Plant Archives Vol. 25, Supplement 1, 2025 pp. 2062-2074 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2025.v25.supplement-1.284

EMERGENCE OF MOLECULAR MARKER FROM RFLP TO SNP: A REVIEW

Rajdeep Jajoriya[1]*, Ritu Sharma[1], Vijay Kumar Meena[1], Amool Singh Vadithya[2], Khushwant Singh[3], Prakash Kumar[4], Saurabh Meemrot[5] and Tejendra Singh Nathawat[6]

[1]Department of Genetics and Plant Breeding, *Rajasthan Agricultural Research Institute,* (SKNAU, Jobner) Durgapura, Jaipur, Rajasthan, 302018, India,
[2]Department of Genetics and Plant Breeding, Punjab Agricultural University, Ludhiana, Punjab, India,
[3]Department of Soil Science and Agricultural Chemistry, University of Agricultural Sciences, Bangalore, India,
[4]Department of Entomology, Swami Keshwanand Rajasthan Agricultural University, Bikaner, Rajasthan, India
[5]Department of Agriculture, Invertis University, Bareilly,U.P.,India
[6]Department of Genetics and Plant Breeding, Institute of Agriculture, Visva-Bharati University, Sriniketan, West Bengal, India
*Corresponding author Email: rajdeepjajoriya@gmail.com
(Date of Receiving : 02-10-2024; Date of Acceptance : 30-11-2024)

ABSTRACT

Genetic markers are polymorphic traits that serve as signs or flags to indicate the genotype of individuals or species. An ideal genetic marker should be polymorphic, codominant, abundant, evenly distributed, neutral, and not influenced by the environment. Genetic markers are broadly classified into morphological, protein-based, and DNA markers. Morphological markers, such as flower color and plant height, are easily observable but limited in number and influenced by the environment. Protein-based markers, including isozymes and allozymes, reflect changes in gene sequences but are tissue-specific and affected by developmental stages. DNA markers, which represent variations in the genomic DNA sequence, are abundant, non-tissue-specific, and not influenced by the environment or developmental stage. The development of PCR revolutionized DNA marker systems, leading to the creation of various markers such as RAPD, SCAR, AFLP, STS, SSRs, ISSRs, CAPS, SSCP, TGGE, SRAP, TRAP, and SCoT. Among these, SSRs or microsatellites have gained popularity due to their high polymorphism, reproducibility, and codominance. Single nucleotide polymorphisms (SNPs) are the most recent and abundant genetic markers, with various types and discovery methods. SNP genotyping can be performed using techniques like molecular beacons, microarrays, primer extension, oligonucleotide ligation, and genotyping by sequencing. InDels, RNA-based markers, and epigenetic markers have also emerged as valuable tools for genetic studies. The choice of genetic marker depends on the specific application, such as linkage mapping, marker-assisted selection, or genetic diversity analysis, and the availability of resources and technology

Keywords **:** Molecular markers, QTL.

Introduction

Gregor Johann Mendel selected seven different qualitative traits having two easily identifiable contrasting forms for his experiments on plant hybridization which laid the foundation of genetics. Subsequently, many qualitative traits were used in genetic studies in a variety of organisms, however, it was soon recognized that many traits of economic importance exhibit continuous variation. These traits were referred to as quantitative traits and their phenotype was largely influenced by the environment. These traits were difficult to score or classify into distinct phenotypic classes, and a complete correspondence between phenotype and genotype couldn't be established. This necessitated the use of markers to study the genotype of an organism or species. Different types of markers were used in genetics and plant breeding to make selections, associate phenotypes with genotypes and map the

genes. Starting with the morphological markers which could be seen with the naked eye (E.g., shape and color of flowers). In common bean a linkage between seed coat color and seed size was reported by Sax in 1923. Subsequently QTLs for several quantitative traits in drosophila were mapped by Thoday 1961 using cytogenetic techniques and genetic analyses.

These were followed by isozyme-based markers which were reliable in genetic studies because of their consistency in expression irrespective of the environment. The isozyme-based markers were soon replaced by DNA based markers as these were highly tissue specific and expression was influenced by the developmental stages or interallelic interactions. The last few decades have witnessed a growing trend in the use of molecular markers to reveal polymorphism at the DNA level. A genetic marker is defined as any trait that is polymorphic, easily and reliably identified, readily followable in segregating generations and indicating the genotype of the individuals. These represent genetic differences between individual organisms or species and act as signs or flags (Collard *et al* 2005).

Characteristics of an ideal genetic marker:

- Polymorphic and multiallelic- this permits classification into more than 2 groups
- Codominant- allows differentiation of heterozygotes from the homozygotes
- Not epistatic- this prevents the interference of identified marker alleles at one locus with that of other marker loci
- Neutral- most of the molecular markers are developed from the non-coding regions of the genome and hence do not bring any effect of their own
- Abundant and evenly distributed over the entire genome
- Non pleiotropic
- Not influenced by environmental variations

Classification of genetic markers: Broadly genetic markers are classified into 3 types

i. Morphological/ visible markers
ii. Protein based markers
iii. DNA markers (the term molecular markers in the present context are used only for DNA based markers)

Visible/ Morphological markers

These represent one of the earliest genetic marker systems and includes shape and color of flowers, shape of fruits, height of the plant, seed color and shape, foliage type etc. The scoring for these markers is easy, inexpensive and rapid and does not involve any sophisticated techniques or equipment. These were used in the construction of linkage maps by studying crosses between parents differing for two or three traits. The gene order and distance between the genes could be found if one or more loci were common between crosses. HD2329 variety of wheat is identified by the presence of brown glumes and crooked neck is a characteristic feature of Kalyan Sona wheat. Their limited number in a species, scorability only during specific developmental stages, exhibition of pleiotropic effects, threshold requirements for their expression inhibit its usage.

Protein based markers

These are electrophoretic variants of proteins and enzymes. Small changes in the coding sequence of genes can give rise to an altered amino acid sequence resulting in variant protein molecules. These variants differ from the wild type molecules in electrical charge and hence will show different mobility.

Isozymes are different forms of an enzyme having the same function but coded by different genes.

Allozymes are variants of an enzyme coded by different alleles of the same gene.

This marker system had an upper hand over morphological markers by virtue of their reflection of changes in the gene sequence, detection using only a small amount of tissue, easy analysis and codominant nature of isozymes, however a few drawbacks also accompany this marker system which includes; only a small number of markers may be polymorphic in the two parents, changes producing functional enzymes and varied electrophoretic mobility can only be detected, different isozymes having same mobility will be observed as a single band and are influenced by the environment and developmental stages. Isozyme markers have been used for Marker Assisted Selection (MAS) eg, Asp1(acid phosphatase) is linked to nematode resistance in tomato. These markers although very informative and more advanced than morphological markers did possess several limitations and have been replaced by DNA based markers.

DNA markers

This marker represents variations in the genomic DNA sequence. DNA markers are abundant, non-tissue specific, not influenced by the developmental stage or environment, and many of them are codominant (*de* Vienne *et al* 2003). An analysis of the fragments generated by restriction digestion of adenovirus DNA for physical mapping in 1974 led to the development of these markers (RFLP). RFLP was the first DNA based

marker. Eventually a number of different DNA marker systems were developed.

