



Molecular Markers in Breeding of Crops: Recent Progress and Advancements

Gudeta Dida*

Ethiopian Institute of Agricultural Research, National Agricultural Biotechnology Research Center, P.O. Box 249, Holeta, Ethiopia

***Corresponding Authors:** Gudeta Dida, Ethiopian Institute of Agricultural Research, National Agricultural Biotechnology Research Center, P.O. Box 249, Holeta, Ethiopia

Abstract: A conventional plant breeder faces the challenge of how to more effectively and efficiently perform selection and accelerate breeding progress to satisfy the requirements of changing demands for crop cultivars. However, with the development and advancement of molecular marker technology, the fate of plant breeding has shifted from year to year. Recently, different types of molecular markers have been developed, and advancements in sequencing technologies have greatly increased plant improvement. To further our understanding of molecular markers, several reviews have been published in recent decades. However, with the advancement of newly emerging technologies and techniques, the reviewers did not discuss several recently emerged technologies and techniques in plant breeding. Therefore, this article is intended to be reviewed as an overview of recent breakthroughs in molecular markers and their applications in plant breeding for early and senior researchers with little or no experience with molecular markers. The progress made in molecular plant breeding, genetics, genomic selection, gene pyramiding, marker assisted selection, QTL mapping, and gene mapping has contributed to a deeper understanding of molecular markers, provided deeper insights into the variability available for crops, and considerably supplemented current breeding techniques. Next-generation sequencing technologies assist in the identification of novel genetic markers for complex and unstructured populations through genotyping-by-sequencing, gene mapping, QTL mapping, and association mapping. Altogether, the classification of molecular markers and their potential application in plants are discussed.

Keywords: Molecular marker, QTL mapping, MAS, genomic selection, genetic marker

1. INTRODUCTION

Plant breeding has a long history of development, especially since the fundamental principles of inheritance were established in the late 19th and early 20th centuries. It has evolved into a major part of agricultural research, which combines elements of science and the arts. Traditional breeding methods have proven to be quite effective in the development of cultivars and germplasm. However, in conventional breeding, subjective analysis and empirical selection continue to play a significant role. Marker assisted breeding techniques have brought great opportunities and prospects for conventional breeding methods. Marker-assisted breeding (MAB), like transgenic breeding or genetic manipulation, cannot replace conventional breeding and is only a supplement to it as a new member of the plant breeding methods. Due to High costs and technical or equipment demands of marker assisted breeding, there will remain significant barriers to widespread use in the near future, particularly in less developed and developing countries (Collard and Mackill, 2008; Ribaut *et al.*, 2010). Therefore, incorporating MAB into traditional breeding efforts will be a promising technique for crop improvement in the future. It can be expected that the drawbacks of MAB will be gradually improved as its theory, technology, and applications are further developed in the near future. This should result in the widespread adoption and usage of MAB in actual breeding projects for a wider range of crop species in different countries (Nadeem *et al.*, 2018).

Genetic markers are broadly grouped into three categories: morphological markers, biochemical markers, and DNA/molecular markers. Morphological and biochemical (monoterpenes, alloenzymes, and other protein marker markers) are types of classical markers. Molecular markers are restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs), expressed sequence tagged polymorphism, single-nucleotide polymorphism

(SNP) and diversity array technology (DArT) markers (Jiang *et al.*, 2013). With the advancement of molecular biology, a new form of marker based on polymorphisms in the DNA sequence, known as molecular markers, has emerged, expanding the possibility for new challenges in plant breeding. The evolution of molecular markers changes the fate of plant breeding. A molecular marker is a gene or DNA sequence that controls a certain gene or characteristic at a known chromosome location. Molecular markers are closely related to the target gene and they act as signs or flags (Collard *et al.*, 2005).

Molecular markers are widely dispersed throughout the genome of an organism, they are not affected by environmental factors, and they can be found in any tissue at any developmental stage. From their development, they were used in agriculture through the construction of genetic maps in crop species, the association between molecular markers and important agronomic traits, the dissection of quantitative traits, and the positional cloning of genes of interest, as well as fingerprinting, and they also increased the efficiency and speed of breeding programs. Molecular markers are the most appropriate tools for evaluating genetic diversity, allowing for the selection of the most suitable parental lines in breeding programs, the management of germplasm collections, and varietal identification, in addition to estimating genetic distances and molecular cloning (Phougat *et al.*, 2018). Furthermore, new advances in molecular markers and genome sequencing provide a good chance to study the genetic diversity of a large germplasm collection (Nawaz *et al.*, 2013). In the study of plant evolution and comparative genomics, genetic diversity assessment is highly useful in understanding the organization of distinct populations (Liu *et al.*, 2015). Genetic markers have been successfully evolved in the characterization of genetic diversity studies and the classification of genetic material. DArT markers and SNP markers are the most frequently used markers for the study of genetic diversity in various crops (Blotch *et al.*, 2017).

Conventional plant breeders face the challenge of how to more effectively and efficiently perform selection and accelerate breeding progress in order to meet the demands of changing crop cultivar markets. Molecular marker-assisted breeding of crops, the application of molecular biotechnologies (DNA markers) to practical breeding and selection, is a novel strategy and a powerful methodology for plant improvement (Salgotra and Stewart, 2020). When compared to traditional breeding procedures, it has a number of advantages. Recently, marker-assisted breeding has received increasing attention and has been extensively used in different crop species. From a plant breeder's point of view, this article addresses the general classification of molecular markers and potential applications of these markers as powerful tools in plant breeding, including marker-assisted selection, marker-assisted gene pyramiding, QTL mapping, gene mapping, genomic selection, and other molecular breeding tools.

