1

Molecular Marker Techniques and Recent Advancements

Dharminder Bhatia and Gagandeep Singh Bajwa

Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, Punjab, India

COPYRIGHTED MATERIAL

1.1 Introduction

Plant selection and systematic breeding efforts led to the development of present-day improved cultivars of crop plants. From a historical perspective, increased crop yield is the result of genetic improvement (Fehr 1984). Markers play an important role in the selection of traits of interest. Markers can be morphological, biochemical, or molecular in nature. Morphological markers are visual phenotypic characters such as growth habit of the plant, seed shape, seed color, flower color etc. Biochemical markers are the isozyme-based markers characterized by variation in molecular form of enzyme showing a difference in mobility on an electrophoresis gel. Very few morphological and biochemical markers are available in plants, and they are influenced by developmental stage and environmental factors. Since a large number of economically important traits are quantitative in nature, which are affected by both genetic and environmental factors, the morphological and biochemical markers-based selection of traits may not be much reliable. The subsequent discovery of abundantly available DNA-based markers made possible the selection of almost any trait of interest. DNA-based markers are not affected by the environment. Besides, these markers are highly reproducible across labs and show high polymorphism to distinguish between two genetically different individuals or species.

In the last four decades, DNA-based molecular marker technology has witnessed several advances from low throughput hybridization-based markers to high-throughput sequencing-based markers. These advances have been possible due to critical discoveries such as polymerase chain reaction (PCR) (Mullis et al. 1986), Sanger sequencing method (Sanger et al. 1977), automation of Sanger sequencing (Shendure et al. 2011), next-generation sequencing (NGS) technologies (Mardis 2008), and development of bioinformatics tools. This chapter will briefly discuss different types of molecular markers while particularly focusing on recent developments in molecular marker technologies. These

Genotyping by Sequencing for Crop Improvement, First Edition.
Edited by Humira Sonah, Vinod Goyal, S.M. Shivaraj, and Rupesh K. Deshmukh.
© 2022 John Wiley & Sons Ltd. Published 2022 by John Wiley & Sons Ltd.

developments have expedited the mapping and cloning of several loci governing important traits, precise trait selection, and transfer into elite germplasm.

1.2 What is a Molecular Marker?

DNA or molecular marker is a fragment of the DNA that is associated with a particular trait in an individual. These molecular markers aid in determining the location of genes that control key traits.

Generally, molecular markers do not represent the gene of interest but act as "flags" or "signs." Similar to genes, all the molecular markers occupy a specific position within the chromosomes. Molecular markers located close to genes (i.e. tightly linked) are referred to as "gene tags."

DNA-based molecular markers are the most widely used markers predominantly due to their abundance. They arise from different classes of DNA mutations such as substitution mutations (point mutations), rearrangements (insertions or deletions), or errors in replication of tandemly repeated DNA. These markers are selectively neutral because they are usually located in noncoding regions of DNA. Unlike morphological and biochemical markers, DNA markers are practically unlimited in number and are not affected by environmental factors and/or the developmental stage of the plant.

DNA markers show genetic differences that can be visualized by using a gel electrophoresis technique and staining ethidium bromide or hybridization with radioactive or colorimetric probes. Markers that can identify the difference between two individuals are referred to as polymorphic markers, whereas those that do not distinguish the individuals are called monomorphic markers. Based on how polymorphic markers can discriminate between individuals, they are described as codominant or dominant. Codominant markers indicate differences in size whereas dominant markers reveal differences based on their presence or absence. The different forms of a DNA marker in the form of band size on gels are known as marker "alleles." Dominant marker has only two alleles whereas codominant markers may have many alleles.

1.3 Classes of Molecular Markers

Based on the method of their detection, DNA markers are broadly classified into three groups: (i) hybridization-based, (ii) PCR-based, and (iii) DNA sequence-based molecular markers. Molecular markers have been discussed earlier in several reviews (Collard et al. 2005; Semagn et al. 2006; Gupta and Rustgi 2004) and book chapters (Mir et al. 2013; Singh and Singh 2015), which readers can also consult for more details. However, a brief description of each of these markers has been presented below.

1.3.1 Hybridization-based Markers

1.3.1.1 Restriction Fragment Length Polymorphism (RFLP)

These are the first molecular markers used by Grodzicker et al. (1975) in adenovirus and Botstein et al. (1980) in human genome mapping. These were first used in plants by Helentjaris et al. (1986). In this type of marker, polymorphism is detected by cutting DNA

into fragments by the use of restriction enzymes followed by hybridization of radioactively labeled DNA probes which are single or low copy DNA fragments and visualized by autoradiography. DNA probes could be genomic clones, cDNA clones, or even cloned genes. The RFLP markers show co-dominance and are highly reliable in linkage analysis and breeding (Semagn et al. 2006). However, this technique requires a large quantity of DNA, labor-intensive, relatively expensive, and hazardous. RFLP shows polymorphism in two different species if they differ due to point mutations, insertion/deletion, inversion, translocation, and duplication.

1.3.1.2 Diversity Array Technology (DArT™)

This is a high-throughput DNA polymorphism analysis method which combines microarray and restriction-based PCR methods. It is similar to AFLP where hybridization is used for the detection of polymorphism. It can able to provide a comprehensive genome coverage even in those organisms not having genome sequence information (Jaccoud et al. 2001). Diversity array technology (DArT) is a solid-state open platform method for analyzing DNA polymorphism. DArT procedure includes (i) Generating a diversity panel and (ii) Genotyping using a diversity panel. The diversity panel is generated using a set of lines representing the breadth of variability in germplasm (~10 lines). An equal quantity of DNA from each representative line is pooled followed by restriction with two to three restriction endonucleases (REs) and ligation of RE-specific adaptors. Later DNA fragments are amplified using adaptor complementary primers. The representation fragments are ligated to vector and transformed into *Escherichia coli* cells. The transformed cells with recombinant DNA are selected and amplified using M13 forward and reverse primer. The amplified DNA is isolated and purified. The purified DNA is coated onto polylysine-coated glass slides to generate a diversity array.

For genotyping, the representation fragments of the target genotypes are prepared in the same as in the diversity panel. The DNA fragments are column purified and fluorescently labeled with two different dyes (Cy3 or Cy5). The labeled DNA fragments are used for hybridization onto the diversity array. Two representative panels – one labeled with Cy3 and another with Cy5 – can be hybridized simultaneously and hybridization signal intensities are measured for each spot. DArT, thus detects DNA polymorphism at several hundred genomic loci in a single array without relying on sequence information.

1.3.2 Polymerase Chain Reaction (PCR)-based Markers

1.3.2.1 Simple-Sequence Repeats (SSRs)

Simple-sequence repeats (SSRs) (Litt and Luty 1989) are also known as microsatellites or short tandem repeats (STRs) or simple sequence length polymorphism (SSLP). These are widely used markers and are also referred to as the mother of all the markers. These are STRs, generally of one to eight nucleotide length. These are found dispersed throughout the genome and are hypervariable. These repeat regions are flanked with unique sequences that are highly conserved. The flanking unique sequences are used to design complementary primers which can be assayed with PCR. SSRs are highly polymorphic and codominant markers. These show polymorphism as a result of the variable number of repeat units. Before the era of genome sequencing, it was difficult to develop SSRs due to the extensive cost and labor involved in the identification of repeat regions and flanking unique

sequences. However, with the availability of genome sequences of several organisms, the development of SSR has become very easy which involves in silico identification of STRs, designing of SSR from flanking unique sequences, and validation through experimentation. SSR markers have shown immense application in population genetic analysis, gene mapping, and cloning due to their abundance in the genome and high polymorphism, and very high reproducibility across labs. SSR-based linkage maps have been developed in several important crop plants such as rice (Temnykh et al. 2000; McCouch et al. 2002; Orjuela et al. 2010), wheat (Roder et al. 1998), maize (Sharopova et al. 2002), potato (Milbourne et al. 1998), etc.

