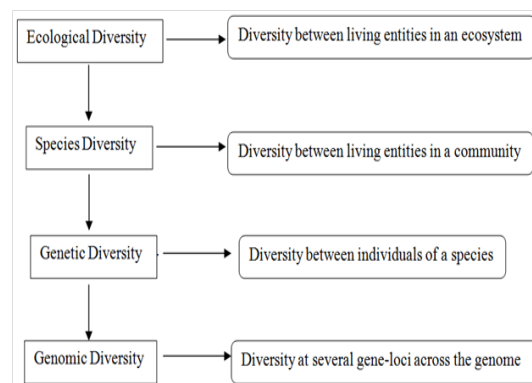

Figure 1 Hierarchy of diversity.

Importance of genetic diversity

Genetic diversity is the base for survival of plants in nature and for crop improvement. Diversity in plant genetic resources provides opportunity for plant breeders to develop new and improved cultivars with desirable characteristics, which include both farmer-preferred traits (high yield potential, large seed, etc.) and breeder-preferred traits (pest and disease resistance and photosensitivity, etc.). From the very beginning of agriculture, natural genetic variability has been exploited within crop species to meet subsistence food requirement. Later the focus shifted to grow surplus food for growing populations. Presently the focus is on both yield and quality aspects of major food crops to provide balanced diet to human beings. With changing climatic scenario, breeding of climate resilient varieties is becoming more important. The existence of genetic diversity represented in the form of wild species, related species, breeding stocks, mutant lines etc. may serve as the source of desirable alleles and may assist plant breeders in breeding climate resilient varieties. The breeding of climate resilient varieties requires novel traits like tolerance towards potential new insect-pests and diseases, extreme heat, extreme cold, and towards various air- and soil- pollutants. For ever-changing breeding goals, different genes need to be reserved in cultivated and cultivable crops species in the form of germplasm resources. Presence of genetic diversity within and between crop plant species permits the breeders to select superior genotypes either to be directly used as new variety or to be used as parent in hybridization programme. Genetic diversity between two parents is essential to realize heterosis and to obtain transgressive segregants. Genetic diversity facilitates breeders to develop varieties for specific traits like quality improvement and tolerance to biotic and abiotic stresses. It also facilitates development of new lines for non-conventional uses like varieties for biofuel in sorghum, maize etc. Diversity is also important with respect to adaptability of crop plants to varied environments with special reference to changing climatic conditions. Some of the germplasm lines harbouring desirable gene in different crops are listed in Table 1.

Table 1 Sources for different traits in different crop plants

S. No	Crop	Variety/germplasm	Trait	Importance	References
1.	Okra	Pusa Sawani	YVMV Tolerance	Biotic stress tolerance	-
2.	Maize/ Sorghum	Brown mid rib lines	Low lignin and fibre content	Forage digestibility/ palatability	-
3.	Chickpea	IPC 2004-52/ PDG 84-10	Fusarium wilt tolerance	Biotic stress Tolerance	IIPR, Kanpur 2009[10]
4.	Cotton	Abadhita, LK 861, Kanchana, Supriya	White fly tolerance	Biotic stress Tolerance	ICAR, 2007[11]
5.	Tomato	F7028	High lycopene content	Nutrition	IIVR, Varanasi[12]
6.	Wheat	Nap Hal/UP 2672	High protein	Nutrition	DWR, Karnal, 2014[13]
7.	Rice	Govindbhog/ Chakhao	Aroma	Palatability	CRRI, Cuttack, 2014[14]