2. CLASSIFICATION OF GENETIC MARKERS

A genetic marker is a gene or DNA sequence with a known location on a chromosome controlling a particular gene that can detect variation in either a protein or DNA sequence. They act as signs or flags to keep track of an individual, a tissue, a cell, a nucleus, DNA sequences, or a gene. Nowadays, genetic markers are used in both basic plant research and plant breeding programs to characterize plant germplasm, for gene isolation, marker-assisted introgression of favorable alleles, the production of improved varieties, and to obtain information about the genetic variation of populations (Henry, 2001; Dida *et al.*, 2021). Genetic markers can be categorized into three broad classes: morphological markers, those based on visual assessable traits; biochemical markers; those based on gene products and molecular markers; and those based on a DNA assay. Each of the various marker systems utilized for various plant breeding techniques has its own specific merits and demerits (Henry, 2001).

2.1. Morphological markers

During conventional breeding programs in the past several decades, breeders used visible markers such as, leaf shape, flower color, pubescence color, pod color, seed color, hilum color, awn type and length, fruit shape, flesh color, stem length, and other important agronomic traits. Morphological markers are very easy to use. There is no requirement for specific instruments, and they do not require any specialized biochemical or molecular methods. Generally, morphological markers represent genetic polymorphisms that are easily visible and identified at specific stages and times (Jiang, 2013). In the history of plant breeding, humans have successfully used different morphological markers to determine the variation for utilization in crop breeding. However, morphological markers may be affected by the environment. They are limited in number and they appear at specific plant growth stages and durations. Generally, it is incompletely linked with the gene of interest. Its phenotypic expression may be

dependent on growth stage. These markers are rare in a natural population and show an extremely low level of polymorphism. In addition, they do not cover the entire genome but are located in certain genomic loci in which the genes are concentrated. Despite these exceptions, morphological markers still remain a relevant and very useful scientific tool in genetic and breeding practices (Chesnokov *et al.*, 2020).

2.2. Biochemical (protein markers)

Biochemical markers, also known as isozymes, are multi-molecular forms of enzymes that are coded by different genes, but have the same functions. They are allelic variations of proteins. Gene and genotypic frequencies can be determined with biochemical markers (Bayley, 1983). The marker reflects the products of various alleles rather than various genes because the difference in electrophoretic mobility is caused by point mutation as a result of amino acid substitution (Xu, 2010). Therefore, biochemical markers can be mapped on to chromosomes and then used as markers to map other genes. Biochemical markers have been successfully utilized in the study of genetic diversity, linkage map construction, population genetic structure, and population gene flow. They are codominant, inherited patterns that are simple to use, simple to assay, cost- and time-effective, and do not require any sophisticated techniques, making them one of the appropriate markers. However, they are fewer in number, they detect less polymorphism, and they are affected by various extraction methodologies; they depend on plant tissues and different plant growth stages (Mondini *et al.*, 2009).

2.3. Molecular markers

Molecular markers are based on naturally occurring polymorphisms in DNA sequences due to base pair deletions, insertions, and substitutions. Molecular markers are superior to both morphological and biochemical markers because they are highly polymorphic, relatively simple to detect, abundant throughout the genome, even in highly inbred cultivars, and completely independent of environmental conditions, and can be detected at any stage of plant development (Bhandari *et al.*, 2017). However, the major disadvantage is the need for technically more complex equipment.

The rapid development of molecular techniques over the last few decades has now offered a good technical approach for plant genotyping or genome analysis. Which technique is best depends on (i) the amount of genetic polymorphism required, (ii) the analytical techniques or statistical approaches available for the techniques, and (iii) the pragmatics of time and material and technique costs (Parker *et al.*, 1998). The discovery of the polymerase chain reaction (PCR) was a landmark in molecular marker evolution and has proved to be a unique process for the development and utilization of a battery of new very sensitive and quick approaches, such as AFLP or microsatellites (SSR) (Paglia and Morgante, 1998). PCR is an *in vitro* technique that uses specific primers to amplify a specific DNA region of a known DNA sequence. Orozco-Castillo *et al.* (1994) observed the power of the PCR technology for the generation of genetic markers for plant breeding.

2.3.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) was developed first and was initially used for human genome mapping (Botstein *et al.*, 1980). Later, RFLP markers are one of the most important tools for plant genome mapping (Jiang 2013), and they are classified as hybridization-based markers. RFLP involves the extraction of genomic DNA followed by its digestion with specific restriction enzymes that cut the DNA into fragments. RFLP results when there is variation in restriction enzyme cleavage sites, arising due to base substitutions, insertions, deletions, or translocations in the genomic DNA (Gupta *et al.*, 2002).

The main advantages of RFLP markers are their high reproducibility, high genomic wide abundance, codominant inheritance, association with a trait of interest, and good transferability, which provide locus specific markers for synteny studies (Kesawat and Das, 2009; Jiang, 2013). However, there are several limitations to the RFLP marker: it requires the presence of a high quantity and quality of template DNA, radioactive toxic reagents, and highly qualified technical persons. RFLP markers can be applied in diversity studies, phylogenetic analysis, and fingerprinting, ranging from individuals within populations to very similar species. It is widely applied in gene mapping because of its high genomic wide abundance, high polymorphism, availability of different restriction enzyme sites and random distribution throughout the genome of an organism (Kesawat and Das, 2009; Madhumati, 2014).