Restriction fragment length polymorphism (RFLP)

It was the first generation of molecular markers and represents the variation in length of the fragments produced by a single restriction enzyme at the same genomic region in different individuals. Grodzicker *et al.* in 1974 used these to determine the location of temperature sensitive mutations in adenovirus on a physical map of restriction fragments. The principle and procedure of RFLP as a molecular marker was given by Botstein *et al.* in 1980 during the construction of human linkage map (Figure 1).The procedure involves isolation of high molecular weight DNA from the sample individuals and their digestion using selected restriction enzymes. The fragments generated are separated on weight basis (gel electrophoresis), denatured and transferred onto a nitrocellulose membrane (Southern Blotting). The fragments are fixed and hybridized with labelled DNA probes (Southern Hybridization). Unbound probes are washed away and bands are visualized using autoradiography.

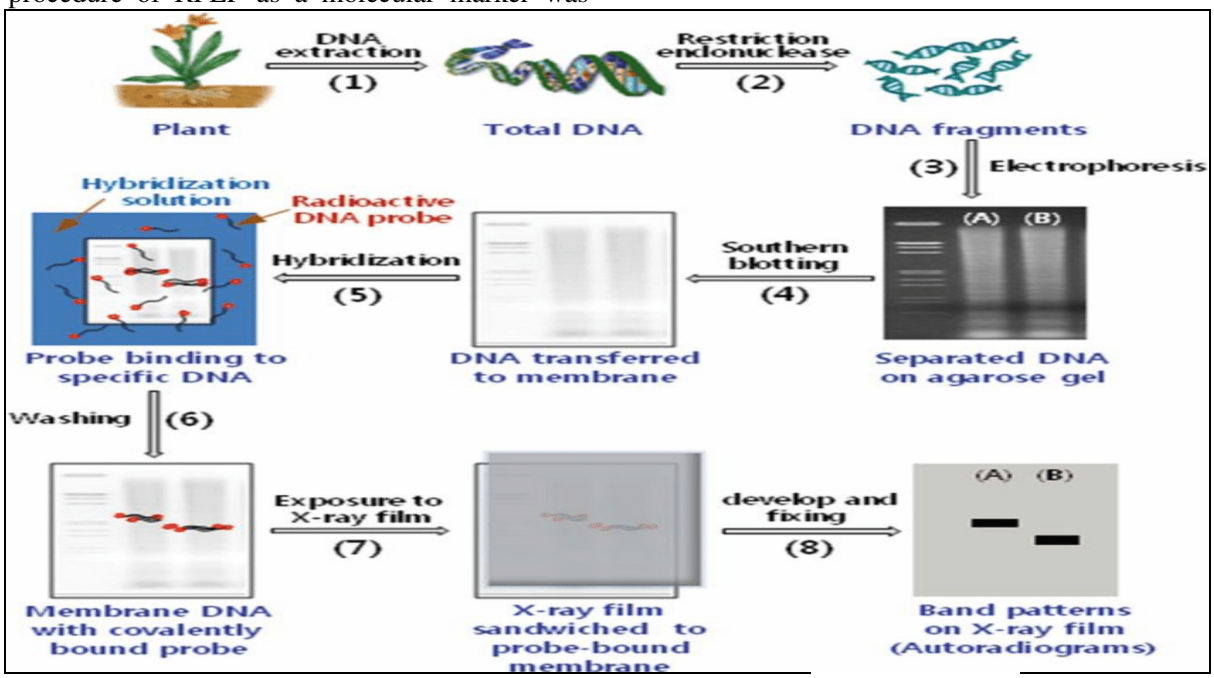

Fig. 1: Schematic representation of RFLP procedure (Tazeb *et al,* 2018)

These were well accepted and widely used markers in the 1980s and 1990s and have several advantages like a number of RFLP loci could be scored and mapped in a mapping population, were highly reproducible, codominant and could be used for the construction of RFLP maps. These markers are no more in common use as it is expensive, labor and time consuming, uses radioactive probes, and its non amenability to automation and high throughput.

Diversity Array Technology (DArT)

It is a high throughput genotyping system which is similar to amplified fragment length polymorphism (AFLP) using microarray based nucleic acid hybridization for detection of polymorphism. It involves two steps construction of a microarray (genetic diversity array or genotyping array) and hybridization of test individuals with the diversity array (genotyping). Construction of diversity array involves bulking of DNA from a group of individuals representing the diversity to be studied followed by digestion with 1-2 restriction enzymes. The fragments generated are ligated with adapters and PCR amplified with primers having 1-3 selection nucleotides. The PCR products are cloned and their DNA inserts are spotted on a microarray (discovery array). Polymorphic fragments in the discovery array are identified by hybridization with labelled genomic DNA fragments of all the individuals used to construct the array. Inserts from those fragments which are polymorphic are now spotted as a microarray and this is called diversity array. Genotyping involves PCR amplification of the individual sample DNA, labelling with flourescence and hybridisation with diversity array. Polymorphism is scored using softwares like DArTsoft, DArTdb etc.

Variable Number of Tandem Repeats (VNTRs)

These are composed of variable number of tandemly repeated sequences (2-60)bp which are distributed throughout the genome. These are classified as minisatellites and microsatellites. Minisatellites are 0.2-2kb long and are made of 11-60bp repeats,

microsatellites are less than 100bp in length and consists of 2-7bp tandem repeats (**Figure 2**). The proterminal region of chromosomes represent high concentration of minisatellite DNAs in humans. Polymorphism in these markers is detected using Southern hybridization with VNTR sequence as the probe. A major application of this is in DNA fingerprinting (Jeffreys *et al* 1985).

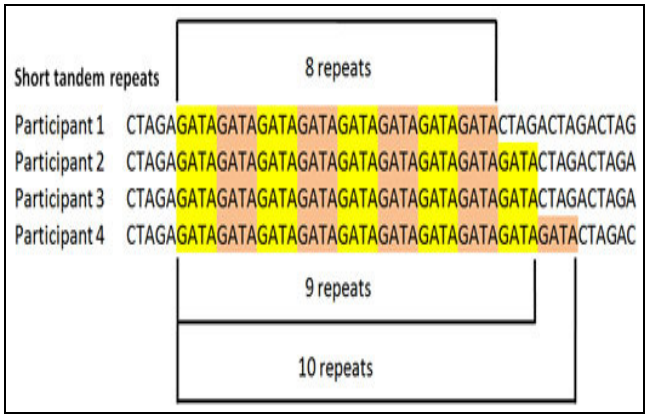

Fig. 2 : Represents the variation in the number of tandem repeats of GATA. (https://www.stewartsociety.org/images/bannockburn-tandem-repeats.jpg)

It was in 1986 that Kary Mullis developed the idea of PCR. This technique could produce microgram quantities of DNA (billions of copies) from a single copy of target DNA or RNA segment in a matter of few hours. A PCR reaction involves denaturing of the template DNA, followed by annealing with primers at the 3́ end and replication of the strands in 5́→3́ by Taq DNA polymerase to produce a double stranded DNA. These steps are repeated for 30-40 times to produce millions of copies of target DNA. Discovery of PCR (Polymerase Chain Reaction) brought revolutionary changes in the way DNA markers are viewed and utilized. A number of PCR based marker systems were developed based on this revolutionary technique starting with RAPD (Randomly Amplified Polymorphic DNA).