Forces affecting genetic diversity

Genetic diversity is primarily a function of sexual recombination. During meiosis, homologous chromosomes undergo crossing over which results in appearance of several new recombinations. Different factors affect the genetic diversity in plants. Evolutionary forces like selection, mutation, migration and genetic drift act continuously and results in continuous changes in allelic frequency in a population and affects the genetic diversity. Domestication or artificial selection favours few alleles at the cost of others resulting in increased frequency of selected alleles. Consequently, domestication reduces the genetic diversity when compared to the diversity in wild. Natural selection also affects the genetic diversity considerably. Directional and stabilizing selection decreases while disruptive selection increases the genetic diversity. Mutation is also reported to increase genetic diversity.[15] Qualitative mutation expresses itself in the form of abrupt changes in morphological/anatomical/biochemical features. Quantitative or micro-mutations have smaller and gradual effects which accumulate over time and bring about changes. Mutation may also bring about several chromosomal aberrations. Smaller sub-lethal or non-lethal aberrations bring about genetic diversity in the form of altered phenotype. Mating system of crop plants also affect genetic diversity. Inbreeding reduces while out breeding increases genetic diversity. Genetic drift can lead to loss of rare alleles thereby reduces genetic diversity. The physical distribution of individuals of a species also affects genetic diversity. Larger the physical distribution of individuals, lesser is the chances of having same genetic make-up. Some techniques like wide-hybridization, hybridization between incompatible types or introgression from previously isolated populations increase the genetic diversity as they result in generation of new phenotypes. In contrast, intra-specific hybridization reduces the genetic diversity.[16] Gene flow within population increases the genetic diversity as new alleles are introduced.

Methods of diversity analysis

Diversity analysis can be carried out using morphological, cytological, biochemical and molecular characterization. Initially, morphological markers were used for diversity analysis and are still in use. These were naturally occurring variants of a particular plant species. Later, cytological and biochemical differences occurring in the genotypes of a species started to be used in genetic diversity assessment. With the advent of genomic tools, molecular markers became the method of choice for genetic diversity assessment.

Morphological markers

These analyses are carried out by raising germplasm lines, purelines, improved varieties etc. in a particular experimental design.

This involves morphological characterization of different entries grown in the field as the morphological characteristics are the strongest determinants of the agronomic value and taxonomic classification of plants.[17] Morphological evaluations are direct, inexpensive, easy and do not require expensive technology. However, the requirements of large tracts of land and human labourers over a period of time make it expensive. They suffer from the constraints of environmental-sensitivity and subjective characterization when compared to other methods. They are mostly dominant/recessive, have some biological effect and some morphological variants cannot survive. Different sets of morphological traits are taken into consideration for different group of crop plants (Table 2).

Table 2 Morphological traits used as markers for different group of plants

Group/ family	Characters
Legumes	Growth habit, Flower colour, Leaf shape, Pod and seed shape, Root Nodule traits
Cereal crops	Stem pigmentation, Panicle length, Grain colour, Grain shape
Fibre crops	Plant height, Basal diameter, Fibre content, Lignin and hemicellulose content
Vegetable crops	Growth habit, Hypocotyl colour, Stem and leaf pubescence
Forage crop	Plant height, Lignin and hemicellulose content
Mustard	Siliqua length, Siliqua beak length, siliqua angle with main raceme, Oil content
Medicinal crops	Herbage yield, Essential oil yield

Cytological markers

It involves study of cytological features like chromosome size, secondary constriction in chromosomes, position of centromere, arm ratio, constitutive heterochromatic patterns, banding characteristics (G, Q, R and N banding), DNA content, total genomic chromosome length, chromosome volume etc. Different cytological features have been applied to assess genetic diversity within and between species in maize,[18] in potato,[19] in lentil, in radish[21] etc. However, these have limited applications in genetic diversity analysis on account of their limited number and low resolution.

Biochemical markers

It involves separation of proteins or their variants (isozymes) into specific banding patterns. The isozymes reflect products of different alleles and not the genes. These isozymes can be mapped onto chromosomes and can be used genetic markers for mapping other genes.[22] This is a rapid method of assessing diversity and requires smaller amount of plant tissue as sample. However, they are limited in number, affected by environmental fluctuations and cannot be used to construct a complete genetic map.

