

## Review Article

# SNP Markers and Their Impact on Plant Breeding

Jafar Mammadov,<sup>1</sup> Rajat Aggarwal,<sup>1</sup> Ramesh Buyyarapu,<sup>1</sup> and Siva Kumpatla<sup>2</sup>

<sup>1</sup> Department of Trait Genetics and Technologies, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

<sup>2</sup> Department of Biotechnology Regulatory Sciences, Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268, USA

Correspondence should be addressed to Jafar Mammadov; jamammadov@dow.com

Received 3 August 2012; Accepted 14 November 2012

Academic Editor: Ian Bancroft

Copyright © 2012 Jafar Mammadov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The use of molecular markers has revolutionized the pace and precision of **plant genetic analysis** which in turn facilitated the implementation of **molecular breeding of crops**. The last three decades have seen tremendous advances in the evolution of marker systems and the respective detection platforms. Markers based on single nucleotide polymorphisms (SNPs) have rapidly gained the center stage of molecular genetics during the recent years **due to their abundance in the genomes and their amenability for high-throughput detection formats and platforms**. Computational approaches dominate SNP discovery methods due to the ever-increasing sequence information in public databases; however, complex genomes pose special challenges in the identification of informative SNPs warranting alternative strategies in those crops. Many genotyping platforms and chemistries have become available making the use of SNPs even more attractive and efficient. This paper provides a review of historical and current efforts in the development, validation, and application of SNP markers in QTL/gene discovery and plant breeding by discussing key experimental strategies and cases exemplifying their impact.

## 1. Introduction

Allelic variations within a genome of the same species can be classified into three major groups that include differences in the number of tandem repeats at a particular locus [microsatellites, or simple sequence repeats (SSRs)] [1], segmental insertions/deletions (InDels) [2], and single nucleotide polymorphisms (SNPs) [3]. In order to detect and track these variations in the individuals of a progeny at DNA level, researchers have been developing and using genetic tools called molecular markers [4]. Although SSRs, InDels, and SNPs are the three major allelic variations discovered so far, a plethora of molecular markers were developed to detect the polymorphisms that resulted from these three types of variation [5]. Evolution of molecular markers has been primarily driven by the throughput and cost of detection method and the level of reproducibility [6]. Depending on detection method and throughput, all molecular markers can be divided into three major groups: (1) low-throughput, hybridization-based markers such as restriction fragment length polymorphisms (RFLPs) [4]; (2) medium-throughput, PCR-based markers that include

random amplification of polymorphic DNA (RAPD) [7], amplified fragment length polymorphism (AFLP) [8], SSRs [9]; (3) high-throughput (HTP) sequence-based markers: SNPs [3]. In late eighties, RFLPs were the most popular molecular markers that were widely used in plant molecular genetics because they were reproducible and codominant [10]. However, the detection of RFLPs was an expensive, labor- and time-consuming process, which made these markers eventually obsolete. Moreover, RFLP markers were not amenable to automation. Invention of PCR technology and the application of this method for the rapid detection of polymorphisms overthrew low-throughput RFLP markers, and new generation of PCR-based markers emerged in the beginning of nineties. RAPD, AFLP, and SSR markers are the major PCR-based markers that research community has been using in various plant systems. RAPDs are able to simultaneously detect polymorphic loci in various regions of a genome [11]. However, they are anonymous and the level of their reproducibility is very low due to the non-specific binding of short, random primers. Although AFLPs are anonymous too, the level of their reproducibility and sensitivity is very high owing to the longer +1 and