RAPD (Randomly Amplified Polymorphic DNA)

It was developed independently by Williams *et al* and Welsch *et al* in 1990. This marker system involves amplification of target DNA using arbitrary primers which serve as both forward and reverse primers and are about 10nt long. Amplification occurs only if the primer anneals at two sites within 2kb of each other on opposite strands such that their 3́ ends face each other.

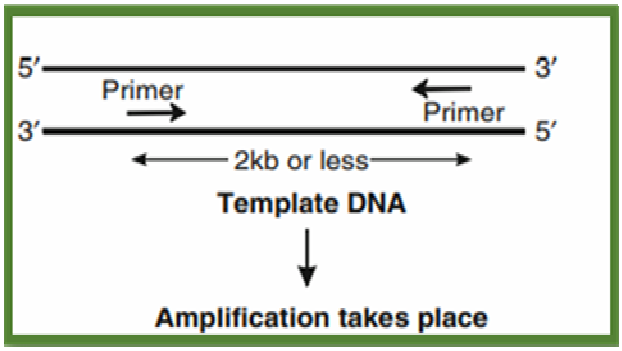

Fig. 3 : Binding of primers to the two strands of DNA for amplification (Singh and Singh, 2015)

The general procedure of this marker system involves isolation of genomic DNA followed by PCR amplification using random primers and detection of polymorphism based on the size of fragments generated using gel electrophoresis (Figure 3). This system does not require prior sequence information, does not involve radioactive probes, are dominant in nature and is used for high density linkage map construction but, the reproducibility of this marker system is very low. There are two variants of RAPD namely, DNA amplification fingerprinting (DAF) and Arbitrary Primed PCR (AP-PCR) differing in only the primer length, annealing conditions and the type of gel used.

DAF amplifies sequences using a short oligonucleotide primer of 5-8bp and visualizes the fragments using polyacrylamide gel electrophoresis (PAGE) along with silver staining. The primer extension conditions are less stringent and these find applications in DNA fingerprinting.

AP-PCR was developed by Welsh and McClelland (1990). Arbitrary primers of 18-32bp are used for amplification and separation of fragments is by PAGE and involves the use of autoradiography making it unpopular.

Sequence Characterized Amplified Regions (SCAR)

This marker system is an extension of RAPD wherein the amplified fragment representing the RAPD marker is eluted, cloned and end sequenced. Forward and reverse primers (20-24nt) are designed based on the terminal sequences and these amplify a single fragment depicting the RAPD marker. They are tested for their ability to detect polymorphism and successful primers form SCAR markers. These are dominant markers and are used in physical and genetic mapping, along with phylogenetic studies (Figure 4).

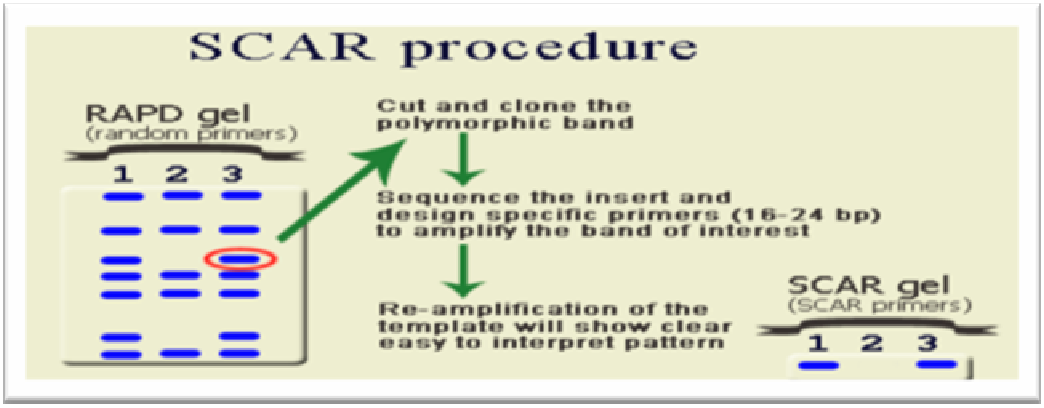

Fig. 4: Procedure for SCAR marker development.
(https://dbgap.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechSTS.shtml)

Amplified Fragment length Polymorphism (AFLP)

This marker technology was developed by Zabeau and Vos in 1993 and represents a combination of RFLP and RAPD markers. It involves digestion of genomic DNA with two restriction enzymes, one a rare cutter and other a frequent cutter giving rise to three types of fragments; one with both ends generated by rare cutter, second with both ends generated by frequent cutter, and third type where one end is generated by rare and the other end by the frequent cutter. Specific adapter molecules are ligated to the cut ends and primers are designed using the adapter sequence. Amplification of the fragments is done using primers with 1-3 selection nucleotides and visualized by denaturing PAGE involving autoradiography. This is a dominant marker system which is highly reproducible and eliminates the need for prior sequence information. It has been used for high resolution mapping, marker assisted selection, line identification and gene cloning although being technically demanding and expensive.

Sequence Tagged Sites (STS)

An unambiguously defined locus in terms of the flanking primer sequence used for its amplification is called an STS (Olsen *et al* 1989). There are different methods for the creation of an STS marker including from the end sequences of a RAPD fragment (SCAR marker), end sequences of AFLP/RFLP fragments, from the unique sequences flanking mini and microsatellites (SSR), etc.

Microsatellites/ Simple Sequence Repeats (SSRs)

Litt and luty in 1989 described Microsatellites/ Short Tandem Repeats/ Simple Sequence Repeats/ Simple Sequence Length Polymorphism/ Sequence Tagged Microsatellite Sites as tandemly repeated sequences of 1-6bp which showed codominant inheritance while working on the human cardiac actin gene. These are special versions of STS markers in which the microsatellite locus is amplified using specific primers designed from the unique sequences flanking the locus. Number of SSRs is variable in different individuals and this variation is detected as bands of varying sizes on gel electrophoresis (Figure 5). The microsatellites or SSRs (Simple Sequence Repeats) or STMS (Sequence Tagged Microsatellites) are one of the most popular and widely used marker systems. These markers are highly polymorphic, reproducible, codominant.

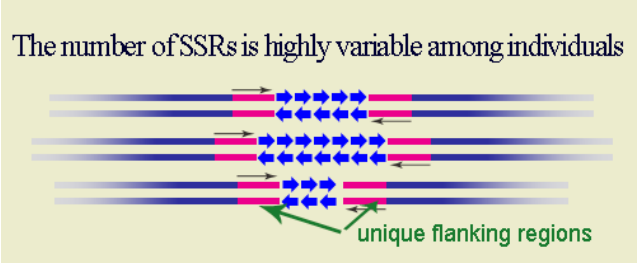

Fig. 5 : Unique sequences flanking the locus by SSR makers
(https://www.ncbi.nlm.nih.gov/probe/docs/techsts)

with medium throughput and amenability to automation leading to its popularity. Although very popular, these are not the best marker systems as they are located in non-coding regions, and may have the presence of null alleles and stutter bands.