1.3.2.2 Sequence-Tagged Sites (STSs)

Sequence-tagged sites (STSs) were first developed for physical mapping of the human genome by Olsen et al. (1989). STS is the short unique sequences developed from polymorphic RFLP probe or AFLP fragment which is linked to desirable traits. The RFLP probes or AFLP fragments showing polymorphism are end-sequenced and primers are designed to specifically amplify these fragments. STS markers are co-dominant and highly reproducible. For example, STS markers have been developed for RFLP markers linked with bacterial blight resistance genes *xa5, xa13,* and *Xa21* (Huang et al. 1997). One major limitation of these types of markers is the reduced polymorphism than the corresponding RFLP probe.

1.3.2.3 Randomly Amplified Polymorphic DNAs (RAPDs)

Williams et al. (1990) first developed these markers to amplify DNA without prior sequence information. In this type of marker, the arbitrary decamer sequences are used as primers at low annealing temperatures for DNA amplification. These markers are referred to as dominant markers because the polymorphism is determined based on the presence or absence of a particular amplified fragment. Polymorphism may also be due to varying brightness of bands at a particular locus due to copy number differences. These markers have been used for constructing linkage maps in several species (Hunt 1997; Laucou et al. 1998) and also for tagging genes of economic importance. However, due to the dominant nature, these may not be appropriate for genetic mapping and marker-assisted selection (MAS). One major limitation of these markers is the lack of repeatability in certain cases. Variations of RAPD include AP-PCR (arbitrarily primed PCR) and DAF (DNA amplification fingerprinting (Table 1.1).

Table 1.1 Details of the other important molecular markers.

Marker	Description
Variable number tandem repeat (VNTR) or minisatellites	A short DNA sequence (10–100 bp) is present as tandem repeats and is a highly variable copy number
DNA amplification fingerprinting (DAF)	A variation of RAPD, where 4–5 bp single and arbitrary primer is used to detect polymorphism
Arbitrary-primed PCR (AP-PCR)	A variation of RAPD, where 18–32 bp long single and arbitrary primer is used to detect polymorphism

Table 1.1 (Continued)

Marker	Description
Inter-simple sequence repeat (ISSR)	Primers are designed based on the repeat region of microsatellites. These primers are used to amplify the region between two microsatellites. The stretches of unique DNA in between or flanking the SSRs are amplified. A single SSR-based primer is used to prime PCR
Selective amplification of microsatellite polymorphic loci (SAMPL)	A modification of ISSR, where SSR-based primer is used along with AFLP primer. The template is identical to the AFLP template and the rare cutter primer is replaced by SSR-based primer
Cleaved amplified polymorphic sequences (CAPS)	These markers are also called PCR-RFLP, where amplified PCR product is digested with endonucleases to reveal polymorphism. These are used when PCR product does not show polymorphism and restriction enzyme site present in amplified PCR product may detect polymorphism
Derived cleaved amplified polymorphic sequences (dCAPS)	A variation of CAPS, where a primer containing one or more mismatches to template DNA is used to create a restriction enzyme recognition site in one allele but not in another due to the presence of SNP. Thus, obtained PCR product is subjected to restriction enzyme digestion to find the presence or absence of the SNP
Single-strand conformational polymorphism (SSCP)	DNA fragments of size ranging from 200 to 800 bp were amplified by PCR using specific primers (20–25 bp), followed by gel-electrophoresis of single-strand DNA to detect nucleotide sequence variation. The method is based on a principle that the secondary structure of single-strand DNA molecule changes significantly if it harbors mutation. This method detects nucleotide variation without sequencing a DNA sample
Denaturing/temperature gradient gel electrophoresis (DGGE, TGGE)	These methods reveal polymorphism due to differential movement of the same genomic double-stranded region with different base-pair composition. As an example, the AT-rich region would have a lower melting temperature than the GC-rich region
Target region amplification polymorphism (TRAP)	This method employs primers designed from the EST database for detecting polymorphism around a selected candidate gene. This includes two primers of 18 bp, one of which is designed from targeted EST and the other is an arbitrary primer

1.3.2.4 Sequence Characterized Amplified Regions (SCARs)

These markers overcome the limitation of RAPDs. In this case, the RAPD fragments that are linked to a gene of interest are cloned and sequenced. Based on the terminal sequences, longer primers (20 mer) are designed. These SCAR primers more specifically amplify a particular locus. These are similar to STS markers in design and application. The presence or absence of the band indicates variation in sequences. The SCAR markers thus are dominant in nature. These, however, can be converted to codominant markers in certain cases by digesting the amplified fragment with tetranucleotide recognizing restriction enzymes. There are several examples where the RAPD markers linked to the gene of importance have been converted to SCAR markers (Joshi et al. 1999; Liu et al. 1999; Kasai et al. 2000; Akkurt et al. 2007; Chao et al. 2018).

1.3.2.5 Amplified Fragment Length Polymorphism (AFLP)

This marker technique was developed by Vos et al. (1995) and is patented by Keygene (www.keygene.com). In this technique, DNA is cut into fragments by a combination of restriction enzymes which are frequent (four bases) and rare (six bases) cutters that generate restriction overhangs on both sides of fragments. This is followed by the annealing of double-stranded oligonucleotide adapters of a few oligonucleotide bases with respective restriction overhangs. The oligonucleotide adapters are designed in such a way that the original restriction sites are not reinstated and also provide the PCR amplification sites. The fragments are PCR amplified and visualized on agarose gel. This method produces many restriction fragments enabling the polymorphism detection. The number of amplified DNA fragments can be controlled by selecting different number or composition of bases in the adapters. The stringent reaction conditions used for primer annealing make this technique more reliable. This method is a combination of both RFLP and PCR techniques and is extremely useful in the detection of polymorphism between closely related genotypes. Like RAPD, AFLP is a dominant marker and is not preferred for genetic mapping studies and MAS. AFLP maps have been constructed in several species and integrated into already existing RFLP maps e.g. tomato (Haanstra et al. 1999), rice (Cho et al. 1997), and wheat (Lotti et al. 2000).

1.3.2.6 Expressed Sequence Tags (ESTs)

These markers are developed by end sequencing (generally 200–300 bp) of random cDNA clones. The sequence thus obtained is referred to as expressed sequence tags (ESTs). A large number of ESTs have been synthesized in several crop plants and are available in the EST database at NCBI (https://www.ncbi.nlm. nih.gov/dbEST/). These markers were originally developed to identify gene transcripts and have played important role in the identification of several genes and the development of markers such as RFLP, SSR, SNPs, CAPS, etc. (Semagn et al. 2006). However, EST-based SSRs show less polymorphism as compared to genomic DNA-based SSRs. Since EST markers are from expressed sequence regions, these are highly conserved among the species and can be used for synteny mapping. Most of these could also be functional genes. A large number of EST markers have been used in rice for developing a high-density linkage map (Harushima et al. 1998) and for chromosome bin mapping in wheat using deletion stocks (Qi et al. 2003). In addition to these, several other molecular marker variants have been developed. The description of those markers is presented in Table 1.1.

