Unlocking Genetic Diversity and Germplasm Characterization with Molecular Markers: Strategies for Crop Improvement

Aniket Bunjkar [a++*], Puneet Walia [a#] and Sanjeet Singh Sandal [a#]

[a] Department of Genetics and Plant Breeding, Lovely Professional University, Phagwara, Punjab, 144411, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Review Article

ABSTRACT

Molecular markers have emerged as the most useful tools in assessment of genetic diversity and characterizing germplasm, for crop improvement. This review paper comprehensively analyzes the applications of molecular markers in diversity analysis and germplasm characterization. It underscores the significance of genetic diversity as the bedrock for plant breeding programs, enabling the development of improved varieties with desirable attributes such as higher yields, stress tolerance, and enhanced nutritional profiles. The paper provides an overview of various types of molecular markers, including hybridization-based markers (e.g., RFLPs) and PCR-based markers (e.g., RAPDs, AFLPs, SSRs, and SNPs). It discusses marker selection strategies, emphasizing the consideration of factors like polymorphism, informativeness, and the potential for multiplexing and

_____

++ *Student;*
# *Assistant Professor;*
Corresponding author: E-mail: aniketbunjkar44@gmail.com;

high-throughput genotyping. Diversity analysis techniques, including principal component analysis (PCA), cluster analysis methods (UPGMA and Neighbor-Joining), and population structure analysis (model-based approaches like STRUCTURE), are detailed. These methods enable the assessment of genetic relationships, identification of subpopulations, and selection of diverse parents based on molecular marker data. The review further explores the applications of molecular markers in germplasm management, conservation, and utilization. It discusses the role of markers in targeted introgression of desirable traits from diverse sources, as well as the integration of genotypic and phenotypic data for association mapping, genomic prediction, and selection. Additionally, the paper addresses emerging technologies, such as next-generation sequencing (NGS) and the integration of molecular markers with other omics data. It also highlights the practical applications and impact of molecular markers in crop improvement programs, including marker-assisted selection (MAS), genomic selection, and the development of genetically modified crops. Finally, the review outlines challenges and future perspectives, including limitations0 of current technologies, the potential of emerging techniques like NGS, and the integration of molecular markers with other omics approaches for a comprehensive understanding of complex traits.

1. INTRODUCTION

Genetic variations in or among the populations, can be effectively used for utilization of plant genetic resources [1]. Historically, plant breeders have relied on evaluating agronomic and morphological traits to assess genetic diversity and other crucial characteristics [2]. However, contemporary breeders increasingly favor diversity analysis using molecular markers due to their proficiency in uncovering variations among genotypes, surpassing the limitations of morphological markers [3].

1.1 Genetic Diversity and Its Importance in Crop Improvement

Term Genetic diversity is differently defined by various authors, "The degree of genetic variation seen in individuals within a species because of processes including recombination, mutations, gene flow, and genetic drift is known as genetic diversity. It symbolizes the diversity of alleles and gene combinations among a group of organisms" [4] or "The range and sum of genetic variation within a population or between populations", in which the term diversity represents the differences among the individuals [5] or "Genetic diversity is the variability of heritable traits present in a population of the given species" [6] or "Genetic diversity is the variety of alleles present in a particular population or among populations and it is reflected in morphological, physiological and behavioral differences between individuals and populations" [7] or "Any measure that can calculate the magnitude of genetic variability present in a population" [8]. Genetic diversity forms the bedrock of plant breeding and

crop improvement initiatives. Comprehending and characterizing the extant genetic variation within crop germplasm is imperative to pinpoint desirable traits for developing enhanced varieties boasting higher yields, resilience against biotic and abiotic stresses, and superior nutritional profiles [9,10]. A vast genetic diversity endows plants with the capacity to adapt to abrupt environmental fluctuations [11].

Plant genetic diversity is invaluable for breeders as it provides a pool of desirable traits to select from when developing new crop varieties or parental lines for hybridization [10]. Utilizing genetically divergent parents in breeding programs allows for the improvement of productivity, disease resistance, and other favorable characteristics in agricultural and horticultural crops [12]. Maintaining high levels of genetic diversity is crucial as it equips breeders with an array of genetic variations to draw upon. By combining diverse genetic backgrounds, breeders can assemble unique gene combinations, leading to the development of superior and adaptable varieties that meet changing environmental conditions and market demands [10,12]. The success of various crop improvement programs depends upon the efficient identification and incorporation of plant genetic resources, such as currently grown cultivars, newly developed varieties, landraces, wild relatives, and germplasm collections [9]. Detailed understanding of genetic diversity and its distribution is important for its effective conservation and utilization, as it guides breeders for the decisions on what to conserve and where to focus efforts [13]. From a plant breeder's perspective, genetic diversity is a must

for developing resilient and adaptable crops that can withstand various abiotic stresses and biotic stresses, it is also important for yield and yield related traits [14].

1.2 Role of Germplasm Characterization in Plant Breeding

Characterization involves describing plant germplasm by evaluating highly heritable traits like morphology, physiology, agronomic features, seed proteins, oils, and molecular markers. This information aids germplasm utilization. Though time-consuming and costly, characterization can occur at any conservation stage [15].

Plant breeders use various methods to characterize germplasm and evaluation of genetic diversity, for identification of diverse parents for breeding programs [16]. These methods include evaluation of phenotypic or morphological traits, molecular approaches and biochemical or allozyme analysis [17]. Germplasm is considered a valuable source for identification of genes related to desirable traits. Once a trait is linked to biosynthetic pathways then alleles responsible for expression of that trait are identified, researchers understand crops on broader base. This knowledge is helpful for making specific parental and allelic combinations for the discovery of superior varieties [18]. By leveraging these characterization methods, breeders can unlock the potential of genetic diversity and develop improved crop varieties with enhanced performance and adaptability.