During this period other marker systems were also developed like ISSR (Inter SSR), CAPS (Cleaved amplified Polymorphic Sequence), SSCP (Single Strand Conformation Polymorphism), TGGE (Temperature Gradient Gel Electrophoresis), SRAP (Sequence Related Amplified Polymorphism), TRAP (target Region Amplification Polymorphism), SCoT (Start Codon Targeted Polymorphism) and are in use although on a limited basis.

Inter Simple Sequence Repeats (ISSR)

Inter SSR represents sequences of 100-3000bp located between two adjacent oppositely oriented microsatellite regions. These are amplified using single primer having the sequence of the SSR locus. The primers can either be anchored or non-anchored. Anchored primers are those in which 1-3 selection nucleotides have been added either on the 3` or 5` end of the SSR locus sequence used as primer. This increases the specificity. Non anchored primers represent only the SSR or microsatellite sequence. The markers generated from non-anchored primers are referred to as SPARs (Single primer amplification reactions) or MP-PCR (Microsatellite Primed PCR) and those from anchored primers are called inter SSR PCR or ASSRs (Anchored Simple Sequence Repeats) or AMP-PCR (Anchored Microsatellite Primed PCR). This marker system has been used for studying genetic identity, strain and clone identification, taxonomic studies etc. The primer development is based on the sequence of the microsatellite locus and hence prior sequence information is not required. These are dominant markers with high throughput but poor producibility.

Cleaved Amplified Polymorphic Sequence (CAPS)

These are also referred to as PCR-RFLPs and detect the length polymorphism generated by restriction digestion of amplified PCR products of different genotypes. These are codominant markers and detect the variations in the restriction site of the enzymes (SNP) which are located within the amplified PCR product. In cases where the amplified PCR products are large, they may fail to reveal polymorphism in the genotypes and in such situations changes in the restriction sites of enzymes can be revealed by digestion with the respective restriction enzymes. The digested amplicons are then visualized by gel electrophoresis and find applications in gene mapping studies. The use of enzymes in this marker system makes it expensive and unsuitable for automation and high throughput.

Single Strand Conformation Polymorphism (SSCP)

This marker system detects variations in the movement of single stranded DNA molecules in different individuals by virtue of their conformational differences (Orita *et al* 1989). The genomic DNA is PCR amplified using specific primers and the last cycle of PCR is followed by denaturation to produce single stranded DNA, this is followed by quenching wherein the reaction is suddenly brought to a low temperature. Quenching results in folding of the single stranded DNA molecules upon themselves to form secondary structures. Any changes in the DNA sequence (SNP) results in changes in the secondary structure conformation which in turn results in differential mobility of the molecules upon electrophoresis. The amount of sequence differences detected by SSCP decreases as the length of the DNA duplex increases. This technique can be used for rapid screening of sequence differences when the precise difference is not needed. It has been used for genetic studies to a limited extent mainly because of the labor intensiveness and lack of automation facilities with this system.

Denaturing/ Temperature Gradient Gel electrophoresis (D/TGGE)

This marker system detects variations in the DNA molecule based on the movement of double stranded DNA in different individuals when denaturing conditions are provided. Regions within a DNA duplex vary in their melting temperatures which depends on their base composition. A-T rich regions have low melting temperatures while the G-C rich regions have higher melting temperatures. The genomic DNA of the individuals to be tested are amplified and loaded onto acrylamide gel provided with denaturing conditions (temperature gradient or chemicals like urea etc.). Initially the DNA migrates as double stranded molecule but further down the gel when they encounter denaturing conditions, unstable regions begin to melt and separate. A point is reached when DNA becomes single stranded producing a branched structure (fork) and there is no further migration in the gel. This point varies with the composition of DNA and hence variant molecules form different bands in the gel. This technique is able to distinguish heterozygotes from homozygotes.

Sequence Related Amplified Polymorphism (SRAP)

Sequence-related amplified polymorphism (Riaz *et al.,* 2001) was originally developed for gene tagging in *Brassica oleracea* L. with ambiguous primers targeting GC-rich exons (forward primers) and AT-rich promoters, introns, and spacers (reverse primers). The primers have a core sequence of 13 to 14 bases long, starting at the 5′ end, the first 10 or 11 bases are "filler" sequences, followed by the CCGG- (forward) or -AATT sequences (reverse). Three selective nucleotides (random) are located at the 3`-end taking the total primer length to 17-18 nucleotides long. Presence/absence of bands using electrophoresis and gel visualization is the most prevalent method for scoring the fragments making these marker systems dominant in nature. SRAPs are capable of elucidating genetic diversity at a number of taxonomic levels, but they are most commonly used to study populations of

interspecific and intraspecific hybrids. The creation of linkage maps and the identification of quantitative trait loci have both been done from SRAP data. As a result, this technique has proven to be beneficial in improving agronomic crops. Many investigations have indicated that SRAP markers and AFLP markers exhibit comparable levels of variation.

Target Region Amplification Polymorphism (TRAP)

The TRAP marker is a reasonably high-throughput marker system that combines Expressed Sequence Tags (EST) data with a bioinformatics approach to build polymorphic markers around target candidate gene sequences (Hu and Vick 2003). The target region amplification polymorphism (TRAP) approach generates markers using two 18-nucleotide primers. The fixed primer is created from the database's targeted EST sequence, while the arbitrary primer is made from an arbitrary sequence having either an AT- or GC-rich core to anneal with an intron or exon, respectively. When separated on a 6.5 percent polyacrylamide sequencing gel, each PCR reaction can create up to 50 scorable fragments with sizes varying from 50 to 900 bp for different plant species. The TRAP method might be useful for genotyping germplasm collections and labelling genes that control desirable agricultural plant agronomic features. These techniques use no prior sequence information, and the markers generated are randomly distributed across the genome.

Start Codon Targeted (SCoT) markers

Start Codon Targeted (SCoT) polymorphisms are dominant and reproducible markers based on the short-conserved regions in plant genes surrounding the ATG translation start (initiation) codon (Joshi *et al.,* 1997 and Sawant *et al.,* 1999). These use a single 18-mer primer in polymerase chain reaction (PCR) assays and a higher annealing temperature (50 °C) than other markers. Conventional gel electrophoresis with agarose gels and staining are used to visualize the markers, making this technique acceptable for the great majority of plant research labs with standard equipment. SCoT markers have lower amounts of recombination with the gene/trait than random markers like RAPDs, ISSRs, or SSRs, allowing them to be employed directly in marker-assisted breeding programs. SCoT markers have been effectively employed in a variety of species, including rice, sugarcane, grape, potato, and others, to assess genetic diversity and structure, identify cultivars, and map quantitative trait loci (QTL) and DNA fingerprinting.