1.4 Sequencing-based Markers

1.4.1 Single-Nucleotide Polymorphisms (SNPs)

Single-nucleotide polymorphisms (SNPs) are more abundant resulted from single-base pair variations. These are evenly distributed in a whole genome that can tag almost any gene or locus of a genome (Brookes 1999). However, the distribution of SNPs varies among species with 1 SNP per 60–120 bp in maize (Ching et al. 2002) and 1 SNP per 1000 bp in humans (Sachidanandam et al. 2001). SNPs are more prevalent in the noncoding region. In the coding region, SNPs could be synonymous or nonsynonymous. In synonymous SNPs, there is no change in the amino acid resulting in no phenotypic differences. However, phenotypic differences could be produced due to modified mRNA splicing (Richard and Beckman 1995). In nonsynonymous SNPs, change in amino acid results in phenotypic differences. SNPs are mostly bi-allelic and cause polymorphism due to nucleotide base substitution. The two types of nucleotide base substitutions result in SNPs. A transition substitution occurs between purines (A, G) or between pyrimidines (C, T). This type of substitution constitutes two-thirds of all SNPs. A transversion substitution occurs between a purine and pyrimidine. SNPs can be detected by the alignment of the similar genomic region of two different species. The SNPs have only two alleles compared to typical multiallele SSLP; however, this disadvantage can be compensated by using the high density of SNPs.

1.4.2 Identification of SNP in a Pregenomic Era

Initially, identification of SNP markers was laborious and expensive and involved allele-specific sequencing (Ganal et al. 2009). This includes sequencing of unigene-derived amplicons using Sanger's method from two or more than two lines. In an experiment, about 350 bp of the RFLP clone, A-519 was end sequenced in soybean and the flanking amplification primers were designed (Coryell et al. 1999). Primers were used to screen for allele diversity using PCR from ten genotypes and the amplicons were sequenced followed by sequence comparison to identify SNP. SNPs were also identified through mining a large number of EST sequences in EST databases, which are generated through improved sequencing technologies (Soleimani et al. 2003). These SNPs are further validated using PCR (Batley et al. 2003). These approaches allowed the identification of mainly gene-based SNPs, but their frequency is generally low. Additionally, SNPs located in low-copy noncoding regions and intergenic spaces could not be identified.

Several assays have been developed for genotyping based on identified SNPs which include, allele-specific hybridization, primer extension, oligonucleotide ligation, and invasive cleavage (Sobrino et al. 2005). Besides, DNA chips, allele-specific PCR, and primer extension were also attractive options since these are suitable for automation and can be used for the development of dense genetic maps. Allele-specific hybridization was used for the identification of polymorphism in 570 genotypes of soybean (Coryell et al. 1999).

1.5 Recent Advances in Molecular Marker Technologies

The improvement of Sanger sequencing technology in the 1990s combined with the beginning of EST and genome sequencing projects in model plants led to the spurt in the identification of variation at the single-base resolution (Wang et al. 1998). From 2005 onward, the

emergence of NGS platforms such as Roche 454, Illumina HiSeq2500, ABI 5500xl SOLiD, Ion Torrent, PacBio RS, Oxford Nanopore, and advances in bioinformatics tools simplified the process of identification of genome-wide SNPs and changed the face of molecular marker technology. NGS-based genotyping platforms such as genotyping-by-sequencing (GBS), whole-genome resequencing (WGR), and high-density SNP arrays helped to type thousands of SNPs in a single reaction in hundreds of individuals.

1.5.1 Genotyping-by-Sequencing (GBS)

GBS is an NGS-based reduced representation sequencing technique for the identification of genome-wide SNPs and genotyping large populations (Bhatia et al. 2013). GBS is a one-step approach for the identification and utilization of markers in a single reaction. It is a complexity reduction procedure where a combination of restriction enzymes is used to separate low copy sequences from high copy repetitive regions. In general, GBS involves the sequencing of fragments generated through restriction digestion of the genome on the NGS platform. In this process, the DNA of the population is digested with RE followed by ligation of RE-specific adaptors containing genotype-specific barcode sequences and sites for binding PCR and sequencing primers (Figure 1.1). The fragments thus generated can be PCR amplified and an equal volume of PCR product from different individuals are pooled in a tube. The fragments in the pool can be selected based on their size and sequenced on

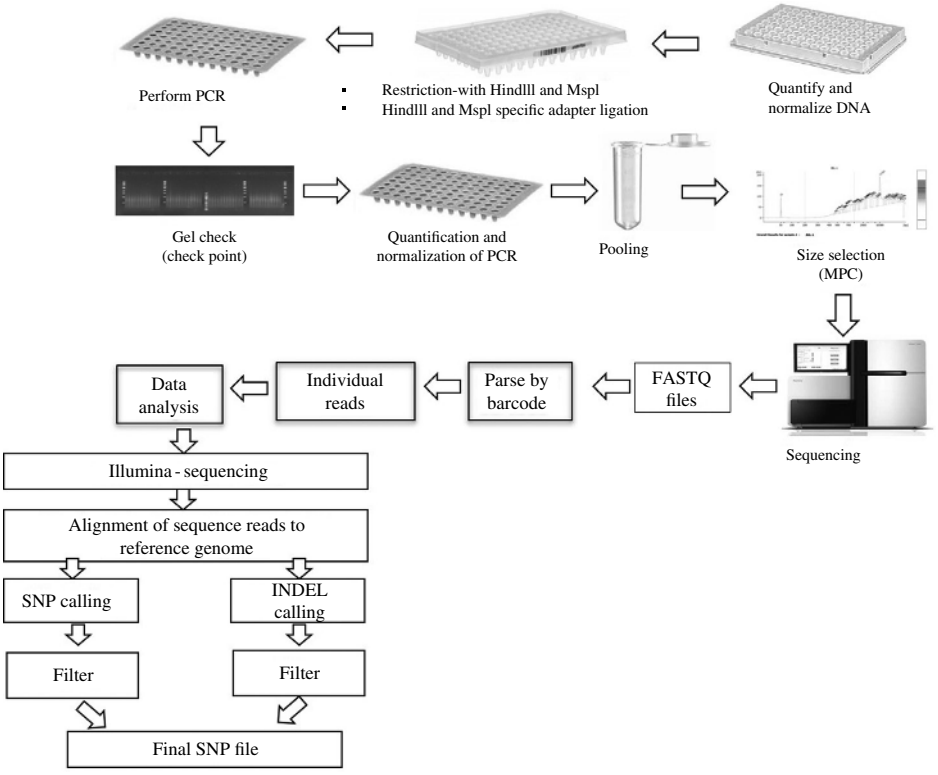

Figure 1.1 An example of GBS and GBS data analysis workflow for identification of SNP markers.

the NGS platform. The choice of restriction enzymes depends upon the complexity and size of the genome. Presently, different versions of GBS are available, which includes RAD-seq (restriction associated DNA sequencing), ddRAD-seq (double-digest restriction associated sequencing), SLAF-seq (specific-locus amplified fragment sequencing), Rest-seq (restriction DNA sequencing), Skim GBS (skim-based GBS) (Bhatia 2020). These versions differ with respect to fragment size selection, the extent of complexity reduction, and genome coverage. Since GBS is a population-dependent genotyping method, to make it cost-effective a low-depth sequencing is adopted which caused a high rate of missing data. The low-depth sequencing makes it an ineffective genotyping approach in heterozygous populations. GBS has low genome coverage due to reduced representation sequencing.