+3 selective primers and the presence of discriminatory nucleotides at 3' end of each primer. That is why AFLP markers are still popular in molecular genetics research in crops with little to zero reference genome sequence available [12]. However, AFLP markers did not find widespread application in molecular breeding owing to the lengthy and laborious detection method, which was not amenable to automation either. Therefore, it was not surprising that soon after the discovery of SSR markers in the genome of a plant, they were declared as "markers of choice" [13], because SSRs were able to eliminate all drawbacks of the above-mentioned DNA marker technologies. SSRs were no longer anonymous; they were highly reproducible, highly polymorphic, and amenable to automation. Despite the cost of detection remaining high, SSR markers had pervaded all areas of plant molecular genetics and breeding in late 90s and the beginning of 21st century. However, during the last five years, the hegemony of medium-throughput SSRs was eventually broken by SNP markers. First discovered in human genome, SNPs proved to be universal as well as the most abundant forms of genetic variation among individuals of the same species [14]. Although SNPs are less polymorphic than SSR markers because of their biallelic nature, they easily compensate this drawback by being abundant, ubiquitous, and amenable to high- and ultra-high-throughput automation. However, despite these obvious advantages, there were only a limited number of examples of application of SNP markers in plant breeding by 2009 [15]. In this paper, we tried to summarize the recent progress in the utility of SNP markers in plant breeding.

## 2. SNP Discovery in Complex Plant Genomes

While SNP discovery in crops with simple genomes is a relatively straightforward process, complex genomes pose serious obstacles for the researchers interested in developing SNPs. One of the major problems is the highly repetitive nature of the plant genomes [16]. Prior to the emergence of next-generation sequencing (NGS) technologies, researchers used to rely on different experimental strategies to avoid repetitive portions of the genome. These include discovery of SNPs experimentally by resequencing of unigene-derived amplicons using Sanger's method [17] and *in silico* SNP discovery through the mining of SNPs within EST databases followed by PCR-based validation [18]. Although these approaches allowed the detection of gene-based SNPs, their frequency is generally low in conserved genic regions, and they were unable to discover SNPs located in low-copy noncoding regions and intergenic spaces. Additionally, amplicon resequencing was an expensive and labor-intensive procedure [15]. As many crops are ancient tetraploids with mosaics of scattered duplicated regions [19], *in silico* and experimental mining of EST databases resulted in the discovery of a large number of nonallelic SNPs that represented paralogous sequences and were suboptimal for application in molecular breeding [20]. Recent emergence of NGS technologies such as 454 Life

Sciences (Roche Applied Science, Indianapolis, IN), HiSeq (Illumina, San Diego, CA), SOLiD and Ion Torrent (Life Technologies Corporation, Carlsbad, CA) has eliminated the problems associated with low throughput and high cost of SNP discovery [21]. Transcriptome resequencing using NGS technologies allows rapid and inexpensive SNP discovery within genes and avoids highly repetitive regions of a genome [22]. This methodology was successfully applied in several plant genomes, including maize [23], canola [24], eucalyptus [25], sugarcane [26], tree species [27], wheat [28], avocado [29], and black currant [30]. Originally developed for human disease diagnostic research, the NimbleGen sequence capture technology (Roche Applied Science, IN) [31] brought the detection of gene-based SNPs in plants into higher throughput and coverage level [32]. This technology consists of exon sequence capture and enrichment by microarray followed by NGS for targeted resequencing. Similar in-solution target capture technologies, such as Agilent SureSelect, are also commercially available for genome/exome mining studies. However, this technology would be efficient only for crops with available reference genome sequence or large transcriptome (EST) datasets, since the design of capture probes requires these reference resources.