CAAT box derived polymorphism (CBDP)

CAAT box-derived polymorphism (CBDP), a new promoter-targeted marker system is based on the nucleotide sequence of plant promoter's CAAT box. CAAT box is a nucleotide pattern with a consensus sequence GGCCAATCT that is positioned 80bp upstream of the eukaryote gene's start codon and plays a key role during transcription (Benoist *et al.,* 1980). The core CAAT nucleotides of the CAAT box are known to be conserved, and C is the most common nucleotide found preceding CAAT nucleotides (Shahmuradov *et al.,* 2003). Primers targeting the CAAT box were developed and utilized in PCR to generate markers. CBDP, like random amplified polymorphic DNA (RAPD), generates markers using a single primer in a polymerase chain reaction (PCR) however, CBDP primers, unlike RAPD primers, are 18 nucleotides long and contain a central CCAAT nucleotides core flanked by a filler sequence at the 5′ end and di- or trinucleotides at the 3′ end. CBDP, has two major advantages over RAPD: the markers are derived from variants found in the gene-rich region of plant genomes, and they are more reproducible.

CDDP (Conserved DNA Derived Polymorphism) markers

CDDP markers are comparable to resistance gene analogue markers, which were created from conserved regions in plant disease resistance genes (Chen *et al* 1998). It uses a single primer and conventional agarose gel electrophoresis to exploit conserved DNA sequences across plant species. These are dominant markers. The development of this marker system involves the selection of genes which are well characterized, the sequences of these genes in diverse plant species are obtained from the database and are subjected to multiple sequence alignment analysis. The conserved sequences are used for primer design which leads to the development of reproducibly polymorphic markers.

This marker technique provides an additional option to RAPD or ISSR techniques, as well as other recently developed techniques such as target region amplified polymorphism, SCoT polymorphism, and CoRAP (Wang *et al* 2009) for QTL mapping, bulked segregant analysis, and genetic diversity applications in plant genetic analysis. CDDP may be particularly effective for targeted QTL mapping.

Intron targeting polymorphism (ITP)

The non-coding regions of genes are known as introns and these offer an appealing source of polymorphism for marker development because of their slow sequence evolution. Intron sequences have

more insertions/deletions and base substitutions than exonic sequences, hence, introns are far more variable than exons, making them ideal for the development of DNA markers (Hawkin, 1988 and Kimura, 1983). Intron length polymorphism has been shown to be a reliable source of informative markers with good interspecific transferability (Poczai *et al.,* 2013). The primers are designed based on the conserved sequences flanking at least two intron/exon junctions. These are codominant markers whose development does not require prior sequence information (**Figure 6).**

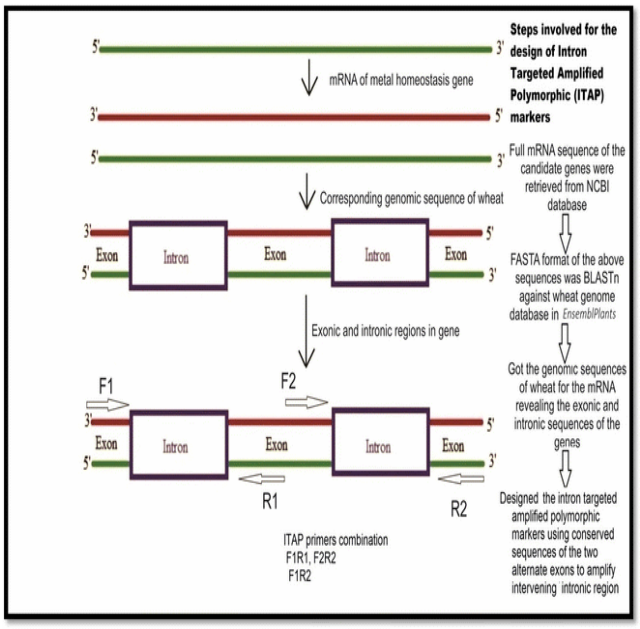

Fig. 6 : Intron targeting polymorphism (Sheikh *et al*, 2018)

Single Nucleotide Polymorphism (SNP)

SNP represents variations at a single nucleotide i.e, a single base change among different individuals. These are commonly biallelic markers arising due to transitions or transversions. Transitions are replacement of a purine for another purine or a pyrimidine for another pyrimidine and transversions represent conversion of a purine to a pyrimidine and vice versa. SNPs are the last-generation of genetic markers that are found in both animal and plant genomes at high frequencies. These are the most common type of DNA variations and therefore the markers based on SNPs have higher densities than other marker systems. The current state of next-generation sequencing (NGS) technology allows for the creation of enormous amounts of sequence data, enabling the identification of high-throughput SNP markers in a species. The development of SNP markers has led to a tenfold increase in genotyping analysis.

Types of SNPs

- Non coding SNPs (ncSNP): SNPs located in non-coding regions or introns.
- Coding SNPs (cSNP): SNPs located in coding regions or exons.
- Synonymous SNPs (synSNP): Exonic SNP that does not lead to a change in the amino acid or polypeptide sequence.
- Non synonymous SNPs (nsSNP): Exonic SNP that alters the amino acid sequence of the protein.
- Functional SNP: Genic SNPs that affect the function of concerned genes
- Reference SNPs (refSNP): SNPs that serve as reference point for defining neighboring SNPs
- In Silico (isSNP): SNPs discovered by mining ESTs or genomic database

Discovery of SNPs

There are a number of methods for SNP discovery. These include:

1. Whole genome resequencing
2. Amplicon resequencing
3. Allele/ SNP mining

Methods of SNP genotyping are categorized as:

Based on nucleic acid hybridization

Molecular Beacons

A specially constructed single-stranded oligonucleotide probe is used to detect SNPs. The oligonucleotide is constructed with complimentary regions at both ends and a probe sequence in the middle. The beacon takes on a hairpin, or stem-loop, configuration in its natural state. A fluorophore is attached to one end of the probe, while a fluorescence quencher is attached to the other. In the unbound natural-state the fluorophore is in close proximity to the quencher due to the stem-loop configuration of the probe, preventing the molecule from generating any fluorescence. The probe sequence is complementary to the genomic DNA under study (Abravaya *et al.,* 2003). If the probe sequence of the molecular beacon encounters its target genomic DNA, it will anneal and hybridize. The hairpin segment of the probe will be denatured in favor of forming a longer, more stable probe-target hybrid. This conformational change permits the fluorophore and quencher to be free of their tight proximity due to the hairpin association, allowing the molecule to fluoresce which can be detected to identify the specific SNP at the locus. If there is any mismatch at the SNP locus the probe will not hybridize with the target genome and hence the hairpin configuration of the beacon is maintained resulting in no fluorescence (Figure 7).