Molecular markers

It involves study of variation among genotypes at DNA/RNA level. Different molecular markers have different characteristics making them suitable for different purposes. They are primarily classified as hybridization-based and PCR-based. Recently, new generation of markers based on sequence or array-platforms have been developed. They can also be classified as neutral markers, genes markers and functional markers based on their activity and expression. Further, these markers may be based on variation in genomic DNA/RNA, ribosomal RNA or organelle genome sequences. Chloroplast microsatellites have been developed[5] and used in assessment of genetic diversity at intra-specific level in wheat,[23] barley,[24] apple,[25] rice,[26] pearlmillet[27] etc. Mitochondrial DNA in plants, in contrast, has been demonstrated to be an unsuitable tool for studying genetic diversity, being quantitatively scarce. Molecular markers are the method of choice for genetic diversity assessment on account of their hyper variability, better genomic coverage, high reproducibility, amenability to automation, being neutral and free from environmental fluctuations. Many studies on genetic diversity have been reported to use both morphological- and molecular- markers simultaneously.

Measures of genetic diversity

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.[28] The COP data matrix can be used to cluster genotypes and produce genealogically similar groups.[29] 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[30] (Falconer & Mackay, 1996):

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

Where, *fxy* is a coefficient of co-ancestry, and *Fx* and *Fy* are inbreeding coefficients of X and Y, respectively. Delannay et al.[31] Murphy et al.[32] and Cox et al.,[33] 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.[34] 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.[35] 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.

Geneticd

Genetic distance was first defined by Nei[36] 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.,*[37] 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 et al.[8] have described in different measures of genetic distance in detail. Euclidean distance between two genotypes can be defined mathematically as below:

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

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.,[38] developed another measure of genetic diversity in inbred lines which can be expressed as below:

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

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:

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

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

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:

$$P = \frac{Np}{Nt} \times 100$$

Where, Np is the number of polymorphic loci and Nt 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:

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

Where, *k* is the number of loci, and ni is the number of alleles at ith locus Polymorphism information content (PIC) is an indirect estimate of number of alleles per locus. This can be expressed as below:

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

Where, Pi is the frequency of ith 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.[39] 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:

Metroglyph analysis

Anderson[40] 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.

D2 Statistics

This technique also called Mahalanobis' generalized distance was developed by Mahalanobis.[41] This technique reduces the number of comparisons among genotypes by classifying them into different clusters. D2 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.

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.[42]

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 $x_1, x_2, \ldots, x_n$ be the original data in a study, then principal components may be defined as:

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

With the condition such that a112 + a122 + ……….. + a1n2 = 1

Similarly other principal components can be defined as:

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

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

Principal coordinate analysis (PCoA)

It is another ordination method, somewhat similar to PCA, was developed by Schoenberg.[43] 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.

Canonical analysis

Bartlett[44,45] 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[46,47] proposed the technique to describe the dependencies between two sets of variants. Seal[48] 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.

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 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.

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[49] compared two methods (Mahalanobis D2 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[50] 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 analyzing 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.[51] have described different programs available. Some of the softwares are briefed below:

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.

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.

Past

Paleontological Statistics software was developed by Hammer et al.[52] 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.

NTSYSpc: (Numerical Taxonomy System for personal computer)

It is a popular program used to analyze genetic diversity from molecular marker data and has been used in different areas of science. It is based on similiarity 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.*[53]

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 diversiry 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.

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 analyze population genetic structure using the target markers/traits. It accepts three types of data *viz.*, codominant data, dominant and quantitative traits. The analysis include 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 neighbor-joining method) and neutrality.

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).