Despite the attractiveness of SNP discovery via transcriptome or exome resequencing, this process is targeted, focusing solely on coding regions. It is obvious that the availability of SNPs within coding sequences is a very powerful tool for molecular geneticists to detect a causative mutation [33]. However, often QTL are located in noncoding regulatory sequences such as enhancers or locus control regions, which could be located several megabases away from genes within intergenic spaces [34]. Discovery of SNPs located within those regulatory elements via transcriptome or exon sequencing is limited. In order to discover SNPs in a genome-wide fashion and avoid repetitive and duplicated DNA, it is very important to employ genome complexity reduction techniques coupled with NGS technologies. Several genome complexity reduction techniques have been developed over the years, including High Cot selection [35], methylation filtering [36], and microarray-based genomic selection [37]. These techniques mainly reduce the number of repetitive sequences but lack the power to recognize and eliminate duplicated sequences, which cause the detection of false-positive SNPs. Unlike the above-mentioned techniques, recently developed genome complexity reduction technologies such as Complexity Reduction of Polymorphic Sequences (CRoPS) (Keygene N.V., Wageningen, The Netherlands) [38] and Restriction Site Associated DNA (RAD) (Florigenics, Eugene, OR, USA) [39] are computationally well equipped and capable of filtering out duplicated SNPs. These systems were successfully applied to discover SNPs in crops with [40] and without reference genome sequences [41].

Although several complexity reduction approaches are being developed to generate data from NGS platforms, it is often challenging to identify candidate SNPs in polyploid crops species such as potato, tobacco, cotton, canola, and wheat. In general, minor allele frequency could be used as

a measure to identify candidate SNPs in diploid species [42]. However, in polyploid crops, you often find loci that are polymorphic within a single genotype due to the presence of either homoeologous loci from the individual subgenomes (homoeologous SNPs) or paralogous loci from duplicated regions of the genome. Such false positive SNPs are not useful for genetic mapping purposes and often lead to a lower validation rate during assays. Successful SNP validation in allopolyploids depends upon differentiation of the sequence variation classes [43]. Use of haplotype information beside the allelic frequency would help to identify homologous SNPs (true SNPs) from those of homoeologous loci (false positives). Bioinformatic programs such as HaploSNPer [44] would facilitate identification of candidate loci for assay design purposes in polyploid crops. Elimination of homoeologous loci for the assay design process would improve the validation rate. Such approaches could also be extended to other complex and highly repetitive diploid genomes such as barley. Complexity reduction approaches, combined with sophisticated computational tools, would expedite SNP discovery and validation efforts in polyploids.

Although CROPS and RAD technologies are powerful tools to detect SNPs in genome-wide fashion, they can hardly be called HTP, because on an average only ~1,000 SNPs pass stringent quality control [40]. While these numbers are enough to generate genetic linkage maps of reasonable saturation and carry out preliminary QTL mapping, they are not adequate to implement genome-wide association studies (GWAS). Depending on the rate of linkage disequilibrium decay, GWAS might require several million genetic landmarks. From this point of view, genotyping-by-sequencing (GBS) technique offers many more opportunities. Discovery of a large number of SNPs using GBS was demonstrated in maize [45] and sorghum [46]. GBS not only increases the sequencing throughput by several orders of magnitude but also has multiplexing capabilities [47]. To eliminate a large portion of repetitive sequences, a type II restriction endonuclease, *ApeKI*, is applied to digest DNA prior to sequencing to generate reduced representation libraries (genome complexity reduction component), which are further subject to sequencing [47]. In polyploid crops, GBS might be challenging, but the associated complexity reduction methods could be used for SNP discovery. For discovery purposes, the availability of a reference genome is not an absolute requirement to implement GBS approach. However, in organisms that do not have a reference genome, GBS-derived SNPs must be validated using one of the techniques that are described in the following section, which might dramatically increase per marker price. Validation needs to be done primarily to discard paralogous SNPs. For organisms with a reference genome sequence, the validation step is replaced by *in silico* mapping of the sequenced fragments to the genome. Although GBS has the potential to discover several million SNPs, one of the major drawbacks of this technique is large numbers of missing data. To solve this problem, computational biologists developed data imputation models such as BEAGLE v3.0.2 [48] and IMPUTE v2 [49], to bring imputed data as close as possible to the real data [50, 51].