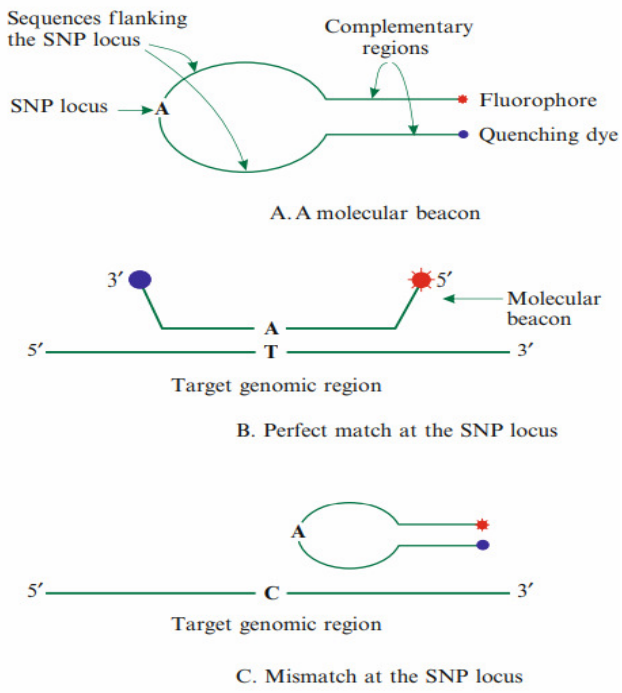

Fig. 7 : SNP genotyping using Molecular Beacons (Singh and Singh, 2015)

Microarray/ SNP chips

Principle: Complementary sequences of DNA can be used to hybridize, immobilized DNA molecules. The three mandatory components of the SNP arrays are:

i. An array containing immobilized allele-specific oligonucleotide (ASO) probes.
ii. Fragmented nucleic acid sequences of target, labelled with fluorescent dyes.
iii. A detection system that records and interprets the hybridization signal.

The silicon chip consists of oligonucleotide probes (ssDNA) attached and their exact positions are known. These probes are allele specific. The sample DNA is PCR amplified and labelled with fluorescent dyes which is then washed over the chip. The specific DNA sample will bind to the specific probe based on complimentary base pairing. The fluorescence is then detected and analysed using software.

Primer extension

Primer extension is a two-step process that first involves the hybridization of a probe to the bases immediately upstream of the SNP nucleotide followed by a 'mini-sequencing' reaction, in which DNA polymerase extends the hybridized primer by adding a base that is complementary to the SNP nucleotide. This incorporated base is detected and determines the SNP allele (Head *et al* 1999, Syvanen 2001). Primer extension is based on highly accurate DNA polymerase enzyme and hence is generally very reliable. Primer extension is able to genotype most SNPs under very similar reaction conditions making it highly flexible. The primer extension method is used in a number of assay formats. These formats use a wide range of detection techniques that include MALDI-TOF Mass spectrometry and ELISA (Enzyme linked Immuno Sorbent Assay) like methods. Probes hybridise to the target DNA just upstream of the SNP nucleotide and a single, ddNTP complementary to the SNP allele is added to the probe's 3' end (the dideoxynucleotide's missing 3'-hydroxyl prohibits additional nucleotide addition). Each ddNTP is labelled with a unique fluorescent signal, allowing all four alleles to be detected in the same experiment. Probes hybridise to target DNA slightly upstream of the SNP nucleotide, and a single ddNTP complementary to the SNP allele is added to the probe's 3' end (the dideoxynucleotide's lack of a 3'-hydroxyl prevents the addition of additional nucleotides). Each ddNTP has its own fluorescent signal, all four alleles can be detected in the same experiment.

Oligonucleotide ligation

The ligation of the 3' end of a DNA fragment to the 5' end of a directly adjacent DNA fragment is catalyzed by DNA ligase. By hybridizing two probes directly over the SNP polymorphism site, ligation can occur if the probes are similar to the target DNA, this approach can be used to interrogate an SNP. Two probes are used in the oligonucleotide ligation assay: an allele-specific probe that hybridizes to the target DNA and places its 3' base directly over the SNP nucleotide, and a second probe that hybridizes the template upstream (downstream in the complementary strand) of the SNP polymorphic site, providing a 5' end for the ligation reaction. The allele-specific probe will fully hybridize to the target DNA if it matches the allele. If the target DNA does not contain an allele complementary to the probe's 3' base, the target DNA will produce a mismatch at the 3' end of the probe and DNA polymerase will not be able to extend from the 3' end of the probe.

Genotyping by Sequencing (SNP)

Genotyping by sequencing (GBS) is a highly efficient strategy for genome-wide SNP detection, and an alternative to microarray chips in order to perform genotyping studies such as genome-wide association studies in the field of genetic sequencing (GWAS). To minimise genome complexity and genotype multiple DNA samples, restriction enzymes are used. Following digestion, PCR is used to increase the number of fragments in the pool, and then GBS libraries are sequenced using next-generation sequencing technology, yielding 100bp single-end reads in most

cases. It's a low-cost option that's been applied in plant breeding. GBS uses a similar methodology to the restriction-site-associated DNA sequencing (RAD-seq) method, with some differences. Elshire *et al* was the first to describe the procedure in 2011. High molecular weight DNA is extracted and digested using a specific Restriction Enzyme which is a frequent cutter. The most popular restriction enzyme is ApeKI. Barcode adapters are then ligated to sticky ends, and PCR amplification is carried out. The sequencing is done using next-generation sequencing technology, which results in single-end reads of roughly 100 bp. Burrows-Wheeler alignment tool (BWA) or Bowtie 2 are used to filter and align raw sequencing data to a reference genome. SNPs from aligned tags are found and all identified SNPs are scored for coverage, depth, and genotypic statistics. Once a large-scale, species-wide SNP production has been run, it is possible to quickly call known SNPs in newly sequenced samples. GBS is a versatile and cost-effective procedure that will allow mining genomes of any species without prior knowledge of its genome structure.

InDels as molecular markers

InDels (insertions and deletions) are structural variations found throughout the genome that can lead to gain or loss of function in the organism. InDels arise as a result of polymerase slippage, transposons, unequal crossing-over, and other factors. Single base pair insertion and deletion, monomeric base pair expansion, and multi base pair expansion are the most common types of InDels. Simple gel-based size separation techniques can genotype InDels, and the lack of stutter bands make InDels a valuable marker system. InDels have also been discovered to be more polymorphic than microsatellite markers in other prior investigations.

RNA based molecular markers

These are similar to DNA based markers which are derived from RNA analysis where instead of DNA m-RNA or cDNA is used.

cDNA-SSCP: estimation of relative abundance of mRNAs encoded by similar homologous genes of polyploid species

RAP-PCR: RNA fingerprinting by arbitrarily primed PCR

cDNA-AFLP: discriminates between genes belonging to the same gene family and are highly homologous. Identification of genes related to novel processes.

Epigenetic markers

Epigenetics is the study of heritable phenotype changes that do not involve alterations in the DNA sequence or it is the study of a change in gene function without any change in the gene base sequence. It can arise due to DNA methylation, RNA interference, and histone modifications (acetylation, methylation, phosphorylation, and ubiquitination) (Figure 8).