Germplasm collections comprise trait-specific accessions exhibiting tolerance to stresses like heat, drought, salinity, cold, as well as desirable traits like high nitrogen and water use efficiency [19]. These accessions offer valuable genetic resources for crop improvement programs.

1.3 Overview of Molecular Marker Technologies

Over the last three decades, major advancements in understanding plant genomes and gene functions have revolutionized plant breeding. Breeders can now engineer crops with desired traits like higher yields and resilience, driving more effective crop improvement [2]. Molecular markers are DNA sequences that highlight genetic variations like insertions, deletions, and mutations across individuals. Their function is to identify and differentiate specific genomic regions, making them invaluable tools for genetic mapping and analysis. [20]. The ability to detect genetic variations with molecular markers has enabled breeders to pinpoint desirable traits and excellent plant genetics. This advancement has driven the development of improved crop varieties with better characteristics.

2. MOLECULAR MARKER-BASED DIVERSITY ANALYSIS

Molecular markers are invaluable tools in the study of genetic diversity. They are used in the analysis of phylogenetic relationships, facilitate the selection of elite varieties, and allow for the comparison of genetic similarities and differences between species. [21]. Various DNA markers have been used as important molecular tools in plants for genetic relation studies among individuals, hybrid validation, varietal identification, phylogenetic relationship between species, gene mapping and quantitative trait loci (QTL) detection in last few decades [22].

Genetic variability within a population can be assessed through: The number (and percentage) of polymorphic genes in the population, number of alleles for each polymorphic gene and the proportion of heterozygous loci per individual [20].

2.1 Types of Molecular Markers

Molecular markers are of two types,

2.1.1 Hybridization-based markers

In molecular genetics, hybridization-based markers are used to visualize DNA profiles. This involves digestion of DNA with restriction enzymes and then hybridizing the resulting fragments to radioactive labeled probe, probe can be a DNA segment of known sequence. Restriction Fragment Length Polymorphism (RFLP) was the first hybridization-based marker [23]. The pioneering work in utilizing RFLP markers for human linkage mapping was conducted by Botstein and colleagues [24].

RFLP markers are regarded as powerful tools for comparative and synteny mapping studies. These markers exhibit a co-dominant inheritance pattern and are highly specific to loci. One of the key advantages of RFLP genotyping is its high reproducibility, coupled with a relatively straightforward methodology that does not necessitate specialized equipment [25]. The simplicity and reliability of RFLP markers have

made them invaluable in genetic research and analysis.

2.1.2 PCR-based markers

Polymerase Chain Reaction (PCR) based markers work on the principle of variations in DNA sequences. Also known as the second generation of molecular markers [26]. Random Amplified Polymorphic DNA, Amplified Fragment Length polymorphism, Simple Sequence Repeats and Single Nucleotide Polymorphism are some major PCR based markers used in the last few decades.

2.1.2.1 Random amplified polymorphic DNA (RAPDs)

The Random Amplified Polymorphic DNA (RAPD) is a PCR-based marker system that was developed independently by Williams et al. (1990) and Welsh and McClelland (1990) [27,28]. The RAPD technique utilizes short, random oligonucleotide primers (8-15 nucleotides) that bind and amplify multiple complementary sequences scattered across DNA, generating a unique banding pattern [25,29]. To visualize the results, an agarose gel electrophoresis system is used. The specific banding pattern on the gel indicates polymorphisms that present either at or between the primer binding sites [25,30]. This technique allows identification of genetic variations without prior information of the target DNA sequence, making it a valuable tool for genetic diversity studies and molecular marker analysis.

2.1.2.2 Amplified fragment length polymorphism (AFLPs)

The limitations of RAPD and RFLP markers were solved with the development of Amplified Fragment Length Polymorphism (AFLP) markers. AFLP is a combination of RFLP and PCR technology, in this method DNA is first digested and then PCR amplification is done [31].

In AFLP technique the digestion of DNA was done by using two restriction enzymes – a frequent cutter and a rare cutter, then the short oligonucleotide sequences are ligated to the resulting fragments, with one oligonucleotide being specific to the frequent cutter site and the other to the rare cutter site [32]. Subsequently, PCR amplification is performed, selectively amplifying only those fragments that have both oligonucleotides attached.

Then by using gel electrophoresis or autoradiography, the banding pattern observed represents the different DNA fragments amplified and can be utilized for analyzing genetic variations [33]. This technique combines the strengths of RFLP and PCR, providing a resilient method for detecting polymorphisms and facilitating diversity studies.

2.1.2.3 Simple sequence repeats (SSRs)

It was first discovered by Litt and Lutty in 1989 and termed it as Microsatellites, in the same year Tautz, also recognize it as a polymorphic DNA marker [34,35]. It consists of 1–6 bp tandem repeat motifs which are abundantly present in the coding and non-coding parts of the genome [36]. Among all the molecular markers developed and used in breeding programs, SSR are considered as markers of Choice [37]. SSRs are regarded as the premier genetic markers for studying genetic variability. This is attributed to their widespread presence throughout genomes, the capability to examine multiple loci, and their capacity to distinguish heterozygous individuals from homozygous ones based on their allele combinations at a specific locus i.e. codominant [38].