GBS is being widely used to capture SNPs and other marker variations by NGS. GBS overtook the conventional genotyping procedures involving the use of traditional markers such as RAPD, AFLP, SSR, and many others in terms of time, labor, and cost involved. As an example, GBS can generate data of thousands of markers in a large population in a week, which can be analyzed in a month (Bhatia et al. 2018). The approach has been utilized in the mapping of several economically important traits in a number of crop plants (Poland and Rife 2012). Most of the developing countries have in-house computational facilities that are being used for GBS analysis. Few online servers are also available, where GBS analysis can be done using in-built pipelines such as cyverse (www.cyverse.org); however, these are unable to analyze the large dataset. Further speed of analysis depends upon the internet speed. Alignment of NGS-based reads and calling SNPs and Indels are the two major steps in GBS analysis, for which several pipelines are available publically such as Stacks, IGST, GB-eaSY, TASSEL-GBS, FAST-GBS, UNEAK, etc. (Wickland et al. 2017).

Another important pipeline widely used for NGS data analysis is dDocent pipeline (www.dDocent.com) which is a simple bash wrapper to quality analysis, assemble, map, and call SNPs from almost any kind of RAD sequencing (Puritz et al. 2014). However, most of these pipelines are hard to code for a student with little bioinformatics background. Most of these pipelines vary with respect to the complexity of the genome and computational space required. Besides there are several bioinformatics tools such as BWA, Bowtie2, SAM tools, GATK, BCFtools including a set of Perl utility scripts (Kagale et al. 2016) that can be used for GBS data analysis. However, there should be knowledge of the installation and usage of these tools for proper utilization in data analysis. With the advancements in NGS approaches, GBS has become a widely used approach in plant breeding and genetics, particularly for understanding complex quantitative traits.

DArT-seq GBS (https://www.diversityarrays.com/technology-and-resources/dartseq/) somehow overcomes the limitation of the missing data point. The technique is an extension of traditional DArT technology where DArT representations are sequenced on the NGS platform. The fragment sequencing enables a dramatic increase in the number of genomic fragments analyzed and an increase in the number of reported markers thus making it a cost-effective technology than the initial DArT method.

1.5.2 Whole-Genome Resequencing (WGR)

WGR with high coverage and depth overcomes the limitations of GBS due to missing data points and heterozygous calls. In general, WGR involves the sequencing of enough DNA fragments (>5×–20×) to cover the whole genome of an organism. Due to sequencing cost,

the technique is suitable in crop plants having smaller genome sizes such as rice. In such cases, GBS can be replaced by resequencing of a larger size population at 5–6× depth. However, WGR for few samples can be done at a much higher read depth of 10–20× as in the case of the BSA-seq approach (Nguyen et al. 2019). One of the important BSA-seq-based approaches is quantitative trait loci (QTL)-seq developed by Takagi et al. (2015) in rice. Later this technique has been widely used in several crop plants. Takagi et al. (2015) developed a pipeline for analysis of the whole genome sequence of bulks and identification of causative variants. WGR has been used in several studies for identification of genome-wide SNPs, genotyping mapping populations for construction of high-density linkage maps and QTL mapping, linkage and genome-wide association studies (GWASs), of reference genome improvement, and genomic selection (Poland and Rife 2012; Bhatia et al. 2013; Chung et al. 2017; Nguyen et al. 2019).

1.5.3 SNP Arrays

Along with GBS, high-density DNA array-based SNP chips or SNP arrays have become a widely used SNP detection platform for high multiplex genotyping. SNP arrays work by hybridization of DNA fragments with allele-specific oligonucleotide probes (SNP probes) and fluorescence-based detection of signals. In general, SNP arrays can be roughly categorized into two types based on SNP detection methods: (i) nonenzymatic differential hybridization including allele-specific hybridization, (ii) enzymatic reactions including primer extension, and mini-sequencing (Ding and Jin 2009). For making SNP arrays, the first step is the identification of genome-wide SNPs by sequencing (preferably WGR) of a large diverse panel. The SNPs arrays may include SNPs from coding (genic) regions only and/or genome-wide SNPs from other noncoding regions. SNPs are in silico validated with several custom tools and final filtered SNPs are identified. The oligonucleotide probes containing SNP alleles are designed and bound on a solid glass plate surface. SNP chips can be custom designed commercially from two widely used platforms: Affymetrics (www.affymetrics.com) as Axiom Affymetrics SNP Chips (Affymetrix/Thermo Fisher Axiom®) or Illumina (https://www.illumina.com/science/technology/microarray.html) as Immunia Infinium assay (Illumina Infinium®). Affymetrics SNP array relies on differential hybridization due to different melting temperatures for matched and mismatched SNPs binding to target DNA sequence. On the other hand, Illumina Infinium assay uses Illumina BeadArray technology that relies on primer extension to distinguish two SNP alleles. The Affymetrix SNP array uses 25-mer for SNP calling while the Illumina BeadArray uses 50-mer for target capture. In rice, a high-resolution 44K Affymetrix array, 50K Infinium array, and 700K high-density rice array are available for rice SNP genotyping (McCouch et al. 2010; Tung et al. 2010; Chen et al. 2013; McCouch et al. 2015). Additionally, high-density SNP arrays have been developed for other crop plants such as maize (Ganal et al. 2011) and sunflower (Bachlava et al. 2012) as well as domestic animal species, including cattle (Gibbs et al. 2009; Matukumalli et al. 2009) and pig (Ramos et al. 2009). One major advantage of SNP arrays is the reproducibility of data points where GBS does have some shortcomings. However, the disadvantage is the less polymorphism as compared to GBS and WGR and detection of only alleles present in the array (Table 1.2).

Table 1.2 Comparison between different marker techniques commonly used in plant research.

	SSR	GBS	WGR	SNP array	KASP™
DNA quality	Moderate	High	High	High	High
PCR-based	Yes	Yes	No	No	No
Allele detection	High	High	High	Low	Low
Polymorphism	High	High	High	Low	Low
Ease to use	Easy	Not easy	Not easy	Easy	Easy
Reproducibility	High	Low	High	High	High
Cost	Moderate	Low to moderate	High	High	moderate
Cost for analysis	High	High	High	Low	Low
Suitability for different approaches					
Genetic diversity analysis	High	Moderate	High (cost concerns)	High	High
Bi-parental QTL mapping	High	High	High	High	High
Genome wide association analysis	Moderate	High	High	High	Low
Genomic selection	Low	Moderate	High (cost concerns)	High	Low

1.5.4 Kompetitive Allele-Specific PCR (KASP™)

KASP™ is a trademark technology of KBiosciences (http://www.kbioscience.co.uk/) or LGC genomics (http://www.lgcgenomics.com) initially developed for in-house genotyping, thereafter evolving as a benchmark technology for SNP genotyping. Any candidate SNP identified through GBS or WGR and associated with any important traits can be validated through KASP assay. Furthermore, any identified candidate SNP associated with a trait of interest can be readily converted into KASP assay to serve as a robust and cost-effective marker to be used as a MAS tool for crop improvement. It works on the principle of competitive allele-specific PCR permitting bi-allelic scoring of SNP, insertion, and deletions (InDels) at a specific location in the genome (Figure 1.2). KASP genotyping reaction consists of DNA sample, KASP assay mix, and universal KASP master mix. Allele-specific two forward primers and common reverse primer all unlabelled constitute KASP assay mix. Allele-specific primers have a unique tail sequence complementary to FRET (fluorescence resonant energy transfer) cassette. Each allele harbors a tail sequence linked to different dyes (FAM™ and HEX™ dyes). KASP master mix has FRET cassettes in the quenched state, a passive reference dye (ROX™), and other components for PCR reaction. During the first round of reaction allele-specific primer binds to template incorporating tail sequence in newly synthesized strands. In the next round of PCR, a complementary strand of the allele-specific tail sequence is generated allowing the FRET cassette to bind enabling an unquenched state, and generating a fluorescent allele-specific signal. In the case of a homozygous DNA sample, only one signal specific to the allele is seen and mixed signal is generated in the case of the heterozygous individual.