2.3.2 Random Amplified Polymorphic DNA (RAPD)

Random Amplified Polymorphic DNA (RAPD) is a type of PCR-based marker that is used to amplify genomic DNA with a single primer of arbitrary nucleotide sequence, usually 10bp long (Williams *et al.*, 1990). RAPD Primers detect polymorphisms with no species specificity and no prior sequence information is required. The polymorphism detected method is used as a genetic marker to construct genetic maps. RAPD polymorphisms arise when genomic regions vary in the presence or absence of complementary primer annealing sites due to insertion or deletion between two priming sites, which gives various lengths of fragments. The main advantages of the RAPD marker include: (i) fast to assay and efficient for diversity analysis and genetic linkage map construction; (ii) no need for high quality and quantity of template DNA; and (iii) simpler and cost-effective. Polymorphism present can be detected in the gel electrophoresis by confirming the presence or absence of specific bands (Jiang, 2013). However, there are various limitations and considerations in RAPD analysis, which include specificity of the marker in genome scanning (Hadrys *et al.*, 1992), reproducibility of amplification products (Williams *et al.*, 1990), and unclear and non-reproducible fragments (Pan *et al.*, 1997).

2.3.3 Amplified Fragment Length Polymorphism (AFLP)

AFLP markers are hybrid markers in RFLP and PCR-based methods for the rapid screening of genetic diversity studies and intra-specific variation. It is a fingerprinting technique for genomic DNA of any origin or complexity and rapidly generates a number of fragmentations of the genomic DNA using specific restriction enzymes. The strength of AFLPs includes their high genomic abundance, generating highly reliable and reproducible data, being highly polymorphic; generating many informative bands per reaction, a small amount of template DNA being needed, and the fact that no sequence information for primer construction is required (Saal and Wricke, 2002). Possible reasons for AFLP polymorphisms are: (i) sequence variations in a restriction site, (ii) insertions or deletions within an amplified fragment, and (iii) differences in the nucleotide sequence immediately adjoining the restriction site. AFLPs have been used for the analysis of genetic diversity, DNA fingerprinting, the construction of linkage maps (Cho *et al.*, 1998), and to locate traits of interest (Hartl *et al.*, 1999).

2.3.4 Microsatellite or Simple Sequence Repeat (SSR)

SSRs are very short motifs (about 1-6bp) usually characterized by a high degree of repetition and occur at many thousand loci in the nuclear genome (Singh *et al.*, 2010). Microsatellites are also present in the chloroplast and mitochondria of eukaryotic and prokaryotic organisms. Several studies have also reported the presence of SSR markers in protein-coding genes and expressed sequence tags (ESTs) (Nadeem *et al.*, 2018). Microsatellites can be mononucleotides, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, or hexanucleotides, and contain a low degree of repetition per locus and are highly polymorphic. The development of microsatellite markers involves the development of small-insert genomic SSR libraries and then the detection of specific microsatellites. After this, the identification of favorable regions for primer design is done, and then PCR is performed. Interpretation of banding patterns is analyzed and assessment of PCR products is performed for determination of polymorphism of the marker (Rode *et al.*, 1998). The higher variability of SSRs among closely related organisms makes them an informative and popular choice of markers for a wide range of applications in population and evolutionary biology (Chen *et al.*, 2009), which include estimating genetic diversity, studying population structure and gene flow, and developing gene mapping. Microsatellite markers have several important advantages, including their abundance with uniform genome coverage, enormous allelic diversity, hyper variability, codominant inheritance, ease of detection by PCR using a pair of flanking primers, and the need for only a small amount of template DNA (Chen *et al.*, 2009). These unique characteristics of this marker make it a favorite for genetic mapping studies, marker assisted selection (MAS), genetic diversity studies, QTL mapping, and other breeding purposes.

2.3.5 Single Nucleotide Polymorphism (SNP) marker

Single nucleotide polymorphism (SNP) markers arise because of point mutations like substitutions or transversions in single nucleotides that occur at a specific position in the genome sequence where each variation is present at a level of more than 1% in the population. SNP markers can provide the simplest and maximum number of markers, and these may be transitions (C/T or G/A) or transversions (C/G, A/T, C/A or T/G) on nucleotide substitution (Nadeem *et al.*, 2018). SNPs are the most abundant form of DNA marker developed and could reach a higher density than any other type of marker (Wurschum

et al., 2013). With the advent of SNP markers, the possibility of simultaneous analysis of a set of loci becomes more real. A SNP is created when a single nucleotide base in a DNA sequence is replaced with a different nucleotide base. The SNP markers are based on the most fundamental alterations of the DNA molecule and mutations in the bases of a unique chain of nitrogenous bases (Adenine, Cytosine, Guanine and Thymine). SNPs are extremely abundant in the genomes of individuals (Wurschum *et al.*, 2013). Recent technological development of DNA sequencing technology has now been widely applied to develop massive genotyping arrays, which allow fast and efficient detection of SNP markers for large numbers of individual organisms (Ganal *et al.*, 2014).

Nowadays, SNP markers are the most common marker type applied in modern plant breeding techniques. They have a wide range of applications in different breeding of crops, such as identification of plant varieties and cultivars, genetic diversity studies, QTL mapping, construction of high density genetic linkage maps, and genome wide association studies (Deplore *et al.*, 2008; Han *et al.*, 2016). These markers are highly, associated with the traits of interest and have a low cost per data point. These facts, together with the development of genomic selection algorithms, have improved the precision of the selection methods employed in plant breeding programs for several plant species. However, in order for a chip to be developed for a novel species, it is important that the SNPs be known a priori (Crossa *et al.*, 2010).

2.3.6 Genotyping by sequencing (GBS)

The GBS marker is a versatile and straightforward approach that is widely used nowadays. GBS was developed in the Buckler lab under the Illumina NGS platform and it is a novel application of NGS protocols for discovering and genotyping SNPs. Advancement in NGS techniques has reduced the sequencing costs, assuring the successful application of GBS for large genome species with a great magnitude of diversity (Elshire *et al.*, 2011). There are two types of GBS techniques: (1) *Digestion of restriction enzyme*: this method is mostly employed in marker-assisted selection (MAS) programs to identify novel markers, and no specific SNPs are identified. In this technique, DNA is digested with one or two specific restriction enzymes prior to the ligation of adapters. (2) *Multiplex enrichment PCR*: In this method, for the amplification of points of interest, particular PCR primers are developed. In contrast to the digestion in the restriction enzyme method, a complete set of SNPs are identified for a genome section.