2.1.2.4 Single nucleotide polymorphism (SNPs)

Single Nucleotide Polymorphisms (SNPs) are individual base variations present in DNA. They can be substitutions (transitions or transversions) or insertions/deletions (InDels) of a single base. SNPs were found abundantly in coding and non-coding regions of genomes, with plants having around 1 SNP per 100-300 bases [39]. Common SNP genotyping methods include RFLP, CAPS, ligation, allele-specific hybridization, primer extension and invasive cleavage. These techniques rely on analyzing sequence databases to identify and characterize SNPs within a given genome. In RFLP, if a SNP creates or eliminates a restriction enzyme binding site on one allele, digestion will yield fragments of varying sizes compared to other alleles [40].

2.2 Marker Selection and Optimization

Marker selection and optimization is the process of carefully picking and refining the set of molecular markers used in genetic studies or breeding programs to ensure they are informative, efficient, and well-suited for the intended application.
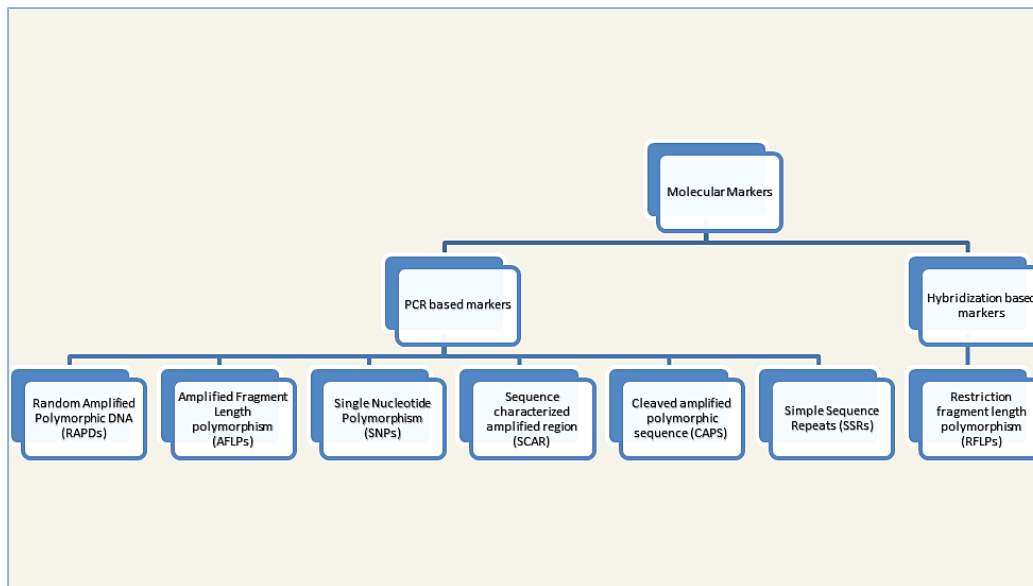
Fig. 1. Types of molecular markers used for diversity studies

2.2.1 Marker polymorphism and informativeness

Polymorphism refers to the presence of different forms or variations of a gene, or DNA sequence in a population. DNA polymorphism involves genetic variations in the DNA sequence among individuals, such as single nucleotide changes, insertions, deletions, or structural modifications. Polymorphisms contribute to genetic diversity, influence traits and disease susceptibility, and aid in understanding population genetics and evolution. Polymorphic DNA markers, regions exhibiting genetic diversity, are major tools for studies like linkage and association mapping [41].

When selecting genetic markers for research purposes, certain characteristics are preferred to ensure informative and reliable results. Markers with a greater number of alleles tend to be more informative and possess greater discriminatory power. Conversely, markers with skewed or unbalanced allele frequencies, where one or a few alleles are highly common while others are rare, tend to be less informative [42,43]. Additionally, markers with rare or low-frequency alleles may not be as useful, especially in smaller population samples, as these rare alleles may not be represented or could be lost due to genetic drift [44].

An even or uniform allele frequency distribution maximizes heterozygosity and provides the highest possible gene diversity or polymorphism information content (PIC) value for a given number of alleles [24,45]. This characteristic makes the marker more powerful for distinguishing genotypes and assessing genetic diversity [46]. Markers with evenly distributed allele frequencies are more reliable for genetic diversity studies, as they are less affected by sampling effects and genetic drift, especially in smaller populations [47].

2.2.2 Multiplexing and high-throughput genotyping

Multiplexing refers to the ability to simultaneously analyze multiple molecular markers in a single reaction or assay [38,48]. This approach offers several advantages, including increased efficiency, higher throughput, reduced cost and labor, and faster data generation [49]. Strategies for multiplexing include the use of fluorescently labeled primers [50] or the design of compatible primer sequences [51]. Careful marker selection is crucial for successful multiplexing to avoid allele size overlap, ensure compatibility with fluorescent dyes, and optimize reaction conditions [38,49]. High-throughput genotyping is the rapid and fully automated analysis of large number of molecular markers and samples using advanced technologies [43,46].

Several high-throughput genotyping platforms including, microarrays [52], next-generation sequencing (NGS) technologies [53], capillary electrophoresis systems [48], and other automated genotyping platforms [43]. These platforms can simultaneously analyze thousands of molecular markers and samples, Facilitating

increased efficiency, high speed, and low cost [43,46]. High-throughput genotyping is necessary for large-scale genetic studies like genome-wide association studies (GWAS) [54], genomic selection (GS) [55,56], and other applications that require extensive genotypic data generation. Crucial considerations include ensuring marker compatibility with the chosen genotyping platform, establishing robust data management and analysis pipelines, and carefully designing experiments to enable efficient data generation and analysis [43,48].