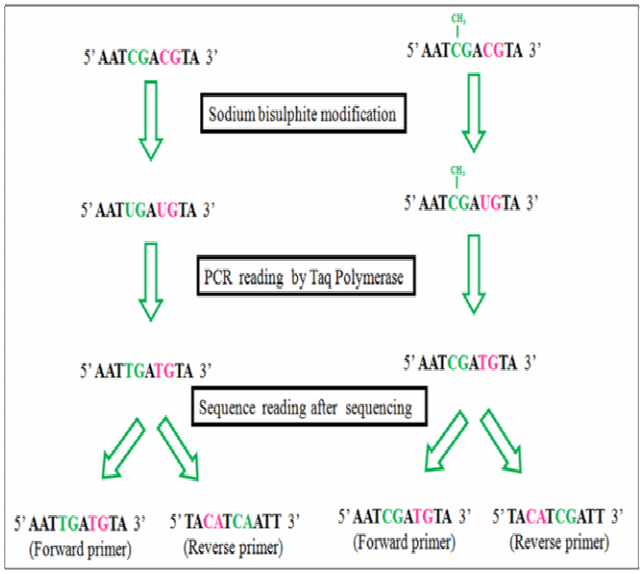

Fig. 8 : Epigenetic changes used as markers (Guerrero-Bosagna, 2013)

Epigenetic changes can be used as markers which is used to detect these variations:

- Whole genome bisulphite sequencing
- Methyl RAD
- Amino-precipitation based method (methyl cap sequencing)

Methylation of DNA usually occurs in the cytosine. Treatment of sodium bisulphite converts normal cytosine to uracil while the methylated cytosine is not changed and remains as such. PCR amplification of this DNA results in replacement of "U" uracil with "T" thymine. The DNA is then sequenced and read methylated regions are observed as cytosine while the non-methylated regions are seen as thymine. This helps in the detection of DNA methylations which are one of the reasons for epigenetic changes.

Conclusion

Over the past few decades, molecular markers have become increasingly important in plant biotechnology and genetics due to their ability to reveal DNA-level polymorphisms. The first DNA markers, RFLPs were widely used until the 1990s. Progress in PCR technology led to the development of new DNA markers, from RAPD to AFLP and beyond, with microsatellites quickly becoming the preferred marker system. With the dawn of the new millennium, DNA sequencing techniques saw transformative

advancements, particularly with the emergence of Next-Generation Sequencing (NGS) tools. Large-scale sequencing efforts uncovered Single Nucleotide Polymorphisms (SNPs) single base-pair variations that are abundant, easily automated, and suited to high-throughput genotyping. The rise of SNPs challenged the dominance of SSR markers. Additionally, transposable element-based markers, DNA methylation markers, and RNA-based markers have been developed to fulfill specific research needs. Some markers, like SRAP, MITE (Miniature Inverted-Repeat Transposable Elements), and TE-AFLP (Transposable Element-AFLP), are still in early usage stages and are not yet widely adopted (Doveri *et al* 2008). However, the current trend strongly favors SNPs as the "marker of choice" (Singh and Singh, 2015). The search for new marker systems remains active, with nearly two dozen marker types now available.

DNA markers have transformed plant breeding through various applications, including Marker-Assisted Selection (MAS) (Islam *et al.,* 2024, Prysiazhniuk *et al* 2024, Vargheese *et al.,* 2024), Marker-Assisted Recurrent Selection (MARS) (Saavedra *et al.,* 2023, Suryendra *et al.,* 2020), Genome-Wide Association Studies (GWAS) (Kumar *et al.,* 2023, Yuan *et al.,* 2020, Akram *et al.,* 2021), and Genomic Selection (GS) (Ma and Cao, 2021). These markers are also widely used to gain comprehensive insights into phylogeny and evolution. For example, Wang *et al*. (2017) studied genetic diversity and population structure in wild soybean using chloroplast and nuclear gene sequences. Various studies have employed molecular markers to examine genetic diversity and hybrid vigor in crop plants like wheat (Singh-Bakala *et al.,* 2024), maize (Arora *et al.,* 2024), and rapeseed (Motallebinia *et al.,* 2024). Recently, SSR markers have been crucial in analyzing diversity and heterosis in rice (Mukta *et al.,* 2024, Rezk *et al.,* 2024, El-Malky *et al.,* 2024). Additionally, SSR and SNP markers have been used to identify doubled haploids (DH) and characterize genotypes of isogenic lines and hybrids, as demonstrated in maize (Tang *et al.,* 2006) and rice (Shahid *et al.,* 2013). The choice of marker system for breeding depends on project goals, budget, marker availability, and reproducibility, as research continues to seek and refine new marker systems.

References

Abravaya, K., Huff, J., Marshall, R., Merchant, B., Mullen, C., Schneider, G. and Robinson, J. (2003). Molecular beacons as diagnostic tools: technology and applications.

Akram, S., Arif, M.A.R. and Hameed, A. (2021). A GBS-based GWAS analysis of adaptability and yield traits in bread wheat (Triticum aestivum L.). *Journal of Applied Genetics*, **62**(1), 27-41.

Arora, A., Das, A.K., Dixit, S., KR, Y., Singh, S.B., Sekhar, J.C., Ravikesavan, R., Sahi, J.P., Kumar, I.S., Mahadevu, P. and Swain, D. (2024). Genetic diversity analysis and heterotic grouping of Indian white maize inbred lines using combining ability and SNP markers. *Crop Science*, **64**(2), 726-737.

Benoist, C., O'hare, K., Breathnach, R. and Chambon, P. (1980). The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Research*, **8**(1), 127-142.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, **32**(3), 314.

Chen, X.M., Line, R.F. and Leung, H. (1998). Genome scanning for resistance-gene analogs in rice, barley, and wheat by high-resolution electrophoresis. *Theoretical and Applied Genetics*, **97**, 345-355.

Collard, B.C., Jahufer, M.Z.Z., Brouwer, J.B. and Pang, E.C.K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica*, **142**, pp.169-196.

de Vienne, D. ed. (2003). *Molecular markers in plant genetics and biotechnology*. CRC Press.

Doveri, S., Lee, D., Maheswaran, M. and Powell, W. (2008). Molecular markers-history, features and applications. 23-67.

El-Malky, M.M. (2024). Heat tolerance and genetic diversity analyses of rice accessions using ssr markers. *sabrao Journal of Breeding & Genetics*, **56**(2).

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, **6**(5), e19379.

Grodzicker, T., Williams, J., Sharp, P. and Sambrook, J. (1974). Physical mapping of temperature-sensitive mutations of adenoviruses. In *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press., **39**, 439-446.

Guerrero-Bosagna, C. (2013). DNA methylation research methods. *Mater. Methods*, *3*(10.13070).

Hawkin, J.D. (1988). A survey on intron and exon lengths. *Nucleic acids research*, **16**(21), 9893-9908.

Head, S.R., Parikh, K., Rogers, Y.H., Bishai, W., Goelet, P. and Boyce-Jacino, M.T. (1999). Solid-phase sequence scanning for drug resistance detection in tuberculosis. *Molecular and cellular probes*, **13**(2), 81-87.

Hu, J., Seiler, G.J., Jan, C.C. and Vick, B.A. (2003). Assessing genetic variability among sixteen perennial Helianthus species using PCR-based TRAP markers. In *Proc of 25th Sunflower Research Workshop, Fargo, North Dakota, USA* (pp. 1-5).

Islam, M.A., Hasan, M.M., Rahman, M.A., Akter, T. and Haque, M.A. (2024). Marker-assisted gene introgression for resistance to Xanthomonas oryzae pv. oryzae in rice for the control of bacterial leaf blight. *Euphytica*, **220**(4), 60.