Threats to genetic giversity

Gene banks across the world maintain a large number of germplasm (about 6million) of important crop plants.[54] Of them, less than 1% has been utilized by breeders.[55] 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 is facing challenge to feed the ever increasing population with diminishing cultivable land. Modern plant breeding has achieved some success in this regard. However, it has resulted in the genetic vulnerability because of narrow genetic base of cultivated varieties in many crops. Hence, there is a need of paradigm shift in plant breeding focussing on diverse genetic resources. Genetic diversity has now been acknowledged as a specific area that can contribute in food and nutritional security. Better understanding of genetic diversity will help in determining what to conserve as well as where to conserve. Genetic diversity of crop plants is the foundation for the sustainable development of new varieties. So there is a need to characterize the diverse genetic resources using different statistical tools and utilize them in the breeding programme. Morphological data in conjunction with molecular data are used for precise characterisation of germplasm resources. With the advent of high throughput molecular marker technologies it is possible to characterize larger number of germplasm with limited time and resources. The analysis is based on statistical tools for better interpretation. The most used statistical tools for morphological data are D2 statistics and PCA because of their easy interpretation. PCoA is very much in use for molecular diversity analysis. POWERMARKER and GenAlEX are mostly used software because of their high informativeness. The diversity indicated by different analysis can further be utilized in heterosis breeding, transgressive breeding and interogression of alien genes for specific traits.

Acknowledgements

Conflict of interest

The author declares no conflict of interest.

References

1. Fu Yong Bi. Understanding crop genetic diversity under modern Plant Breeding. *Theoretical and Applied Genetics*. 2015;128(11):21312142.

2. Keneni G, Bekele E, Imtiaz M, et al. Genetic vulnerability of modern crop cultivars: causes, mechanism and remedies. *Int J Plant Res*. 2012;2(3):69–79.

3. Ogwu MC, Osawaru, ME, Ahana CM. Challenges in conserving and utilizing plant genetic resources (PGR). *Int J Genet and Mol Bio*. 2014;6(2):16–22.

4. Rao VR, Hodgkin T. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tissue and Organ Cult*. 2002;68(1):1–19.

5. Mondini L, Noorani A, Pagnotta MA. Assessing plant genetic diversity by molecular tools. *Diversity*. 2009;1(1):19–35.

6. Aremu CO. Genetic Diversity:A review for need and measurements for intraspecie crop improvement. *J Microbiol Biotech Res*. 2011;1(2):80–85.

7. Balzarini M, Teich I, Bruno C, et al. Making genetic biodiversity measurable:A review of statistical multivariate methods to study variability at gene level. *Revista de la Facultad de Ciencias Agrarias*. 2011;43(1):261–275.

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

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

10. Indian Institute of Pulse Resaecrh 25 Years of Pulses Research at IIPR, India; 2009.

11. *Indian Council of Agricultural Research*. Research Achievements of AICRPs on Crop Science. India: ICAR; 2007.

12. Germplasm collection. Indian council of agricultural research, India; 2016.

13. Directorate of Wheat Research. Annual Report. Karnal, India; 2014.

14. Central Rice Research Institute. Annual Report. Cuttack, India; 2014.

15. Yilmaz A, Boydak E. The effects of cobalt–60 applications on yield components of cotton *(Gossypium barbadense L.)*. *Pak J Bio Sci*. 2006;9(15):2761–2769.

16. Osawaru ME, Ogwu MC, Aiwansoba RO. Hierarchical approaches to the analysis of genetic diversity in plants: a systematic overview. *University of Mauritius Res J*. 2015;21:1–33.

17. Cholastova T, Knotova D. Using morphological and microsatellite (SSR) markers to assess the genetic diversity in alfalfa *(Medicago sativa L.)*. *Int J of Biol*. 2012;6(9):781–787.

18. Albert PS, Gao Z, Danilova TV, et al. Diversity of chromosomal karyotypes in maize and its relatives. *Cytogenet Genome Res*. 2010;129(1–3):6–16.

19. Das AB, Mohanty IC, Mahapatra D, et al. Genetic variation of Indian potato *(Solanum tuberosum L.)* genotypes using chromosomal and RAPD markers. *Crop Breeding and Applied Biotech*. 2010;10(3):238–246.

20. Pal T, Ghosh S, Mondal A, et al. Evaluation of genetic diversity in some promising varieties of lentil using karyological characters and protein profiling. *J Gen Engineering and Biotech– In Press*. 2016;14(1):39–48.