GBS was originally created to examine high-resolution associations in maize, but it is currently employed in a variety of other species with complicated genomes. The primary advantages of GBS are: (I) lower costs than previous approaches, making it a novel technique for identifying SNPs in various animals and crops. (II) This method yields good results in the characterization of germplasm, population studies, and crop breeding (Poland *et al.*, 2012). (III) GBS generates a large number of SNPs that can be used for genotyping and genetic analysis. (IV) Fewer samples are required, and (V) fewer PCR and purification sets are required (Beissinger *et al.*, 2013).

2.3.7 Diversity array Technology (DArT Seq)

DArT Seq is a technology that allows a great opportunity for the genotyping of polymorphic loci (in hundreds to thousands), that are dispersed throughout the genome. It is highly reproducible microarray hybridization technology. No preceding sequence information is required for the identification of loci for a trait of interest. (Wenzl *et al.*, 2004). The most significant advantages of this technology are its high throughput and low cost. A single-reaction assay can genotype thousands of genomic loci to find polymorphic markers with this method. Genotyping can be done with as little as 50–100ng of genomic DNA. The scoring and the discovery of markers are both done on the same platform. There is no need for a specific genotyping assay after the identification of a marker, except to begin assembling polymorphic markers into an array of a single genotype. Genotyping arrays with polymorphic markers are routinely used for genotyping (Nadeem *et al.*, 2018).

3. USES OF MOLECULAR MARKERS IN PLANT BREEDING

3.1. Genetic diversity assessment

Plant genetic diversity analysis is a critical component of plant genetics, breeding, conservation, and evolution (Peterson *et al.*, 2014). It enables plant breeders to create new and improved cultivars with desirable traits, such as farmer-preferred traits (high yield potential and large seed size, for example)

and breeder-preferred traits (pest and disease resistance, photosensitivity, for example) (Govandraj *et al.*, 2015). Recent advances in molecular markers and genome sequencing provide an excellent opportunity to characterize the genetic diversity of a large germplasm collection (Liu *et al.*, 2015). In the study of plant evolution and comparative genomics, genetic diversity assessment is highly useful in understanding the structure of distinct populations (Baloch *et al.*, 2017).

In addition, to design any conservation strategy, analyzing the genetic diversity using different marker systems is vital for sustainable use, the efficient utilization of plant germplasm for improvement purposes and conservation strategy. Therefore, knowledge about the population structure and genetic relationships of any plant is important for conservation and efficient utilization of these genotypes in future breeding program. The analysis of genetic diversity within and among populations routinely involves the use of different genetic markers. Different genetic markers (such as (i) morphological, (ii) biochemical characterization (allozyme) and (iii) Molecular markers have been used to determine genetic diversity and classify genetic material with great effectiveness. DArT markers and SNP markers are the most frequently used markers for determining genetic diversity in various crops (Baloch *et al.*, 2017). Markers can have the same modes of inheritance as other features, such as dominant/recessive or codominant inheritance. In general, codominant markers are more informative than dominant markers (Govandraj *et al.*, 2015).

3.2. Gene Pyramiding

Gene pyramiding is a technique for combining (stacking) multiple desired genes from different donor parents into a single plant through recombinant DNA technology in a short period of time. In gene pyramiding, genes controlling various traits are introduced into a single cultivar at the same time. Gene pyramiding is a method of improving a widely-grown elite variety's few undesirable features by replacing unsatisfactory genes with better genes. Although standard plant breeding procedures allow for gene pyramiding, phenotypic selection and identification of a single plant carrying multiple genes is extremely challenging. Recombination and the number of meiotic cycles might result in the loss of a gene of interest, complicating plant breeding (Iyer *et al.*, 2007). Functional markers can improve the prospects of gene pyramiding for different desired traits (Chukwu *et al.*, 2019), as demonstrated by the plethora of functional markers associated with a multitude of morphological traits, quality improvements, and resistance to biotic and abiotic stresses for use in many major crops. (Azhaguvel *et al.*, 2006; Chukwu *et al.*, 2019).

3.3. Association mapping

Association mapping is a significant relationship of molecular markers with a phenotypic trait, involves searching for genotype-phenotype correlations. It was used to observe variation in multiplex characters through ancient and developmental process in population. Association mapping technology was generated in plants through transmutation, variation in genes, cross breeding and population selection (Hart and Clark, 1997). The Inherited traits allowing individual discretion of progenitors that permitting descendent for chromosomal mutation and transformation from this new advanced technique.

Association mapping is the statistical covariance between the marker's polymorphism and the trait of interest (Zhang *et al.*, 2016). When compared to linkage mapping, it saves time and gives higher mapping resolution with a higher number of recombination events. Due to the availability of more genetic variations with a broader background, association mapping allows for the identification of a large number of alleles with traditionally measured phenotypic data. Association mapping has also helped scientists to check the variation found in germplasm. The development of the SNP technique made possible the study of the whole genome, the construction of genetic maps and the finding of desirable QTLs in plants (Malik *et al.*, 2014). Li *et al.* (2016) evaluated 5000 inbred lines of maize from 30 joint linkage association mapping populations using 365 SNPs for genome wide association, and these SNPs related to drought resistance traits were observed in 354 candidate genes, of which fifty-two of these genes showed considerable differential expression in the inbred line B73 under water-deficit conditions.