The key advantages of high-throughput genotyping lie in its capability to generate large volumes of genetic data rapidly and cost-effectively, facilitating advanced research.

2.3 Diversity Analysis Techniques

Diversity analysis techniques are statistical and computational methods used to analyze genetic variation data obtained from molecular markers [44,54]. These techniques help understand genetic relationships, population structure, and evolutionary patterns within and among groups of individuals or populations. By exploring genetic similarities and differences, diversity analysis techniques provide insights into genetic diversity, gene flow, adaptation and population differentiation [57]. Commonly employed techniques include,

2.3.1 Principal component analysis (PCA)

PCA is a data reduction and visualization method used to identify genetic clusters, outliers, and population relationships based on marker genotypes or allele frequencies [58,59]. It is a important tool for analyzing and interpreting large data sets generated from various molecular marker techniques [44].

2.3.1.1 Genetic diversity analysis

PCA can be used to study the genetic diversity within and between germplasms or breeding populations. By reducing the high-dimensional molecular marker data to a few principal components, it becomes easier to visualize the patterns of genetic variation and identify distinct groups or clusters of genotypes [44].

2.3.1.2 Population structure analysis

PCA can help identify subpopulations or genetically distinct groups within a breeding population, which is crucial for association mapping studies and controlling population structure [60,61].

2.3.1.3 Identification of redundant genotypes

PCA can help identify genetically redundant or highly similar genotypes within a germplasm collection, which can aid in the efficient management and conservation of genetic resources [62].

2.3.1.4 Selection of diverse parents

PCA can be used for selection of genetically diverse parents for hybridization and breeding programs, ensuring the maximum exploitation of genetic variation and the potential for transgressive segregation [62,63].

2.3.2 Cluster analysis (e.g., UPGMA, neighbor-joining)

Cluster analysis is a widely used technique in plant breeding and genetic diversity studies for grouping genotypes based on their similarity or dissimilarity in molecular marker data. Two commonly employed clustering methods, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ).

2.3.2.1 UPGMA (Unweighted pair group method with arithmetic mean)

UPGMA is a hierarchical clustering method that constructs a dendrogram based on the pairwise genetic distances or similarity measures between genotypes [64]. The key steps involved in UPGMA are:

a) Calculation of a distance or similarity matrix based on the molecular marker data.
b) Identification of the two most similar (or closest) genotypes or clusters.
c) Joining these two genotypes or clusters to form a new cluster and calculating the average distance between this new cluster and all other genotypes or clusters.
d) Repeating step 3 until all genotypes or clusters are joined into a single hierarchical tree.

UPGMA assumes a constant rate of evolution across all lineages, making it suitable for analyzing closely related genotypes or populations [44].

2.3.2.2 Neighbor-joining (NJ)

The Neighbor-Joining method which is introduced by Saitou and Nei, is a widely used clustering approach in plant breeding for genetic

diversity studies. It constructs an unrooted tree based on the pairwise genetic distances between genotypes. The key steps involved in NJ are:

a) Calculation of a distance matrix based on the molecular marker data.
b) Identification of the two genotypes or clusters that are "neighbors" (i.e., have the smallest sum of branch lengths).
c) Joining these two neighbors to form a new node and recalculating the distances between this new node and all other genotypes or clusters.
d) Repeating step 3 until all genotypes or clusters are joined into a single unrooted tree.

Unlike UPGMA, NJ does not assume a constant rate of evolution and can better handle situations where the rate of evolution varies among lineages [65,66].
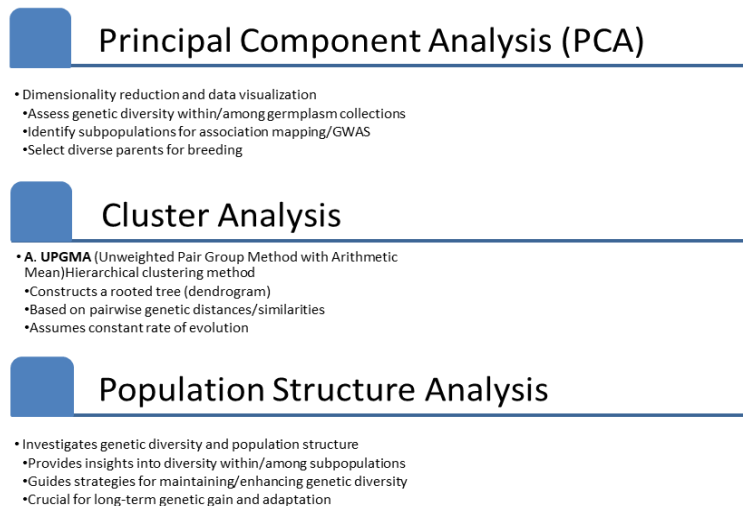
2.3.3 Population structure analysis

In diversity analysis and plant breeding programs. In diversity analysis and various plant breeding programs, population structure analysis based on molecular markers plays a pivotal role.

Population structure analysis using molecular markers can provide information about genetic diversity within and among subpopulations [44]. This information can be useful making strategies for maintaining and enhancing genetic diversity in breeding programs, which is crucial for long-term genetic gain and adaptation to changing environments.