21. Chen F, Liu H, Yao Q, et al. Genetic variations and evolutionary relationships among radishes *(Raphanus sativus L.)* with different flesh colors based on red pigment content, karyotype and simple sequence repeat analysis. *African J of Biotech*. 2015;16(50):3270–3281.

22. Xu Y. *Molecular Plant Breeding*. South Asia: CABI; 2009.

23. Mori N, Kondo Y, Ishii T, et al. Genetic diversity and origin of timopheevi wheat inferred by chloroplast DNA fingerprinting. *Breed Sci*. 2009;59:571–578.

24. Neale DB, Saghai–Maroof MA, Allard RW, et al. Chloroplast DNA diversity in populations of wild and cultivated barley. *Genetics*. 1988;120(4):1105–1110.

25. Coart E, VAN Glabeke S, DE Loose M, et al. Chloroplast diversity in the genus Malus:new insights into the relationship between the European wild apple *(Malus sylvestris (L.) Mill.)* and the domesticated apple *(Malus domestica Borkh.)*. *Mol Eco*. 2006;15(8):2171–2182.

26. Li Wen–Jia, Kang Gong–Ping, Zhang B, et al. Chloroplast DNA genetic diversity between Asian cultivated rice *(Oryza Sativa L.)* and different types of cytoplasmic male sterile rice. *African J Agril Res*. 2012;7(25):3705–3711.

27. Clegg MT, Rawson JRY, Thomas K. Chloroplast DNA variation in Pearl Millet and related species. *Genetics*. 1984;106(3):449–461.

28. Martin JM, Blake TK, Hockett EA. Diversity among North American Spring Barley cultivars based on coefficient of parentage. *Crop Sci*. 1991;31(5):1131–1137.

29. Bered F, Barbosa–Neto JF, De Carvalho FIF. Genetic variability in common wheat germplasm based on coefficients of parentage. *Gen and Mol Biol*. 2002;25(2):211–215.

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

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

32. Murphy JP, Cox TS, Rodgers DM. Cluster analysis of red winter wheat cultivars based upon coefficients of parentage. *Crop Sci*. 1986;26(4):672–676.

33. 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.

34. Gopal J and Oyama K. Genetic base of Indian potato selections as revealed by pedigree analysis. *Euphytica*. 2005;142(1–2):23–31.

35. Maw Sun Lin. Genetic base of japonica rice varieties released in Taiwan. *Euphytica*. 1991;56(1):43–46.

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

37. 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.

38. 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.

39. Singh S and Pawar IS. Theory and Application of Biometrical Genetics. India: CBS Publishers; 2005.

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

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

42. Aremu CO, Adebayo MA, Ariyo OJ, et al. Classification of genetic diversity and choice of parents for hydrization in cowpea vigna unguiculata (L) walip for humid savanna ecology. *African J of Biotech*. 2007;6(20):2333–2339.

43. 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.

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

45. Bartlett MS. Multivariate analysis. *J Roy Statistics Soc B*. 1947;9:170–197.

46. Hotelling H. The Most predictable criterion. *J Educational Psychology*. 1935;26:139–142.

47. Hotelling H. Simplified calculation of Principal Components. *Psychometrical*. 1936;1(1):27–35.

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

49. Chandra S. Comparison of Mahalanobis's method and Metroglyph technique in the study of genetic divergence in *Linum usitatissimum L*. germplasm collections. *Euphytica*. 1977;26(1):141–148.

50. 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.

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

52. Hammer Ø, Harper DAT, Paul DR. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica.* 2001;4(1):1–9.

53. Rohlf FJ (1998) NTSYS pc:Numerical Taxonomy System, Version 2.1. Exeter Publishing, Setauket, USA.

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

55. Upahyaya HD, Furman BJ, Dwivedi SL, et al. Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea. *Plant Genet Resour.* 2006;4(1):13–19.