3.4. QTL mapping

The regions in which genes are present in the genome and genes associated with specific quantitative traits are called QTLs (Collard *et al.*, 2005). Genetic factors that are responsible for some of the observed measurable phenotypic variation in a population for a quantitative trait are called quantitative

trait loci (QTLs). Conceptually, it can be a single gene or may be a cluster of interrelated genes for the trait. A QTL, while comparable to a gene, simply refers to an area of the genome that contains one or more functional genes. Among such quantitative traits: yield, plant length, days to flowering, seed size, etc., are the most important ones. Selection for quantitative traits is difficult because the relationship between observed trait values in the field (the phenotype) and the underlying genetic constitution (the genotype) is not straight-forward. QTL mapping is the process of creating a linkage map and doing QTL analysis to discover specific regions associated with traits of interest (Paterson *et al.*, 1996). A quantitative trait locus (QTL) is succeeded from the process of crossing over, which allows analysis of genes and markers in progeny.

Quantitative traits are typically controlled by many genes, but each contributes only a small part of the observed variation. The environmental variations resulting from differences in growing conditions, further create the problem of understanding the relationship between phenotype and genotype. In practice, this problem is typically dealt with by evaluating large and replicated trials, which allow the identification of genotypic differences through statistical analysis. Plant breeders would like to utilize the quantitative traits for genetic factors that are responsible for the observed variability in quantitative traits. In a process known as QTL mapping, the association between reported trait values and the presence/absence of alleles of markers that have been mapped onto a linkage map is evaluated in a process known as QTL mapping. When it is evident that the observed association is not the result of a random process, it is stated that a QTL has been identified.

Identification of molecular makers associated with QTLs involves three basic steps, namely, scoring individuals of a random segregating population for a QTL trait; determination of the molecular genotype of each member of the population; and determination of association between any of the markers and the quantitative trait. The first step is to make a cross and generate marker data. In the next step, they generate linkage maps of molecular markers. Subsequently, collect phenotypic measurements of the QTL trait across the environments in replicated trials. Finally, the mapping of QTL is done. The most common method of determining the association between a marker and QTL is done by analyzing phenotypic observations of traits and scoring of molecular data by one-way analysis of variance and regression analysis. For each marker, the presence of a specific fragment of DNA is considered a marker class, and all individuals (in a segregating population) possessing that marker class are considered to be positive for that class. If the variance attributable to a specific class is significant, the molecular marker that was used to identify that class is known to be linked to a QTL. Regression values are calculated for all the markers that have shown association with the quantitative trait, which reflects the amount of total genetic variation that is explained by the specific molecular marker.

Molecular markers are efficient and accurate methods of breeding technique for introgression of any lines, and they allocate selection in every breeding cycle for the introgression of genes of interest in an accurate method. It has been applied to various crops such as wheat, rice, cotton, oil seeds, and forage species and represents an additional tool in breeding techniques for enhancing yield and growth under biotic and abiotic conditions (Rauf *et al.*, 2016). In rice, MAB has been employed in the mapping of QTLs that control abiotic tolerance such as drought, salinity, and submergence (Al-Ashkar *et al.*, 2019). Prince *et al.* (2015) conducted SSR-based QTL mapping on a RIL population of rice derived from a cross of IR20 and Nootripathu for physiological and yield traits in a drought environment. Furthermore, three QTLs were detected on chromosome 1 (RM8085), chromosome 4 (I12S) and chromosome 6 (RM6836), which can be appropriately utilized to transfer into elite rice lines for stable yield production and growth under drought stress conditions. Merchuk-Ovnat *et al.* (2016) reported three QTLs for yield and biomass in RILs identified from a cross of wild emmer wheat (*Triticum turgidum ssp.*) and durum (*T. turgidum ssp.*) and bread wheat (*T. aestivum*) on chromosomes 1BL, 2BS, and 7AS where wild emmer wheat was used as a source of drought resistance genotype.

3.5. Marker-assisted selection (MAS)

MAS is a molecular breeding strategy that involves phenotypic selection based on a marker's genotype. It aids in avoiding the challenges associated with traditional plant breeding approaches. It has totally changed the standard of selection (Tabor *et al.*, 2002). Plant breeders mostly employ MAS to find acceptable dominant or recessive alleles over generations, as well as to find the most desirable individuals among segregating progeny (Francia *et al.*, 2005). The success rate could be considerably increased by increasing the number of markers connected with QTL (Singh *et al.* 2021). The marker to

be applied should be highly linked to the gene of interest for efficient and accurate marker selection. Plant breeding and traditional genetics have both benefited from the use of molecular markers.

In earlier times, the development of molecular markers, QTL mapping, and fine mapping of precise genes were considered to be difficult and time-consuming processes. However, with the advent of next-generation sequencing, it has made the development of molecular markers easier. The emergence of these molecular markers has further facilitated the development of high-density genetic maps, which enabled the mapping of target genes. Furthermore, they enabled the identification of suitable parents for molecular breeding and were also used to identify desirable offspring during the early stages of development (Jiang, 2013). MAS is an efficient and accurate method of breeding technique for the introgression of any lines, and it allocates selection in every breeding cycle for the introgression of genes of interest in an accurate method. It has been applied to various crops such as wheat, rice, cotton, oil seeds, and forage species and represents an additional tool in breeding techniques for enhancing yield and growth under biotic and abiotic conditions (Rauf *et al.*, 2016).

Under continual recurrent selection, MAS plant breeding approaches for transferring complex features into varieties require a long time. Phenotypic selection and analysis of complex traits may be difficult because of unclear phenotypic selection and analysis due to the vast number of genes that contribute to the traits. With the use of genetic markers, desirable alleles of complex traits are gathered and thus made more efficient. Several recurrent selection cycles are required to accumulate desirable QTL alleles in the breeding population for the enhancement of complex traits (Mullis *et al.*, 1986), which is aided by the use of different molecular markers. As a result, MAS reduces the number of breeding cycles required while improving the precision of complicated trait selection. When parents used in MAS are crossed when informed by markers, an ideal genotype can be obtained after only a few successive generations of backcrossing. Semagn *et al.* (2006) found that functional markers aided genetic gain in the development of hybrid inbred lines and the improvement of complex agronomic traits. Figure 1 shows the most important MAS steps.