Softwares that are used for population structure analysis and cluster analysis are as follows:-

Table 1. Software's used for diversity analysis

Sr. no.	Analysis Type	Software
1.	Cluster Analysis	NTSYSpc [67]
		Powermarker [68]
		DARwin [69]
		PAST [70]
2.	Population Structure Analysis	STRUCTURE [71]
		fastSTRUCTURE [72]
		BAPS [73]
		BAYES [74]
		GENECLUST [75]
		TESS [75]
		GENELAND [76]
		InStruct [77]
3.	Principal Component Analysis	DARwin [69]
		NTSYSpc [78]
		PAST [79]

Principal Component Analysis (PCA)
- Dimensionality reduction and data visualization
- Assess genetic diversity within/among germplasm collections
- Identify subpopulations for association mapping/GWAS
- Select diverse parents for breeding

Cluster Analysis
- **A. UPGMA** (Unweighted Pair Group Method with Arithmetic Mean)Hierarchical clustering method
- Constructs a rooted tree (dendrogram)
- Based on pairwise genetic distances/similarities
- Assumes constant rate of evolution

Population Structure Analysis
- Investigates genetic diversity and population structure
- Provides insights into diversity within/among subpopulations
- Guides strategies for maintaining/enhancing genetic diversity
- Crucial for long-term genetic gain and adaptation

Fig. 2. Diversity analysis techniques

3. APPLICATIONS OF MOLECULAR MARKERS IN GERMPLASM CHARACTERIZATION

3.1 Genetic Relationship and Phylogenetic Studies

The study of genetic relation between genes or species via the comparison of homologous DNA or protein sequences is known as molecular phylogeny. Genetic divergence from molecular evolution is shown by differences. To comprehend the evolutionary history and relatedness of taxa, researchers recreate phylogenetic trees [80].

3.1.1 Interspecific and intraspecific relationships

Molecular marker is the common tool for examining genetic relationship within and across species (intraspecific and interspecific) [80]. By utilizing marker data to determine genetic distances or similarities, these links may be valuable information about taxonomy, gene flow, and evolutionary history [57].

3.1.2 Identification of core collections and mini-core collections

Subsets of germplasm collections called core and mini-core collections show the greatest genetic diversity found in a species or gene pool [81,82]. DNA molecular markers, which are developed based on DNA polymorphisms are used in measuring genetic diversity and are notr affected by environmental interactions [83] hence they are more suitable for evaluating genetic diversity than phenotypic traits and in constructing a core collection [81].

3.2 Germplasm Management and Utilization

3.2.1 Maintenance and conservation of genetic resources

Molecular markers play a vital role in the maintenance and conservation of plant genetic resources [84]. They facilitate the accurate identification and characterization of accessions and management of germplasm collections. Furthermore, molecular markers are essential for assessing genetic diversity ensuring the effective conservation of genetic variability [82].

3.2.2 Targeted introgression of desirable traits

For targeted introgression of desirable traits from diverse germplasm sources, molecular markers are used in plant breeding programs [85]. Through marker-assisted selection (MAS) and backcrossing strategies, breeders can efficiently transfer and incorporate beneficial traits, such as biotic stress resistance, abiotic stress tolerance and improved yield, from wild relatives or diverse accessions into elite cultivars [86,87].

4. INTEGRATION OF MOLECULAR MARKERS WITH PHENOTYPIC DATA

4.1 Combining Genotypic and Phenotypic Data

The integration of marker data with morphological data is crucial for understanding the genetic basis of complex traits and enabling marker-assisted breeding [88]. By combining these two data, breeders can identify quantitative trait loci (QTLs) associated with traits of interest, estimate the effects of specific alleles or haplotypes, and develop statistical models for predicting phenotypes based on genotypic information [89]. Knowledge of genetic diversity in germplasm is vital for selecting parents to develop new varieties. Molecular markers provide a robust evaluation of diversity, unaffected by environmental factors, and enable thorough exploration of genetic variation across genomes. Leveraging molecular markers allows breeders to select divergent parents, maximize desirable trait combinations, characterize germplasm, and accelerate the development of improved varieties [90].

4.2 Genome-wide Association Studies (GWAS)

Genome-wide association studies (GWAS) are a great tool for analyzing the genetic makeup of complex phenotypes because they combine phenotypic data with high-density molecular markers [91]. It involves looking across the whole genome for correlations between genetic markers and trait variations, which makes it possible to find the genes or genomic areas that govern traits of interest [92,93].

4.3 Genomic Prediction and Selection

Genomic prediction and selection methods integrate molecular marker data with

morphological data to predict the breeding values or genetic potentials of individuals based on their genotypes [55,94]. These approaches utilize statistical models and machine learning algorithms to capture the combined effects of multiple markers across the genome, enabling the selection of superior individuals without extensive phenotyping [95,96]. Genomic selection has a high impact on plant breeding by accelerating the breeding cycle and improving genetic gains.

5. CHALLENGES AND FUTURE PERSPECTIVES

5.1 Limitations of Current Molecular Marker Technologies

While molecular markers have revolutionized plant genetics and breeding, the existing technologies have certain limitations. These include relatively low throughput and high costs associated with marker development and genotyping, limited genomic coverage, and challenges in interpreting and utilizing marker data for complex traits [43]. Additionally, the presence of genetic complexity, such as polyploidy and structural variations, can pose challenges for accurate marker-based analysis [57].

5.2 Emerging Technologies

The emergence of next-generation sequencing (NGS) technology has opened various possibilities for marker discovery, genotyping, and genomic analysis [43,53]. NGS-based approaches, such as Genotyping-by-sequencing (GBS) and whole-genome resequencing enable the efficient discovery and analysis of thousands to millions of molecular markers across the entire genome in a single process [97,98]. These technologies offer higher resolution, reduced costs, and increased throughput, helping in genome-wide association studies (GWAS) and genomic selection (GS) in various crop improvement programs [99].