Jeffreys, A.J., Wilson, V. and Thein, S.L.(1985). Hypervariable 'minisatellite'regions in human DNA. *Nature*, **314**(6006), 67-73.

Joshi, C.P., Zhou, H., Huang, X. and Chiang, V.L. (1997). Context sequences of translation initiation codon in plants. *Plant molecular biology*, **35**, 993-1001.

Kimura, M. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, **312**(1154), 343-354.

Kumar, M., Kumar, S., Sandhu, K.S., Kumar, N., Saripalli, G., Prakash, R., Nambardar, A., Sharma, H., Gautam, T., Balyan, H.S. and Gupta, P.K. (2023). GWAS and genomic prediction for pre-harvest sprouting tolerance involving sprouting score and two other related traits in spring wheat. *Molecular Breeding*, **43**(3), 14.

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

Ma, J. and Cao, Y. (2021). Genetic dissection of grain yield of maize and yield-related traits through association mapping and genomic prediction. *Frontiers in Plant Science*, **12**, 690059.

Motallebinia, S., Sofalian, O., Asghari, A., Rasoulzadeh, A. and Achachlouei, B.F. (2024). Genetic diversity and morpho-physiological assessment of drought tolerance in rapeseed (*Brassica napus* L.) cultivars. *Plant Genetic Resources*, **22**(3), 181-190.

Mukta, S., Bappy, M.N.I., Bhuiyan, J., Zohora, F.T. and Afrin, D. (2024). Assessment of genetic diversity in Bangladeshi rice (*Oryza sativa* L.) varieties utilizing SSR markers. *Gene Reports*, p.102051.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor symposia on quantitative biology* . Cold Spring Harbor Laboratory Press, **51**, 263-273.

Olsen, D.B. and Eckstein, F. (1989). Incomplete primer extension during in vitro DNA amplification catalyzed by Taq polymerase; exploitation for DNA sequencing. *Nucleic acids research*, **17**(23), 9613-9620.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences*, **86**(8), 2766-2770.

Poczai, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J.P. and Hyvönen, J. (2013). Advances in plant gene-targeted and functional markers: a review. *Plant methods*, **9**, 1-32.

Prysiazhniuk, L., Honcharov, Y., Melnyk, S. and Kliachenko, O. (2023). The selection of maize parent lines within marker assisted selection (MAS) by crtRB1-3'TE marker for Steppe zone of Ukraine.

Rezk, A.A., Mohamed, H.I. and El-Beltagi, H.S. (2024). Genetic variability and diversity analysis in Oryza sativa L. genotypes using quantitative traits and SSR markers. *Saudi Journal of Biological Sciences*, **31**(3), 103944.

Riaz, A., Li, G., Quresh, Z., Swati, M.S. and Quiros, C.F. (2001). Genetic diversity of oilseed Brassica napus inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. *Plant breeding*, **120**(5), 411-415.

Saavedra, L.M., Caixeta, E.T., Barka, G.D., Borém, A., Zambolim, L., Nascimento, M., Cruz, C.D., Oliveira, A.C.B.D. and Pereira, A.A. (2023). Marker-assisted recurrent selection for pyramiding leaf rust and coffee berry disease resistance alleles in Coffea Arabica L. *Genes*, **14**(1), 189.

Sawant, S.V., Singh, P.K., Gupta, S.K., Madnala, R. and Tuli, R. (1999). Conserved nucleotide sequences in highly expressed genes in plants. *Journal of Genetics*, **78**, 123-131.

Sax, K. (1923). The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. *Genetics*, **8**(6), 552.

Shahid, M.Q., Chen, F.Y., Li, H.Y., Wang, S.Z., Chen, P.F., Lin, S.Q., Liu, X.D. and Lu, Y.G. (2013). Double-Neutral Genes, Sa-n and Sb-n, for Pollen Fertility in Rice to Overcome Indica× Japonica Hybrid Sterility. *Crop Science*, **53**(1), 164-176.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M. and Solovyev, V.V. (2003). PlantProm: a database of plant promoter sequences. *Nucleic acids research*, **31**(1), 114-117.

Sheikh, I., Sharma, P., Verma, S.K., Kumar, S., Kumar, N., Kumar, S., Kumar, R., Vyas, P. and Dhaliwal, H.S. (2018). Development of intron targeted amplified polymorphic markers of metal homeostasis genes for monitoring their introgression from Aegilops species to wheat. *Molecular Breeding*, **38**, 1-11.

Singh, B.D. and Singh, A.K. (2015). Marker-assisted plant breeding: principles and practices.

Singh-Bakala, H., Srivastava, P., Singh, R., Sharma, A., Chhuneja, P. and Bains, N.S. (2024). Enhancing Genetic Diversity in Spring Restorer Germplasm for Hybrid Wheat Breeding Using Winter Wheat Genotypes. *Plant Molecular Biology Reporter*, 1-16.

Suryendra, P.J., Revathi, P., Singh, A.K. and Viraktamath, B.C. (2020). Marker assisted recurrent selection for genetic male sterile population improvement in rice. *Electronic Journal of Plant Breeding*, **11**(01), 149-155.

Tang, F., Tao, Y., Zhao, T. and Wang, G. (2006). In vitro production of haploid and doubled haploid plants from pollinated ovaries of maize (Zea mays). *Plant cell, tissue and organ culture*, **84**, 233-237.

Tazeb, A., Box, P.O. and Markos, D. (2018). Molecular marker techniques and their novel applications in crop improvement: A review article. *Global J Molecular Sci*, **13**(1), 01-16.

Thoday, J.M. and Boam, T.B. (1961). Regular responses to selection I. Description of responses. *Genetics Research*, **2**(2), 161-176.

Vargheese, R.L., Saravanan, S., Hepziba, S.J., Kumari, S.M.P., Pushpam, A.K., Kanagarajan, S. and Pillai, M.A. (2024). Marker-assisted introgression to improve the oleic acid content in the TMV 7 groundnut (Arachis hypogaea L.) variety suitable for the oil industry. *BMC Plant Biology*, **24**(1), 733.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T.V.D., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. and Zabeau, M. (1995.) AFLP: a new technique for DNA fingerprinting. *Nucleic acids research*, **23**(21), 4407-4414.

Wang, Q., Zhang, B. and Lu, Q. (2009). Conserved region amplification polymorphism (CoRAP), a novel marker technique for plant genotyping in Salvia miltiorrhiza. *Plant Molecular Biology Reporter*, **27**, 139-143.

Wang, Y., Ghouri, F., Shahid, M.Q., Naeem, M. and Baloch, F.S. (2017). The genetic diversity and population structure of wild soybean evaluated by chloroplast and nuclear gene sequences. *Biochemical Systematics and Ecology*, **71**, 170-178.

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

Williams, J.G., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic acids research*, **18**(22), 6531-6535.

Yuan, J., Wang, X., Zhao, Y., Khan, N.U., Zhao, Z., Zhang, Y., Wen, X., Tang, F., Wang, F. and Li, Z. (2020). Genetic basis and identification of candidate genes for salt tolerance in rice by GWAS. *Scientific reports*, **10**(1), 9958.

Zabeau, M. and Vos, P. (1993). European patent application, publication no: EP 0534858.