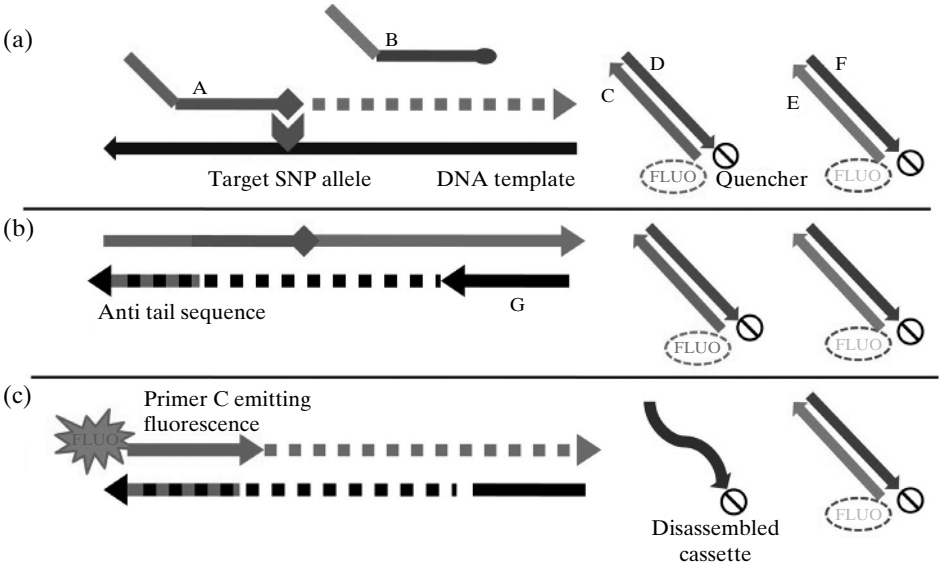

Figure 1.2 Steps in KASP reaction: (a) annealing: allele-specific primer binds to target SNP, (b) extension: anti tail sequence generation leading to disassembly of allele-specific FRET cassette, and (c) fluorescent emission: sample emitting fluorescence on exposing to a specific wavelength. *Source:* The figure is reproduced from Rosas et al. (2014) available with CC-BY-4.0.

It can be carried out in 96- to 1536-well plate format. Application of KASP includes QC (quality control) analysis, QTL mapping, allele mining (Semagn et al. 2013), and MAS. However, it may not be a suitable platform for genome-wide association mapping and genomic selection due to fewer data points. KASP markers have been utilized extensively for MAS in major crops like rice (Yang et al. 2019; Kang et al. 2019), wheat (Makhoul et al. 2020; Fang et al. 2020; Grewal et al. 2020), and maize (Zhang et al. 2017; Su et al. 2016).

1.6 SNP Databases

SNP databases correspond to a publically available archive of genetic information of economically important species. Recent developments in genome sequencing technologies have ushered the era of cost-effective, high-throughput genomics resulting in the creation of huge datasets of sequence information. In a similar manner, developments in computing facilities and data sciences had enabled us to compare, categorize, and compute relationship matrices among and between species leading to the creation of databases. As numerous labs are working on the same species, collaborative consortiums were established to avoid redundancy leading to high-quality SNP databases containing a range of molecular variations constituting SNP's, insertions, and deletions (InDels), trait-specific characterized SNP's and called variants. Table 1.3 enlists some important SNP databases corresponding to humans, model genetic organisms, and important crop species. These public databases serve as an important resource for crop improvement for genetic diversity analysis, establishing a genetic association and linkage disequilibrium studies.

Table 1.3 List of important online SNP databases.

SNP database	Organism	URL
dbSNP	Human	http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp
Ensembl	Human	http://www.ensembl.org/Homo_sapiens/index.html
1001 Genomes	Arabidopsis	https://1001genomes.org/
CropSNPdb	Brassica crops, wheat	http://snpdb.appliedbioinformatics.com.au
SNP-Seek Database	Rice 3K panel	https://snp-seek.irri.org/
MaizeSNPDB	1210 Maize inbred lines	http://150.109.59.144:3838/MaizeSNPDB/
CerealDB	Wheat	http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php

1.7 Application of Molecular Markers

1.7.1 Application of Molecular Markers in Crop Improvement

Molecular markers have several applications in genetic studies and crop improvement programs. These have been used in the development of saturated linkage maps, gene/QTL mapping, map-based cloning of genes, orthologous gene mapping, and marker-assisted transfer of targeted genes/QTLs in the background of different cultivars/lines. Saturation of linkage maps refers to increased marker density to cover the entire chromosomal region. In general, when molecular markers are arranged on a linkage map with less than 1 cM distance apart, is considered as saturated linkage map. The development of saturated linkage maps could only be possible with the availability of molecular markers. These maps are prerequisite for gene/QTL mapping, map-based cloning of genes, and MAS. Several molecular markers-based saturated linkage maps have been developed in crop plants including rice (Harushima et al. 1998; McCouch et al. 2002; IRGSP 2005; Zhu et al. 2017; Kumar et al. 2018), wheat (Somers et al. 2004; Song et al. 2005; Poland et al. 2012; Li et al. 2015; Hussain et al. 2017), maize (Sharopova et al. 2002; Zhou et al. 2016; Su et al. 2017), and tomato (Tanksley et al. 1992; Haanstra et al. 1999; Sim et al. 2012). In one of the studies, a rice genetic map helped to enrich the genetic region of Ph1 locus of wheat and facilitated the identification of candidate genes governing the locus (Sidhu et al. 2008).

Plant breeders have relied heavily on generating new gene combinations and selecting these new gene combinations empirically. Though phenotype-based selection has largely been successful, MAS has improved the efficiency and precision of selection. MAS can be practiced more efficiently for characters whose phenotypic selection is difficult. As an example, selecting a fertility restorer gene in segregating generations need test crossing before subsequent backcrossing. However, if such genes are tagged with molecular markers, desirable plants with fertility restorer genes (in heterozygous

condition), can be identified and backcrossed. Similarly, if desirable genes conferring tolerance to abiotic stresses are tagged, these can be selected easily in segregating generations. Also, genetic markers can be assayed in nontarget areas such as growth chambers, greenhouses, or off-season nurseries, thus permitting more rapid progress. The efficiencies of scale and time accorded by DNA markers are valuable in breeding horticultural plants where fewer individuals might save several hectares and fewer generations may save several decades (Paterson et al. 1991). MAS has been used efficiently in (i) gene pyramiding (Huang et al. 1997; Singh et al. 2001; Bhatia et al. 2011; Kumar et al. 2013; Yasuda et al. 2015), (ii) marker-assisted alien introgressions (Jena et al. 1992; Brar and Dhaliwal 1997; Elkot et al. 2015), and (iii) simultaneous identification and pyramiding of QTLs from primitive cultivars and wild species (Tanksley and Nelson 1996; Tanksley et al. 1996; Fulton et al. 1997; Xiao et al. 1996, 1998).