5.3 Integrating Molecular Markers with other Omics Data

Combining molecular marker data with other omics data like transcriptomics, proteomics, and metabolomics is a new area of research. By combining information from these different omics areas, researchers can better understand the molecular basis of complex traits and make more accurate predictions for genetic improvement [55,96].

5.4 Practical Applications and Impact on Crop Improvement

Molecular markers have revolutionized crop improvement by enabling marker-assisted selection, genomic selection strategies, and the development of genetically modified crops with desirable traits [100,101]. MAS enables the efficient introgression of desirable traits from diverse germplasm sources, while genomic selection accelerates the breeding cycle and increases genetic gains [86,95]. Additionally, molecular markers have facilitated the identification and manipulation of key genes or regulatory regions, leading to the development of improved crop varieties with enhanced yield, nutritional quality, and resilience to biotic and abiotic stresses [102,103].

6. CONCLUSION

Molecular markers have emerged as invaluable tools for diversity analysis and germplasm characterization in plant breeding programs. This review paper highlights the use of principal component analysis (PCA), cluster analysis methods (UPGMA and Neighbor-Joining), and population structure analysis (model-based approaches like STRUCTURE) to assess genetic diversity, identify subpopulations, and select diverse parents based on molecular marker data. Molecular markers act as direct method for measurement of genetic variation at the DNA level, providing effective means for conservation, and utilization of genetic resources. They play a crucial role in various applications, including association mapping, genomic selection, introgression breeding, and monitoring genetic diversity for long-term genetic gain and adaptation.

7. FUTURE PROSPECTUS

Future research directions in the utilizing molecular markers for assessing diversity and characterizing germplasm (plant genetic resources) now involves the use of advanced high-throughput sequencing technologies, genotyping methods based on sequencing data, exploration of novel molecular marker systems (epigenetic markers, structural variations), development of advanced statistical and computational methods for data analysis, incorporation of molecular markers and

population structure information into genomic selection and prediction models, application to orphan crops and underutilized species, integration with other omics approaches (transcriptomics, proteomics, metabolomics), and fostering international collaborations and data-sharing initiatives. All these advancements in plant breeding will promote the sustainable use of genetic resources and contribute to global food security.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Cole CT. Genetic variation in rare and common plants. Annual Review of Ecology, Evolution, and Systematics. 2009;34(1):213-237.
2. Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Baloch FS. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology & Biotechnological Equipment. 2018;32(2): 261-285.
3. Zebire DA. Applications of molecular markers in Genetic Diversity Studies of maize. Nigerian Journal of Biotechnology. 2020;37(1):101-108.
4. Brown WL. Genetic diversity and genetic vulnerability—An appraisal. Economic botany. 1983;37:4-12.
5. Ennos R. Genetic variation and conservation of British native trees and shrubs: Current knowledge and policy implications; 2000.
6. Swingland IR. Biodiversity, definition of. Encyclopedia of biodiversity. 2001;1:377-391.
7. Frankham R, Briscoe DA, Ballou JD. Introduction to conservation genetics. Cambridge University Press; 2002.
8. Hughes AR, Inouye BD, Johnson MT, Underwood N, Vellend M. Ecological consequences of genetic diversity. Ecology letters. 2008;11(6):609-623.
9. Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, Glenn KC. Genetic diversity is indispensable for plant breeding to improve crops. Crop Science. 2021;61(2):839-852.
10. Bhandari HR, Bhanu AN, Srivastava K, Singh MN, Shreya HA. Assessment of genetic diversity in crop plants-an overview. Adv. Plants Agric. Res. 2017; 7(3):279-286.
11. Raza A, Razzaq A, Mehmood SS, Zou X, Zhang X, Lv Y, Xu J. Impact of climate change on crops adaptation and strategies to tackle its outcome: A review. Plants. 2019;8(2):34.
12. Raffard A, Santoul F, Cucherousset J, Blanchet S. The community and ecosystem consequences of intraspecific diversity: A meta-analysis. Biological Reviews. 2019;94(2):648-661.
13. Ramanatha Rao V, Hodgkin T. Genetic diversity and conservation and utilization of plant genetic resources. Plant cell, tissue and organ culture. 2002;68:1-19.
14. Dwivedi SL, Ceccarelli S, Blair MW, Upadhyaya HD. Are AK, Ortiz R. Landrace germplasm for improving yield and abiotic stress adaptation. Trends in plant science. 2016;21(1):31-42.
15. Food and agriculture organization of the United Nations. Genebank Standards for Plant Genetic Resources for Food and Agriculture; 2014.
Retrieved from: https://www.fao.org/3/ i3704e/i3704e.pdf
16. Begna T, Begna T. Role and economic importance of crop genetic diversity in food security. International Journal of Agricultural Science and Food Technology. 2021;7(1):164-169.
17. Salgotra RK, Chauhan BS. Genetic diversity, conservation, and utilization of plant genetic resources. Genes. 2023; 14(1):174.
18. Saad NSM, Neik TX, Thomas WJ, Amas JC, Cantila AY, Craig RJ, Batley J. Advancing designer crops for climate resilience through an integrated genomics approach. Current Opinion in Plant Biology. 2022;67:102220.
19. Mondal R, Kumar A, Gnanesh BN. Crop germplasm: Current challenges, physiological-molecular perspective, and advance strategies towards development of climate-resilient crops. Heliyon. 2023; 9(1).
20. Mondini L. Noorani A, Pagnotta MA. Assessing plant genetic diversity by molecular tools. Diversity. 2009;1(1):9-35.
21. Mirzaei S. Application of molecular markers in plant sciences; An overview. Central Asian Journal of Plant Science Innovation. 2021;1(4):192-200.
22. Amom T, Nongdam P. The use of molecular marker methods in plants: a

review. International Journal of Current Research and Review. 2017;9(17):1-7.