### 3. SNP Validation and Modern Genotyping Platforms and Chemistries

The availability of reference sequence and sophisticated software does not always guarantee that the discovered SNP can be converted into a valid marker. In order to insure that the discovered SNP is a Mendelian locus, it has to be validated. The validation of a marker is the process of designing an assay based on the discovered polymorphism and then genotyping a panel of diverse germplasm and segregating population. Compared to the collection of unrelated lines, a segregating population is more informative as a validation panel because it allows the inspection of the discriminatory ability and segregation patterns of a marker which helps the researcher to understand whether it is a Mendelian locus or a duplicated/repetitive sequence that escaped the software filter [40].

The most popular HTP assays/chemistries and genotyping platforms that are currently being used for SNP validation are Illumina's BeadArray technology-based Golden Gate (GG) [52] and Infinium assays [53], Life Technologies' TaqMan [54] assay coupled with OpenArray platform (TaqMan OpenArray Genotyping system, Product bulletin), and KBiosciences' Competitive Allele Specific PCR (KASPar) combined with the SNP Line platform (SNP Line XL; <http://www.kbioscience.co.uk>). These modern genotyping assays and platforms differ from each other in their chemistry, cost, and throughput of samples to genotype and number of SNPs to validate. The choice of chemistry and genotyping platform depends on many factors that include the length of SNP context sequence, overall number of SNPs to genotype, and finally the funds available to the researcher, because most of these chemistries still remain cost intensive. Comparative analyses of these four genotyping assays and platforms were described in Kumpatla et al. [55].

Though all genotyping chemistries and platforms are applicable to generate genotypic data in polyploid crops, analysis of SNP calls is somewhat challenging in polyploids due to multiallele combinations in the genotypes. SNPs in polyploid species can be broadly classified as simple SNPs, hemi-SNPs, and homoeo-SNPs. Here, we describe simple, hemi-, and homoeo-SNPs using an example of allele calls in tetraploid and diploid cotton species (Figure 1). Genomes of tetraploid cotton species, *Gossypium hirsutum* (AD<sub>1</sub>) and *G. barbadense* (AD<sub>2</sub>), consist of two subgenomes A and D, where A genome was derived from diploid progenitors, such as *G. herbaceum* (A<sub>1</sub>) and *G. arboreum* (A<sub>2</sub>), and D genome resulted from another diploid progenitor *G. raimondii* (D<sub>5</sub>). Simple, or true SNPs are markers that detect allelic variation between homologous loci of the same subgenome of two tetraploid samples. For example, in Figure 1(a), a SNP marker clearly detects polymorphism within A subgenomes of *G. hirsutum* (AD<sub>1</sub>) and *G. barbadense* (AD<sub>2</sub>) and separates samples into homozygous A (blue) and B (red) clusters. This marker does not discriminate polymorphism in D subgenome, because the D genome allele is absent there (pink dot in *G. raimondii*). In contrast to simple SNPs, hemi-SNPs detect allelic variation in the homozygous state in one sample and the heterozygous state in the other sample.