1.7.2 Role of Molecular Markers in Germplasm Characterization

Molecular markers are also used in DNA fingerprinting for varietal identification, germplasm evaluation, phylogenetic and evolutionary studies, etc. The molecular marker-based DNA fingerprinting data are useful for the characterization of plant germplasm accessions, quantification of genetic diversity, and protection of proprietary germplasm (Smith and Smith 1992). Molecular markers have been utilized to distinguish closely related crop cultivars (Melchinger et al. 1991; Paull et al. 1998), in sex identification of dioecious plants (Parasnis et al. 1999). They are also used to understand evolutionary relationships within and between species, genera, or higher taxonomic groups. Such studies involve large number of markers to study similarities and differences among taxa (Paterson et al. 1991). Although phylogeny has been established for many plant species based on morphological markers, biochemical markers, and chromosome homology, the genetic markers have enhanced our understanding of phylogeny. In one important study, DNA-based markers enabled the designation of GG for *Oryza granulata* and HHJJ for *Oryza ridleyi (*Aggarwal et al. 1997*)*.

1.7.3 Deployment of Molecular Markers in Plant Variety Protection and Registration

The current system of plant variety protection and registration using assessment of DUS (Distinctness, Uniformity, and Stability) characteristics predominantly relies upon morphological traits which are quite laborious, time-consuming, requires skills, expertise, and evaluation under special designs for most quantitative traits. With the advent of novel breeding technologies, new varieties differ only for few traits or at few loci which make the process of detecting distinctness in varieties, a challenging task. Even with increasing numbers of plant varieties, DUS testing is becoming quite expensive. In a review by Jamali et al. (2019), DNA-based molecular markers have been proposed to be a reliable alternative for conducting DUS testing. Molecular markers not only cut cost, time and labor but will also help in the proper sharing of Plant Breeder's Rights with an assessment of few distinct traits, particularly in essentially derived varieties.

1.8 Summary

The DNA-based molecular markers are widely recognized for their enormous potential in plant breeding and genetic studies. The past few years have seen remarkable developments in the field of molecular markers technology particularly with the emergence of NGS technologies. SNPs have become the choice of markers of present and future based on their genome-wide abundance, high polymorphism, amenability to high-throughput automation, and easier analysis. SNPs can be utilized in different genotyping platforms such as GBS, DArT, WGR, SNP arrays, and KASP. Any genotyping platform can be chosen based on the objective of the study and cost concerns. As an example, GBS, WGR, SNP arrays, and KASP can generate similar type of results for genetic diversity analysis; however, KASP and GBS could be cheaper than others. While the development of KASP will depend upon the availability of SNPs particularly in SNP databases, but GBS can be done without any previous information. It has been observed that these recent marker genotyping technologies have accelerated the crop improvement programs particularly in the identification and utilization of novel genes and QTLs.

References

Aggarwal, R.K., Brar, D.S., and Khush, G.S. (1997). Two new genomes in the *oryza* complex identified on the basis of molecular divergence analysis using total genomic hybridization. *Molecular & General Genetics* 254: 1–12.

Akkurt, M., Welter, L., Maul, E. et al. (2007). Development of SCAR markers linked to powdery mildew (*Uncinula necator*) resistance in grapevine (*Vitis vinifera* L. and *Vitis* sp.). *Molecular Breeding* 19: 103–111.

Bachlava, E., Taylor, C.A., Tang, S. et al. (2012). SNP discovery and development of a high-density genotyping array for sunflower. *PLoS One* 7 (1): e29814.

Batley, J., Barker, G., O'Sullivan, H. et al. (2003). Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiology* 132 (1): 84–91.

Bhatia, D. (2020). Advanced quantitative genetics technologies for accelerating plant breeding. In: *Accelerated Plant Breeding*, vol. I (eds. S.S. Gosal and S.H. Wani). Springer Verlag https://doi.org/10.1007/978-3-030-41866-3_5.

Bhatia, D., Sharma, R., Vikal, Y. et al. (2011). Marker assisted development of bacterial blight resistant, dwarf and high yielding versions of two traditional Basmati rice cultivars. *Crop Science* 51: 759–770.

Bhatia, D., Wing, R.A., and Singh, K. (2013). Genotyping by sequencing, its implications and benefits. *Crop Improvement* 40: 101–111.

Bhatia, D., Wing, R.A., Yu, Y. et al. (2018). Genotyping by sequencing of rice interspecific backcross inbred lines identifies QTLs for grain weight and grain length. *Euphytica* 214: 41. https://doi.org/10.1007/s10681-018-2119-1.

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 (314): 331.

Brar, D.S. and Dhaliwal, H.S. (1997). Molecular markers and their application in crop improvement. In: *Proceedings 3rd Agricultural Science Congress* (eds. M.S. Bajwa et al.). New Delhi, India: Natl. Acad. Agr. Sci.

Brookes, A.J. (1999). The essence of SNPs. *Gene* 234: 177–186.

Chao, W.Z., Tang, C.H., Zhang, J.S. et al. (2018). Development of a stable SCAR marker for rapid identification of *Ganoderma lucidum* Hunong 5 cultivar using DNA pooling method and inter-simple sequence repeat markers. *Journal of Integrative Agriculture* 17 (1): 130–138.

Chen, H., Xie, W., He, H. et al. (2013). A high density SNP genotyping array for rice biology and molecular breeding. *Molecular Plant* 7: 541–553.

Ching, A., Caldwell, K.S., Jung, M.T., and Dolan, M. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3 (1): 19. https://doi.org/10.1186/1471-2156-3-19.

Cho, Y.G., McCouch, S.R., Kuiper, M. et al. (1997). Integration of AFLP markers into an RFLP and SSLP map of rice (*Oryza sativa* L.). Rice Genet. *Newsletter* 14: 106–109.

Chung, Y.S., Choi, S.C., Jun, T.-H., and Kim, C. (2017). Genotyping by sequencing: a promising tool for plant genetics research and breeding. *Horticulture, Environment and Biotechnology* 58 (5): 425–431.

Collard, B.C.Y., 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: 169–196.

Coryell, V.H., Jessen, H., Schupp, J.M. et al. (1999). Allele-specific hybridization markers for soybean. *Theoretical and Applied Genetics* 98: 690–696.

Ding, C. and Jin, S. (2009, 2003). High-throughput methods for SNP genotyping. In: *Single Nucleotide Polymorphisms, Methods in Molecular Biology*, vol. 578 (ed. A.A. Komar). Humana Press, A Part of Springer Science+Business Media, LLC https://doi.org/10.1007/978-1-60327-411-1_16.

Elkot, A.F.A., Chhuneja, P., Kaur, S. et al. (2015). Marker assisted transfer of two powdery mildew resistance genes *PmTb7A.1* and *PmTb7A.2* from *Triticum boeoticum* (Boiss.) to *Triticum aestivum* (L.). *PLoS One* 10 (6): e128297.

Fang, T., Lei, L., Li, G. et al. (2020). Development and deployment of KASP markers for multiple alleles of *Lr34* in wheat. *Theoretical and Applied Genetics* 133: 2183–2195.