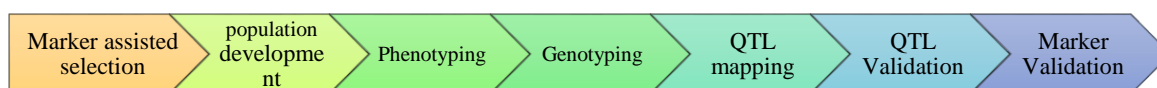


Figure1. Some important steps involves in MAS.

3.6. Genomic selection (GS)

Genomic selection is a more advanced version of marker-assisted selection that was created by Meuwissen *et al.* (2001). It is a strategy that can predict the genetic values of selected candidates based on genome-estimated breeding values (GEBVs), which are anticipated from a high density of markers spread across the genome. GEBV is an estimation model that uses phenotypic data in conjunction with marker and pedigree data to improve prediction accuracy. Grain yield and its components, quality attributes, and abiotic stress resistance are all complex traits that vary with the production of desired phenotypes by selection. By increasing the accuracy of selection, genomic selection of complex traits and high-throughput phenotyping have revolutionized breeding (Ingvarsson and Street, 2011). The main approach of Genomic selection is described in Figure 2.

3.7 Mapping of mutations

Researchers and breeders can now quickly link phenotypic variation to genome sequence differences thanks to advances in genome sequencing technologies. Mapping of a mutation is the first step toward isolating and cloning the corresponding normal gene and identifying its encoded protein. Next-generation sequencing combines single nucleotide polymorphism detection, mutation location, and the possible identification of causative sequence variants to provide genetic mapping (Leshchiner *et al.*, 2012). CandiSNP is a user-friendly tool that will aid in the discovery of new mutations via forward-genetic mutant screens (Etherington *et al.*, 2014).

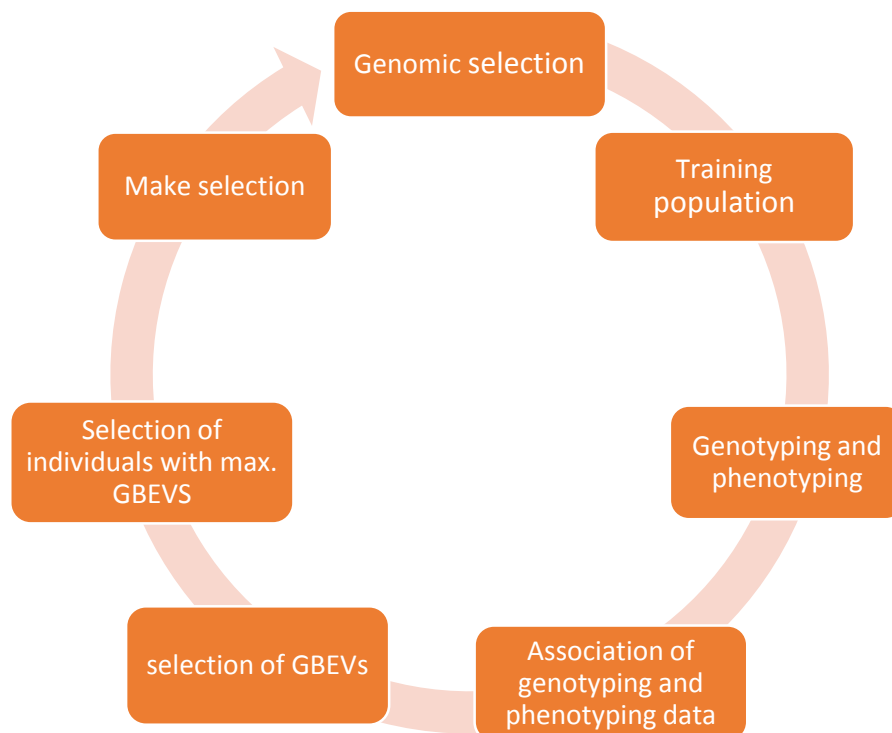


Figure2. General methodology of genomic selection.

4. CONCLUSION

Recently, molecular markers have been widely utilized in crop improvement due to their simplicity, accuracy, reproducibility, and precise location. It is not affected by any environmental factors that reduce its breeding cycle. For many years, they have been continuously used in plant breeding techniques, from RFLP to SNPs, and a diversity of array-technology-based markers. The emergence of sequencing technologies has resulted in the development of NGS platforms that may be low-cost while providing high throughput. The coming years will possibly see continued innovations in molecular marker technology to make it more precise, productive, powerful, and cost-effective in order to discover the underlying biology of various super traits of interest. Recently, many markers have been developed and are available. Of these, SSR, GBS, SNP, and diversity array technology-based markers are mostly used in crop breeding programs. Molecular markers are not only used in plant breeding; they are also used in various areas of plant biology like systematics, population genetics, evolutionary biology, conservation genetics, advances in genomics, and identification of the wild progenitors of domestic species.

ACKNOWLEDGMENTS

The author would like to thank “Jesus” for his kind support for manuscript editing and figure preparation

REFERENCES

- [1] Al-Ashkar I, Alderfasi A, El-Hendawy S, Al-Suhaibani N, El-Kafaf S, Seleiman MF (2019). Detecting salt tolerance in doubled haploid wheat lines. *Agronomy* 9:211
- [2] Azhaguvel P, Saraswathi D, Sharma A, Varshney RK (2006). Methodological advancement in molecular markers to delimit the gene(s) for crop improvement. In: *Advances in Molecular Markers for Crop Improvement*; Azhaguvel P Eds, Global Science Books: Isleworth, UK, pp. 460–469.
- [3] Baloch FS, Alsaleh A, Shahid MQ et al (2017). A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from Central Fertile Crescent. *PLoS One* 12(1):e0167821
- [4] Bayley DC (1983). Isozymic variation and plant breeders' rights. In: Tanksley SD, Orton TJ, editors. *Isozymes in plant genetics and breeding*. Amsterdam: Elsevier p. 425–440.
- [5] Beissinger TM, Hirsch CN, Sekhon RS et al (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics*, 193(4):1073–1081.