23. Dhutmal RR, Mundhe AG, More AW. Molecular marker techniques: A Review. Int. J. Curr. Microbiol. Appl. Sci. 2018;6: 816-825.

24. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics. 1980;32(3):314.

25. Jiang GL. Molecular markers and marker-assisted breeding in plants. Plant breeding from laboratories to fields. 2013;3:45-83.

26. Singh BD, Singh AK. Marker-assisted plant breeding: principles and practices; 2015.

27. Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research. 1990;18(22), 6531-6535.

28. Welsh J, McClelland M. Fingerprinting genomes using PCR with arbitrary primers. Nucleic acids research. 1990;18(24):7213-7218.

29. Babu KN, Sheeja TE, Minoo D, Rajesh MK, Samsudeen K, Suraby EJ, Kumar IPV. Random amplified polymorphic DNA (RAPD) and derived techniques. Molecular Plant Taxonomy: Methods and Protocols. 2021;219-247.

30. Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F. Van de Wiel C, Karp A. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular breeding. 1997;3:381-390.

31. Lynch M, Walsh B. Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer. 1998;1:535-557.

32. Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M, Zabeau M. AFLP: A new technique for DNA fingerprinting. Nucleic acids research. 1995;23(21):4407-4414.

33. Madhumati B. Potential and application of molecular markers techniques for plant genome analysis. Int J Pure App Biosci. 2014;2(1):169-188.

34. Litt M, Luty JA. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American journal of human genetics. 1989;44(3):397.

35. Tautz D. Hypervariability of simple sequences as a general source for

36. Tautz D, Ren M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic acids research. 1984;12(10):4127-4138.

37. Mir RR, Varshney RK. 10 Future prospects of molecular markers in Plants. Molecular markers in plants. 2013;169.

38. Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Petit RJ. Current trends in microsatellite genotyping. Molecular ecology resources. 2011;11(4): 591-611.

39. Xu Y. Molecular plant breeding. Cabi; 2010.

40. Sobrino B, Brión M, Carracedo A. SNPs in forensic genetics: A review on SNP typing methodologies. Forensic science international. 2005;154(2-3):181-194.

41. Hildebrand CE, Torney DC, Wagner RP. Informativeness of polymorphic DNA markers. Los Alamos Science. 1992; 20(20):100-102.

42. Vieira MLC, Santini L, Diniz AL, Munhoz CDF. Microsatellite markers: what they mean and why they are so useful. Genetics and molecular biology. 2009; 32(3):308-328.

43. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Molecular plant. 2017;10(8):1047-1064.

44. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop science. 2003;43(4):1235-1248.

45. Nei M. Analysis of gene diversity in subdivided populations. Proceedings of the national academy of sciences. 1973; 70(12):3321-3323.

46. Mammadov J. Aggarwal R, Buyyarapu R, Kumpatla S. SNP markers and their impact on plant breeding. International journal of plant genomics; 2012.

47. Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF. Core hunter: an algorithm for sampling genetic resources based on multiple genetic measures. BMC bioinformatics. 2009;10(1):1-12.

48. Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop

improvement. Molecular Breeding. 2014; 33(1):1-14.

49. Hayden MJ, Nguyen TM, Waterman A, Chalmers KJ. Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. BMC genomics. 2008;9(1):1-12.

50. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology. 2000;18(2):233-234.

51. Henegariu O, Heeresen NA. Dlouhy SR. Vance GH, Vogt PH. Multiplex PCR: Critical parameters and step-by-step protocol. Biotechniques. 1997;23(3):504-511.

52. Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A. Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proceedings of the National Academy of Sciences. 2004; 101(26):9915-9920.

53. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics. 2011;12(7):499-510.

54. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Han B. Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genetics. 2010;42(11): 961-967.

55. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J. Genomic selection in plant breeding: Methods, models, and perspectives. Crop Science. 2017;57(3):1058-1067.

56. Huang, X., Han, B., & Sunderman, D. Analysis of genetic diversity in sweet potato (Ipomoea batatas) germplasm resources explored by genomic markers. Genetics and Molecular Biology. 2015; 34(4)590-596.

57. Dufresne F, Stift M, Vergilino R, Mable BK. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. Molecular Ecology. 2019;23(1):40-69.

58. Jombart T, Pontier D, Dufour AB. Genetic markers in the playground of multivariate analysis. Heredity. 2009;102(4):330-341.

59. Meirmans PG. Seven common mistakes in population genetics and how to avoid them. Molecular Ecology. 2015;24(13): 3223-3231.

60. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics. 2006;38(8):904-909.

61. Zou C, Fu Y, Song W, Mei X, Liu J, Tang H. Principal component analysis combined with a model-based approach dissects genetic structure of germplasm populations: A case study of rice. Crop Science. 2021;61(5):3030-3041.

62. Chakravarthi BK, Naravaneni R. SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa*. L). African Journal of Biotechnology. 2006; 5(9):684-688.