Fehr, W. (1984). *Genetic Contributions to Yield Gains of Five Major Crop Plants*. Madison, Wisconsin: Special publication No. 7 Crop Science Society of America.

Fulton, T.M., Nelson, J.C., and Tanksley, S.D. (1997). Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of cultivated tomato into *L. esculentum* followed through three successive backcross generations. *Theoretical and Applied Genetics* 95: 895–902.

Ganal, M.W., Altmann, T., and Roder, M.S. (2009). SNP identification in crop plants. *Current Opinion in Plant Biology* 12 (2): 211–217.

Ganal, M.W., Durstewitz, G., Polley, A. et al. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One* 6.

Gibbs, R.A. et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324: 528–532.

Grewal, S., Othmeni, M., Walker, J. et al. (2020). Development of Wheat-Aegilops caudata introgression lines and their characterization using genome-specific KASP markers. *Frontiers in Plant Science* 11: 606.

Grodzicker, T., Williams, J., Sharp, P., and Sambrook, J. (1975). Physical mapping of temperature sensitive mutants of adenovirus. *Cold Spring Harbor Symposia on Quantitative Biology* 39: 439–446.

Gupta, P.K. and Rustgi, S. (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. *Functional & Integrative Genomics* 4: 139–162.

Haanstra, J.P.W., Wye, C., Odinot, P. et al. (1999). An integrated high density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* x *L. pennellii* F_2 populations. *Theoretical and Applied Genetics* 99: 254–271.

Harushima, Y., Yano, M., Shomura, A. et al. (1998). A high density rice genetic linkage map with 2275 markers using a single F_2 population. *Genetics* 148: 479–494.

Helentjaris, T., Slocum, M., Wright, S. et al. (1986). Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theoretical and Applied Genetics* 61: 650–658.

Huang, N., Angeles, E.R., Domingo, J. et al. (1997). Pyramiding bacterial blight resistance gene in rice: marker assisted selection using RFLP and PCR. *Theoretical and Applied Genetics* 95: 313–320.

Hunt, G.J. (1997). Construction of linkage maps with RAPD markers. In: *Fingerprinting Methods Based on Arbitrarily Primed PCR* (eds. M.R. Micheli and R. Bova). Berlin, Heidelberg: Springer Lab Manuals. Springer https://doi.org/10.1007/978-3-642-60441-6_22.

Hussain, W., Baenziger, P.S., Belamkar, V. et al. (2017). Genotyping-by-Sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. *Scientific Reports* 7: 16394.

IRGSP (2005). The map-based sequence of the rice genome. *Nature* 436: 793–800. (Suppl. Table 18).

Jaccoud, D., Peng, K., Feinstein, D., and Kilian, A. (2001). Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29: e25.

Jamali, S.H., Cockram, J., and Hickey, L.T. (2019). Insights into deployment of DNA markers in plant variety protection and registration. *Theoretical and Applied Genetics* 132: 1911–1929.

Jena, K.K., Khush, G.S., and Kochert, G. (1992). RFLP analysis of rice introgression lines. *Theoretical and Applied Genetics* 84: 608–616.

Joshi, S.P., Ranjekar, P.K., and Gupta, V.S. (1999). Molecular markers in plant genome analysis. *Current Science* 77: 230–240.

Kagale, S., Koh, C., Clarke, W.E. et al. (2016). Analysis of genotyping by sequencing data. *Methods in Molecular Biology* 1374: 269–284.

Kang, J.-W., Shin, D., Cho, J.H. et al. (2019). Accelerated development of rice stripe virus-resistant, near-isogenic rice lines through marker-assisted backcrossing. *PLoS One* 14 (12): e0225974.

Kasai, K., Morikawa, Y., Sorri, V.A. et al. (2000). Development of SCAR markers to the PVY resistance gene *Ryadg* based on a common feature of plant disease resistance genes. *Genome* 43: 1–8.

Kumar, A., Dixit, S., and Henry, A. (2013). Marker-assisted introgression of major QTLs for grain yield under drought in rice. In: *Translational Genomics for Crop Breeding Volume 2: Abiotic Stresses, Yield and Quality* (eds. R.K. Varshney and R. Tuberosa). Wiley.

Kumar, K., Sarao, P.S., Bhatia, D. et al. (2018). High-resolution genetic mapping of novel brown planthopper resistance locus, *Bph34* in *Oryza sativa* X *Oryza nivara* (Sharma &

Shastry) derived interspecific F_2 population. *Theoretical and Applied Genetics* 131: 1163–1171.

Laucou, V., Haurogne, K., Ellis, N., and Rameau, C. (1998). Genetic mapping in pea. 1. RAPD-based genetic linkage map of *Pisum sativum*. *Theoretical and Applied Genetics* 97: 905–915.

Li, H., Vikram, P., Singh, R.P. et al. (2015). A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics* 16: 216.

Litt, M. and Luty, J.M. (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: 397–401.

Liu, Z., Sun, Q., Ni, Z., and Yang, T. (1999). Development of SCAR markers linked to the *Pm21* gene conferring resistance to powdery mildew in common wheat. *Plant Breeding* 118: 215–219.

Lotti, C., Salvi, S., Pasqualone, A. et al. (2000). Integration of AFLP markers into an RFLP-based map of durum wheat. *Plant Breeding* 119 (5): 393–401.

Makhoul, M., Rambla, C., Voss-Fels, K.P. et al. (2020). Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. *Theoretical and Applied Genetics* 133 (8): 2413–2430.

Mardis, E.R. (2008). Next generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9: 387–402.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D. et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 4: e5350.

McCouch, S.R., Teytelman, L., Xu, Y. et al. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Research* 9: 199–207.

McCouch, S.R., Zhao, K., Wright, M. et al. (2010). Development of genome-wide SNP assays for rice. *Breeding Science* 60: 524–535.

McCouch, S.R., Wright, M.H., Tung, C.-W. et al. (2015). Open access resources for genome-wide association mapping in rice. *Nature Communications* 7: 10532.

Melchinger, A.E., Messemer, M.M., Lee, M. et al. (1991). Diversity and relationship among US maize inbreds revealed by RFLPs. *Crop Science* 31: 669–678.

Milbourne, D., Meyer, R.C., Collins, A.J. et al. (1998). Isolation, characterization and mapping of simple sequence repeat loci in potato. *Molecular & General Genetics* 259: 233–245.

Mir, R.R., Hiremath, P.J., Lizarazu-Riera, O., and Varshney, R.K. (2013). Evolving molecular marker technologies in plants: from RFLPs to GBS. In: *Diagnostics in Plant Breeding*, https://doi.org/10.1007/978-94-007-5687-8__11 (eds. T. Lübberstedt and R.K. Varshney). Dordrecht: Springer Science Business Media.

Mullis, K., Faloona, F., Scharf, S. et al. (1986). Specific enzymatic amplication of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposium on Quantitative Biology.* 51: 263–273.

Nguyen, K.L., Grondin, A., Courtois, B., and Gantet, P. (2019). Next-Generation sequencing accelerates crop genome discovery. *Trends in Plant Science* 24: 263–274.

Olsen, M., Hood, L., Cantor, C., and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* 245: 1434–1435.

Orjuela, J., Garavito, A., Bouniol, M. et al. (2010). A universal core genetic map for rice. *Theoretical and Applied Genetics* 120: 563–572.

Parasnis, A.S., Ramakrishna, W., Chowdari, K.V. et al. (1999). Microsatellite (GATA)n reveals sex-specific differences in papaya. *Theoretical and Applied Genetics* 99: 1047–1052.