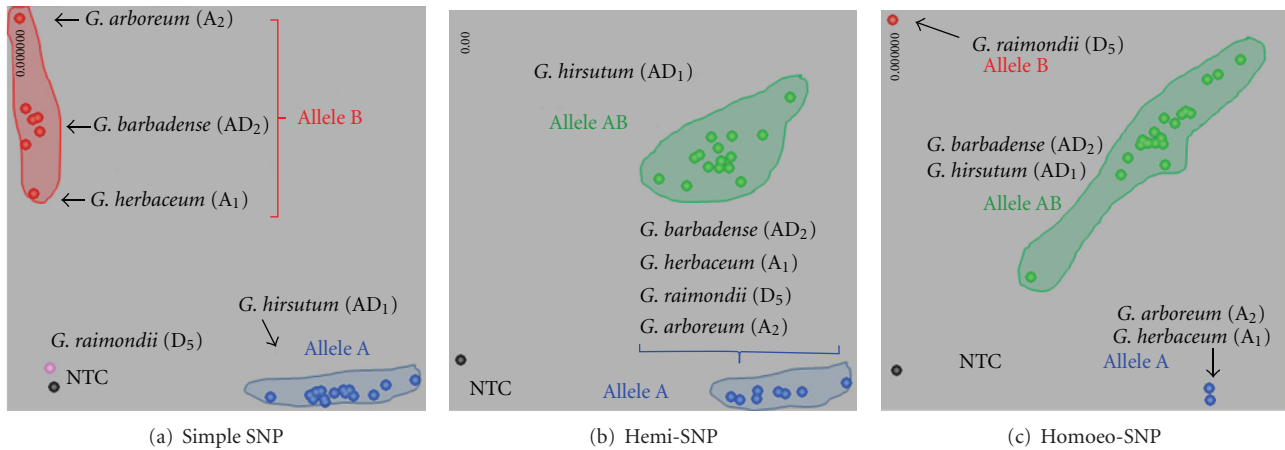


FIGURE 1: Segregation patterns of simple, hemi-, and homoeo-SNPs assayed using KASPar chemistry across tetraploid cotton species [*G. hirsutum* (AD<sub>1</sub>) and *G. barbadense* (AD<sub>2</sub>)] and their diploid progenitors [*G. arboreum*, *G. herbaceum* (A subgenome), and *G. raimondii* (D-subgenome)]. (a) Simple, or true SNP detects allelic variation between homologous loci of A subgenome of *G. hirsutum* (AD<sub>1</sub>) and *G. barbadense* (AD<sub>2</sub>). (b) Hemi-SNPs detect allelic variation in homozygous state in *G. barbadense* (AD<sub>2</sub>) and heterozygous state in *G. hirsutum* (AD<sub>1</sub>). (c) Homoeo-SNP detects homoeologous and/or paralogous loci both in A and D subgenomes, which are monomorphic between *G. barbadense* (AD<sub>2</sub>) and *G. hirsutum* (AD<sub>1</sub>). Allele calls depicted in blue, red, green, pink, and black represent alleles A, B, and AB, no amplification or missing locus, and no template control (NTC), respectively.

In Figure 1(b), SNP marker detects both alleles (A and B) in *G. hirsutum* (heterozygous green cluster) and one allele A in *G. barbadense* (a homozygous blue cluster) and could be *vice versa*. Homoeo-SNPs detect homoeologous and possibly paralogous loci both in A and D subgenomes and result in monomorphic loci in tetraploid species (right image). In Figure 1(c) A genome progenitors (*G. herbaceum* and *G. arboreum*) had allele A (blue) and D genome progenitor (*G. raimondii*) had allele B (red), but both tetraploid species (*G. hirsutum* and *G. barbadense*) were grouped into heterozygous AB (green) cluster. As homoeo-SNPs can detect paralogous loci, the diploid progenitors both have different alleles.

Simple SNPs as well as hemi-SNPs are useful markers for genetic mapping and diversity screening studies. Simple SNPs segregate like the markers in diploids in most of the mapping populations and would account for approximately 10–30% of total polymorphic SNPs in various polyploid crop species. Hemi-SNPs form a major category (30–60%) of polymorphic SNPs in a polyploid crop species and could be used for genetic mapping purposes in F<sub>2</sub>, RIL, and DH populations. Homoeo-SNPs are of lesser value for mapping purposes as most of the genotypes result in heterologous loci due to polymorphism between the homoeologous genomes or duplicated loci within each of the polyploid genotypes [56].

#### 4. Application of SNP Markers in Gene/QTL Discovery

**4.1. Biparental Approach.** Genetic mapping studies involve genetic linkage analysis, which is based on the concept of genetic recombination during meiosis [57]. This encompasses developing genetic linkage maps following genotyping

of individuals in segregating populations with DNA markers covering the genome of that organism. Since their discovery in the 1980s, DNA-based markers have been widely used in developing saturated genetic linkage maps as well as for the mapping and discovery of genes/QTL. With the large-scale availability of the sequence information and development of HTP technologies for SNP genotyping, SNP markers have been increasingly used for QTL mapping studies. This is primarily, because SNPs are highly abundant in the genomes and, therefore, they can provide the highest map resolution compared to other marker systems [58, 59]. A review of the selected examples of QTL and gene discovery using SNP markers is presented below.