- [6] Bhandari HR, Bhanu AN, Srivastava K, Singh MN, Shreya (2017). Assessment of genetic diversity in crop plants. An overview: *Advance Plants Agricultural Research*, 7(3):00255
- [7] Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32:314-331
- [8] Chen PH, Pan YB, Chen RK, Xu LP and Chen YQ (2009). SSR marker based analysis of genetic relatedness among sugarcane cultivars (*Saccharum spp.* hybrids) from breeding programs in China and other countries. *Sugar Technology* 11: 347-354
- [9] Chesnokov YV, Kosolapov VM & Savchenko IV (2020). Morphological Genetic Markers in Plants. *Russian Journal of Genetic,s* 56:1406–1415
- [10] Cho YG, McCouch SR, Kuiper M, Kang MR, Pot J, Groenen JTM and Eun MY (1998). Integrated map of AFLP, SSLP, and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 97: 370-380
- [11] Chukwu SC, Rafii MY, Ramlee SI, Ismail SI, Oladosu Y, Okporie EE, Onyashi G, Utobo E, Ekwu L, Swaray S et al (2019). Marker-assisted selection and gene pyramiding for resistance to bacterial leaf blight disease of rice (*Oryza sativa* L.). *Biotechnology and Biotechnological Equipment*, 33: 440–455.
- [12] Collard BC, Jahufer MZ, Brouwer JB et al (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, 142(1–2):169–196.
- [13] Collard BC, and Mackill DJ (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491):557-572.
- [14] Crossa J, de Los Campos G, Pérez P, Gianola D, Burgueno J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J and Arief V (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186:713–724.
- [15] Chukwu SC, Rafii MY, Ramlee SI, Ismail SI, Hasan MM, Oladosu YA, Magaji UG, Akos I, Olalekan KK (2019). Bacterial leaf blight resistance in rice: A review of conventional breeding to molecular approach. *Molecular Biology Reports*, 46:1519–1532.
- [16] Delourme R, Piel N, Horvais R, Pouilly N, Domin C, Vallee P (2008). Molecular and phenotypic characterization of near isogenic lines at QTL for quantitative resistance to *Leptosphae maculans* in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics*, 117: 1055–1067.
- [17] Dida G, Bantte K and Disasa T (2021). Molecular characterization of Arabica Coffee (*Coffea arabica* L.) germplasms and their contribution to biodiversity in Ethiopia. *Plant Biotechnology Reports*, 15(6):791-804.
- [18] Elshire RJ, Glaubitz JC, Sun Q, et al (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 6(5):e19379.
- [19] Etherington GJ, Monaghan J, Zipfel C, MacLean D (2014). Mapping mutations in plant genomes with the user-friendly web application CandiSNP. *Plant methods*, 10(1): 41.
- [20] Francia E, Tacconi G, Crosatti C et al (2005). Marker assisted selection in crop plants. *Plant Cell Tissue and Organ Culture*, 82 (3):317–342
- [21] Ganal MW, Wieseke R, Luerksen H, Durstewitz G, Graner EM, Plieske J and Polley A (2014). High-throughput SNP profile of genetic resources in crop plants using genotyping arrays, PP 113-130. In: R. Tuberosa, A. Graner and E. Frison (eds). *Genomics of Plant Genetic Resources*. Dordrecht: Springer.
- [22] Govindaraj M, Vetriventhan M and Srinivasan M (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics research international*
- [23] Gupta PK, Varshney RK, Prasad M (2002). *Molecular Markers: Principles and Methodology*. pp. 9-54. In: Jain SM, Brar DS, Ahloowalia BS (eds). *Molecular Techniques in Crop Improvement*. Springer, Dordrecht.
- [24] Hadrys H, Balick M and Schierwater B (1992). Applications of Random Amplified Polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology*, 1: 55-63.
- [25] Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, Yang Z, Zhao L Zhou G Wang Z, Huang L and Zhang Z (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist*, 209 (2):871-884.
- [26] Hartl DL, Clark AG and Clark AG (1997). *Principles of population genetics* (Vol. 116) Sunderland: Sinauer associates.