Paterson, A.H., Tanksley, S.D., and Sorrels, M.E. (1991). DNA markers in crop improvement. *Advances in Agronomy* 46: 39–90.

Paull, J.G., Chalmers, K.J., Karakousis, A. et al. (1998). Genetic diversity in Australian wheat varieties and breeding material based on RFLP data. *Theoretical and Applied Genetics* 96: 435–446.

Poland, J. and Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genitics* 5: 92–102.

Poland, J.A., Brown, P.J., Sorrells, M.E., and Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* 7 (2): e32253.

Puritz, J.B., Hollenbeck, C.M., and Gold, J.R. (2014). dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *Peer Journal* 2: e431.

Qi, L., Echalier, B., Friebe, B., and Gill, B.S. (2003). Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. *Functional & Integrative Genomics* 3: 39–55. https://doi.org/10.1007/s10142-002-0063-5.

Ramos, A.M., Crooijmans, R.P.M.A., Affara, N.A. et al. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* 4: e6524.

Richard, I. and Beckman, J.S. (1995). How neutral are synonymous codon mutations? *Nature Genetics* 10: 259.

Roder, M.S., Korzun, V., Wendehake, K. et al. (1998). A microsatellite map of wheat. *Genetics* 149: 2007–2023.

Rosas, J.E., Bonnecarrère, V., and Pérez de Vida, V.F. (2014). One-step, codominant detection of imidazolinone resistance mutations in weedy rice (*Oryza sativa* L.). *Electronic Journal of Biotechnology* 17: 95–101.

Sachidanandam, R., Weissman, D., Schmidt, S.C. et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409 (6822): 928–933.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain chain-terminating inhibitors. *Proceedings of the National Academy Sciences of the United States of America* 74: 5463–5467.

Semagn, K., Bjornstad, A., and Ndjiondjop, M.N. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology* 5 (25): 2540–2568.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2013). Single nucleotide polymorphism genotyping using *Kometitive allele* specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* 33: 1–14.

Sharopova, N., McMullen, M.D., Schultz, L. et al. (2002). Development and mapping of SSR markers for maize. *Plant Molecular Biology* 48: 463–481.

Shendure, J.A., Porreca, G.J., Church, G.M. et al. (2011). Overview of DNA sequencing strategies. *Current Protocols in Molecular Biology* 7 (1): 1–7.1.23. https://doi.org/10.1002/0471142727.mb0701s96.

Sidhu, G.K., Rustgi, S., Shafqat, M.N. et al. (2008). Fine structure mapping of a gene-rich region ofwheat carrying *Ph1*, a suppressor of crossing over between homoeologous

chromosomes. *Proceedings of the National Academy Sciences of the United States of America* 15: 5815–5820.

Sim, S.-C., Durstewitz, G., Plieske, J. et al. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One* 7 (7): e40563. https://doi.org/10.1371/journal.pone.0040563.

Singh, B.D. and Singh, A.K. (2015). *Marker-Assisted Plant Breeding: Principles and Practices*. https://doi.org/10.1007/978-81-322-2316-0. Springer (India) Pvt. Ltd.

Singh, S., Sidhu, J.S., Huang, N. et al. (2001). Pyramiding three bacterial blight resistance genes (*xa*5, *xa*13 and *Xa*21) using marker assisted selection into indica cultivar PR106. *Theoretical and Applied Genetics* 102: 1011–1015.

Smith, J.S.C. and Smith, O.S. (1992). Fingerprinting crop varieties. *Advances in Agronomy* 47: 85–140.

Sobrino, B., Briona, M., and Carracedoa, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International* 154: 181–194.

Soleimani, V.D., Baum, B.R., and Johnson, D.A. (2003). Efficient validation of single nucleotide polymorphisms in plants by allele-specific PCR, with an example from barley. *Plant Molecular Biology Reporter* 21: 281–288.

Somers, D.J., Isaac, P., and Edwards, K. (2004). A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 109: 1105–1114.

Song, Q.J., Shi, J.R., Singh, S. et al. (2005). Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics* 110: 550–560.

Su, A., Song, W., Xing, J. et al. (2016). Identification of genes potentially associated with the fertility instability of S-Type cytoplasmic male sterility in maize via bulked segregant RNA-Seq. *PLoS One* 11 (9): e0163489.

Su, C., Wang, W., Gong, S. et al. (2017). High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. *Frontiers in Plant Science* 8: 706. https://doi.org/10.3389/fpls.2017.00706.

Takagi, H., Tamiru, M., Abe, A. et al. (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. *Nature Biotechnology* 33: 445–449.

Tanksley, S.D. and Nelson, J.C. (1996). Advanced backcross QTL analysis: a method for simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics* 92: 191–203.

Tanksley, S.D., Ganal, M.W., Prince, J.P. et al. (1992). High density molecular linkage maps of tomato and potato genomes. *Genetics* 132: 1141–1160.

Tanksley, S.D., Grandillo, S., Fulton, T.M. et al. (1996). Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and Applied Genetics* 92: 213–224.

Temnykh, S., Park, W.D., Ayres, N. et al. (2000). Mapping and genome organization of mirosatellite sequences in rice (*Oryza sativa*) Theor. *Applied Genetics* 100: 697–712.

Tung, C.W., Zhao, K., Wright, M.H. et al. (2010). Development of a research platform for dissecting phenotype genotype associations in rice (*Oryza* spp.). *Rice* 3: 205–217.

Vos, P., Hogers, R., Bleeker, M. et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407–4414.

Wang, D.G., Fan, J.B., Siao, C.J. et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077–1082.

Wickland, D.P., Battu, G., Hudson, K.A. et al. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics* 18: 586.

Williams, J.G.K., Kublelik, A.R., Livak, K.J. et al. (1990). DNA polymorphism's amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531–6535.

Xiao, J., Grandillo, S., Ahn, S.N. et al. (1996). Genes from wild rice improved yield. *Nature* 384: 223–224.

Xiao, J., Li, J., Grandillo, S. et al. (1998). Identification of trait improving QTL alleles from a wild rice relative *Oryza rufipogon*. *Genetics* 150: 899–909.

Yang, G., Chen, S., Chen, L. et al. (2019). Development of a core SNP arrays based on the KASP method for molecular breeding of rice. *Rice* 12: 21.

Yasuda, N., Mitsunaga, T., Hayashi, K. et al. (2015). Effects of pyramiding quantitative resistance genes *pi21, Pi34*, and *Pi35* on rice leaf blast disease. *Plant Disease* 99: 904–909.

Zhang, X., Yang, Q., Rucker, E. et al. (2017). Fine mapping of a quantitative resistance gene for gray leaf spot of maize (*Zea mays* L.) derived from teosinte (*Z. mays* ssp. *parviglumis*). *Theoretical and Applied Genetics* 130: 1285–1295.

Zhou, Z., Zhang, C., Zhou, Y. et al. (2016). Genetic dissection of maize plant architecture with an ultra-high density bin map based on recombinant inbred lines. *BMC Genomics* 17: 178. https://doi.org/10.1186/s12864-016-2555-z.

Zhu, M., Liu, D., Liu, W. et al. (2017). QTL mapping using an ultra-high-density SNP map reveals a major locus for grain yield in an elite rice restorer R998. *Scientific Reports* 7: 10914. https://doi.org/10.1038/s41598-017-10666-7.