63. Malosetti M, Ribaut JM, van Eeuwijk FA. The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. Frontiers in physiology. 2007;4:44.

64. Sneath PHA. The principles and practice of numerical classification. Numerical taxonomy. 1973;573.

65. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987;4(4):406-425.

66. Takezaki N, Nei M. Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. Genetics, 2008; 178(1):385-392.

67. Ghasemi AR, Golparvar AR, Isfahani MN. Analysis of genetic diversity of sugar beet genotypes using random amplified polymorphic DNA marker. Genetika. 2014; 46(3):975-984.

68. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005; 21(9):2128-2129.

69. Perrier X. DARwin software; 2006. Available:http://darwin.cirad.fr/darwin.

70. Kumar H, Kaur G, Banga S. Molecular characterization and assessment of genetic diversity in sesame (*Sesamum indicum* L.) germplasm collection using ISSR markers. Journal of crop improvement. 2012 Jul 1;26(4):540-57.

71. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2):945-959.

72. Raj A, Stephens M, Pritchard JK. Fast structure: Variational inference of population structure in large SNP data sets. Genetics. 2014;197(2):573-589.

73. Corander J, Waldmann P, Marttinen P, Sillanpää MJ. BAPS 2: Enhanced possibilities for the analysis of genetic population structure. Bioinformatics. 2004; 20(15):2363-2369.

74. Pella J, Masuda M. The gibbs and split merge sampler for population mixture analysis from genetic data with incomplete baselines. Canadian Journal of Fisheries and Aquatic Sciences. 2006;63(3):576-596.

75. François O, Ancelet S, Guillot G. Bayesian clustering using hidden Markov random fields in spatial population genetics. Genetics. 2006;174(2):805-816.

76. Guillot G, Santos F, Estoup A. Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. Bioinformatics. 2008;24(11): 1406-1407.

77. Gao H, Williamson S, Bustamante CD. A markov chain monte carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics. 2007;176(3):1635-1651.

78. Ma M, Lei E, Meng H, Wang T, Xie L, Shen D, Lu B. Cluster and principal component analysis based on SSR markers of Amomum tsao-ko in Jinping County of Yunnan Province. In AIP Conference Proceedings. (Vol. 1864, No. 1). AIP Publishing; 2017.

79. Hammer O. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron. 2001;4:9.

80. Patwardhan A, Ray S, Roy A. Molecular markers in phylogenetic studies-a review. Journal of Phylogenetics & Evolutionary Biology. 2014;2(2):131.

81. Odong TL, Jansen J, Van Eeuwijk FA, Van Hintum TJ. Quality of core collections for effective utilisation of genetic resources review, outlook and guidelines. Biotechnology and Genetic Engineering Reviews. 2013;29(1):135-156.

82. Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: An overview of its analytical perspectives. Genetics Research International. 2015;1-14.

83. Liu X, Liu H, Ma L, Li X, Xu C, Su H, Fan Y. Construction of sugarcane hybrids core collection by using stepwise clustering sampling approach with molecular marker data. Acta Agronomica Sinica. 2014; 40(11):1885-1894.

84. Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, Varshney RK. Emerging genomic tools for legume breeding: current status and future prospects. Frontiers in Plant Science. 2016;7:455.

85. Collard BCY, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences. 2008; 363(1491):557-572.

86. Xu Y, Crouch JH. Marker-assisted selection in plant breeding: from publications to practice. Crop Science. 2008;48(2):391-407.

87. Langridge P, Waugh R. Harnessing the potential of germplasm collections. Nature Genetics. 2019;51(2):200-201.

88. Xu Y, Li P, Yang Z, Xu C. Genetic mapping of quantitative trait loci in crops. The Crop Journal. 2017;5(2):175-184.

89. Bernardo R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Science. 2016;56(1):53-65.

90. Kumawat G, Singh G, Gireesh C, Shivakumar M, Arya M, Agarwal DK, Husain SM. Molecular characterization and genetic diversity analysis of soybean (*Glycine max* (L.) Merr.) germplasm accessions in India. Physiology and molecular biology of plants. 2015;21:101-7.

91. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: A review. Plant Methods. 2013;9(1):1-9.

92. Huang X, Han B. Natural variations and genome-wide association studies in crop plants. Annual review of plant biology. 2014;65:531-551.

93. Juliana P, Singh RP, Singh PK, Crossa J, Huerta-Espino J, Lan C, Sorrells ME. Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. Theoretical and Applied Genetics. 2018;131(8):1691-1701.

94. de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 2013;193(2):327-345.

95. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819-1829.

96. Hickey JM, Chiurugwi T, Mackay I, Powell W, Eggen A, Kilian A, Gorjanc G. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. Nature Genetics. 2017;49(9):1297-1303.

97. Deschamps S, Llaca V, May GD. Genotyping-by-sequencing in plants. Biology. 2012;1(3):460-483.

98. He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z. Genotyping-by-sequencing BS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science. 2014;5:484.

99. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Han B. High-throughput genotyping by whole-genome resequencing. Genome Research. 2014; 24(7):1183-1183.

100. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: Features and applications. Trends in Biotechnology. 2009;23(1):48-55.

101. Langridge P, Waugh R. Harnessing the potential of germplasm collections. Nature Genetics. 2019;51(2):200-201.

102. Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. Science. 2010; 327(5967):818-822.

103. Abberton M, Batley J, Bentley A, Bryant J, Cai H, Cairns J, Edwards, D. Global agricultural intensification during climate change: a role for genomics. Plant Biotechnology Journal. 2016;14(4):1095-1098.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/116714