- [27] Hartl L, Mohler V, Zeller FJ, Hsam SLK and Schweizer G (1999). Identification of AFLP markers closely linked to the powdery mildew resistance genes *Pm1c* and *Pm4a* in common wheat (*Triticum aestivum* L.). *Genome*, 42: 322-329.
- [28] Henry RJ (2001). The DNA Fingerprinting of Plants: Plant Genotyping CAB1 Publisher Wallingford, UK 512p.
- [29] Ingvarsson PK, Street NR (2011). Association genetics of complex traits in plants. *New Phytologist*, 189(4):909–922.
- [30] Iyer AS, McCouch SR (2007). Functional markers for xa5 mediated resistance in rice (*Oryza sativa* L.). *Molecular Breeding*, 19: 291–296.
- [31] Jiang G (2013). 'Molecular Markers and Marker-Assisted Breeding in Plants', in: SB. Andersen (eds.), Plant Breeding from Laboratories to Fields, IntechOpen, London 10.5772/52583
- [32] Kesawat MS and Das BK (2009). Molecular markers: It's application in crop improvement. *Journal of Crop Science and Biotechnology*, 12: 169-181.
- [33] Leshchiner I, Alexa K, Kelsey P, Adzhubei I, Austin-Tse CA et al (2012). Mutation mapping and identification by whole-genome sequencing. *Genome Research*, 22: 1541-1548.
- [34] Li C, Sun B, Li Y, Liu C, Wu X, Zhang D, Shi Y, Son Y, Buckler ES, Zhang Z and Wang T (2016). Numerous genetic loci identified for drought tolerance in the maize nested association mapping populations. *BMC genomics*, 17(1):1-11.
- [35] Liu W, Shahid MQ, Bai L et al (2015). Evaluation of genetic diversity and development of a core collection of wild rice (*Oryza rufipogon* Griff.) populations in China. *PLoS One*, 10 (12):e0145990.
- [36] Madhumati B (2014). Potential and application of molecular markers techniques for plant genome analysis. *International Journal of Pure and Applied Bioscience* 2(1):169–188.
- [37] Malik W, Ashraf J, Iqbal MZ, Ali Khan A, Qayyum A, Ali Abid M, Noor E, Qadir Ahmad M and Hasan Abbasi G (2014). Molecular markers and cotton genetic improvement: current status and future prospects. *The Scientific World Journal*
- [38] Merchuk-Ovnat L, Barak V, Fahima T, Ordon F, Lidzbarsky GA, Krugman T and Saranga Y (2016). Ancestral QTL alleles from wild emmer wheat improve drought resistance and productivity in modern wheat cultivars. *Frontiers in plant science* 7:452
- [39] Meuwissen TH, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819–1829.
- [40] Mondini L, Noorani A and Pagnotta M (2009). Assessing plant genetic diversity by molecular tools. *Diversity* 1(1):19-35.
- [41] Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harb. Symp. Quantum Biology* 512:63–273.
- [42] Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N and Özkan H (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment* 32(2):261-285.
- [43] Nawaz MA, Sadia B, Awan FS et al (2013). Genetic diversity in hyper glucose oxidase producing *Aspergillus niger* UAF mutants by using molecular markers. *International Journal of Agriculture and Biology* 15(2):362–366.
- [44] Orozco-Castillo C, Chalmers KJ, Waugh R and Powell W (1994). Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics* 87(8):934-940.
- [45] Paglia G and Morgante M (1998). PCR-based multiplex DNA fingerprinting techniques for the analysis of conifer genomes. *Molecular Breeding* 4:173-177.
- [46] Pan YB, Burner DM, Ehrlich KC, Grisham MP and Wei Q (1997). Analysis of primer-derived non-specific amplification products in RAPD-PCR. *Bio techniques*, 22: 1071-1077.
- [47] Parker PG, Snow AA, Schug MD, Booton GC and Fuerst PA (1998). What molecules can tell us about populations: choosing and using a molecular marker. *Ecology*, 79: 361-382.
- [48] Peterson GW, Dong Y, Horbach C, Fu YB (2014). Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity* 6: 665–680.

- [49] Phougat D, Panwar I, Puina, M, Sethi S (2018). Microsatellite markers based characterization in advance breeding lines and cultivars of bread wheat. *Journal of Environmental Biology* 39:339–346.
- [50] Poland JA, Rife TW (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5(3):92–102.
- [51] Prince SJ, Beena R, Michael GS, Senthivel S and Chandra BR (2015). Mapping consistent Rice (*Oryza sativa* L.) yield QTLs under drought stress in target rain fed environments. *Rice* 8:25.
- [52] Rauf S, Al-Khayri JM, Zaharieva M, Monneveux P, Khalil F (2016). Breeding Strategies to Enhance Drought Tolerance in Crops. In: Al-Khayri J, Jain S, Johnson D (eds) *Advances in Plant Breeding Strategies: Agronomic, Abiotic and Biotic Stress Traits*. Springer, Cham.
- [53] Roder MS, Korzun V, Wendehake K et al (1998). A microsatellite map of wheat. *Genetics* 149(4):2007–2023.
- [54] Saal B and Wricke G (2002). Clustering of amplified fragment length polymorphism markers in a linkage map of rye. *Plant Breeding* 121(2):117-123.
- [55] Salgotra RK and Stewart CN (2020). Functional markers for precision plant breeding. *International journal of molecular sciences* 21(13):4792.
- [56] Semagn K, Bjornstad A, Ndjiondjop MN (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology* 5(25): 2540-2568.
- [57] Singh M, Nara U, Kumar A et al (2021). Salinity tolerance mechanisms and their breeding implications. *Journal of Genetic Engineering and Biotechnology* 19: 173
- [58] Singh RK, Mishra SK, Singh SP, Mishra N and Sharma ML (2010). Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids. *Australian Journal of Crop Science* 4:116-125.
- [59] Tabor HK, Risch NJ, Myers RM (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Review Genetics*. 3(5):391–397.
- [60] Wenzl P, Carling J, Kudrna D et al (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proceeding National Academy of Science USA* 101(26):9915–9920.
- [61] Williams JG, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research* 18(22):6531-6535.
- [62] Wurschum T, Langer SM, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E and Reif JC (2013). Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theoretical and Applied Genetics* 126(6):1477-1486.
- [63] Xu Y (2010). *Molecular plant breeding*. CAB International.
- [64] Zhang P, Zhong K, Shahid MQ et al (2016). Association analysis in rice: from application to utilization. *Frontier in Plant Science* 7:1202.

Citation: Gudeta Dida, (2022). “Molecular Markers in Breeding of Crops: Recent Progress and Advancements” *International Journal of Research Studies in Agricultural Sciences (IJRSAS)*, 8(5), pp.1-12
DOI: <http://dx.doi.org/10.20431/2454-6224.0805001>

Copyright: © 2022 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.