**4.1.1. Examples in Rice.** A recent study on QTL analysis in rice for yield and three-yield-component traits, number of tillers per plant, number of grains per panicle, and grain weight compared a SNP-based map to that of a previous RFLP/SSR-based QTL map generated using the same mapping population [42]. Using the ultra-high-density SNP map, the authors showed that this map had more power and resolution relative to the RFLP/SSR map. This was clearly evident by the analysis of the two main QTL for grain weight, *kgw3a* (GS3) and *kgw5* (GW5/qSW5). Using the SNP bin map, GW5/qSW5 QTL for grain width was accurately narrowed down to a 123 kb region as compared to the 12.4 Mb region based on the RFLP/SSR genetic map. Likewise, GS3 QTL for grain length was mapped to a 197 kb interval in comparison to 6 Mb region with the RFLP/SSR genetic map. Beside the power and the resolution, maps based on high-density SNP markers are also highly suitable for fine mapping and cloning of QTL and at times SNPs on these maps are also functionally associated with the natural variation in the trait. In another QTL mapping



project, SNP and InDel markers were used to fine map *qSH1* gene, a major QTL of seed shattering trait in rice [60]. The QTL were initially detected using RFLP and RAPD markers on F<sub>2</sub> plants. Using large BC<sub>4</sub>F<sub>2</sub> and BC<sub>3</sub>F<sub>2</sub> populations in fine mapping approach with SNP and InDel markers, the authors mapped the functional natural variation to a 612 bp interval between the QTL flanking markers and discovered only one SNP. They further showed that this SNP in the 5' regulatory region of the *qSH1* gene caused loss of seed shattering. Fine mapping approach was also taken to positionally clone the rice bacterial blight resistance gene *xa5*, by isolating the recombination breakpoints to a pair of SNPs followed by sequencing of the corresponding 5 kb region [61]. Several studies have shown that the SNPs and InDels are highly abundant and present throughout the genome in various species including plants [62–64]. SNP genotyping is a valuable tool for gene mapping, map-based cloning, and marker assisted selection (MAS) in crops [65]. A study was conducted to assess the feasibility of SNPs and InDels as DNA markers in genetic analysis and marker-assisted breeding in rice by analyzing these sequence polymorphisms in the genomic region containing *Piz* and *Piz-t* rice blast resistance genes and developing PCR-based SNP markers [65]. The authors discovered that SNPs were abundant in the *Piz* and *Piz-t* (averaging one SNP every 248 bp), while InDels were much lower. This dense distribution of SNPs helped in developing SNP markers in the vicinity of these genes. Advancements in rice genomics have led to mapping and cloning of several genes and QTL controlling agronomically important traits, enabled routine use of SNP markers for MAS, gene pyramiding, and marker-assisted breeding (MAB) [66–68].

**4.1.2. Examples in Maize.** SNP markers have facilitated the dissection of complex traits such as flowering time in maize. Using a set of 5000 RILs, which represent the nested association mapping (NAM) population and genotyping with 1,200 SNP markers, the authors discovered that the genetic architecture of flowering time is controlled by small additive QTL rather than a single large-effect QTL [69]. The same NAM population was used for mapping resistance to northern leaf blight disease [70]. Twenty-nine QTL were discovered and candidate genes were identified with genome-wide NAM approach using 1.6 million SNPs. Proprietary SNP markers developed by companies are being predominantly used in their private breeding programs. A study from Pioneer Hi-Bred International Inc. reported identifying a high-oil QTL (*qHO6*) affecting maize seed oil and oleic acid contents. This QTL encodes an acyl-CoA:diacylglycerol acyltransferase (*DGAT1-2*), which catalyzes the final step of oil synthesis [71].

**4.1.3. Examples in Wheat.** Recent advances in wheat genomics have led to the implementation of high-density SNP genotyping in wheat [72–75]. Gene-based SNP markers were developed for *Lr34/Yr18/Pm38* locus that confers resistance to leaf rust, stripe rust, and powdery mildew diseases [76]. These markers serve as efficient tools for MAS and MAB of

disease resistant wheat lines. Another economically important wheat disease, Fusarium head blight (FHB), has been extensively studied. Several QTL controlling FHB resistance have been identified, with the most important being *Fhb1* [77]. Recently, SNP markers were mapped between the known flanking markers for *Fhb1* [78]. These new markers would be useful for MAS and fine mapping towards cloning the *Fhb1* gene. MAS in wheat has been extensively applied for simple traits that are difficult to score [79].

**4.1.4. Examples in Soybean.** In order to improve the effectiveness of MAS and clone soybean aphid resistance gene, *Rag1*, fine mapping was done to accurately position the gene, which was previously mapped to a 12 cM interval [80]. The authors mapped the gene between two SNP markers that corresponded to a physical distance of 115 kb and identified several candidate genes. Similarly, another aphid resistance gene, *Rag2*, originally mapped to a 10 cM interval, was fine mapped to a 54 kb interval using SNP markers that were developed by resequencing of target intervals and sequence-tagged sites [81]. In another study that used a similar approach, the authors identified SNP markers tightly linked to a QTL conferring resistance to southern root-knot nematode by developing these SNP markers from the bacterial artificial chromosome (BAC) ends and SSR-containing genomic DNA clones [82]. In all of these examples the main idea behind the identification of closely linked SNP markers was to enhance the efficiency and cost effectiveness through MAS and increase the resolution within the target locus.

**4.1.5. Examples in Other Crops.** In a study conducted in canola to map the *fad2* and *fad3* gene, single nucleotide mutations were identified by sequencing the genomic clones of these genes and subsequently SNP markers were developed [83]. Allele-specific PCR assays were developed to enable direct selection of desirable *fad2* and *fad3* alleles in marker-assisted trait introgression and breeding. In barley, SNP markers were identified that were linked to a covered smut resistance gene, *Ruh.7H*, by using high-resolution melting (HRM) technique [84]. In sugar beet, an anchored linkage map based on AFLP, SNP, and RAPD markers was developed to map QTL for *Beet necrotic yellow vein virus* resistance genes, *Rz4* [85] and *Rz5* [86]. A consensus genetic map based on EST-derived SNPs was developed for cowpea that would be an important resource for genomic and QTL mapping studies in this crop [87]. In one of the post-genomic era studies in 2002, the fine mapping and map-based cloning approaches were used to clone the *VTC2* gene in Arabidopsis [88]. The authors fine mapped the gene interval from ~980 kb region to a 20 kb interval with SNP and InDel markers. Additional nine candidate genes were identified in that interval and subsequently the underlying mutation was discovered. Although only a few examples that demonstrate the application of SNP markers in QTL mapping and genomic studies have been mentioned here, several other studies have been published in this area. Recent advances in HTP genotyping technologies and sequence

information will further pave the way for rapid identification of causative variations and cloning of QTL of interest for use in MAB.

**4.2. Genome-Wide Association Study Approach.** GWAS is increasingly becoming a popular tool for dissecting complex traits in plants [89–92]. The idea behind GWAS is to genotype a large number of markers distributed across the genome so that the phenotype or the functional alleles will be in LD with one or few markers that could then be used in the breeding program. However, due to limited extent of LD, a greater number of markers are required for sufficient power to detect linkage between the marker and the underlying phenotypic variation. Several studies on association mapping in plants have been published and reviewed in the past [89, 90, 92, 93]. A few selected examples on the GWAS and candidate gene association (CGA) studies that utilized SNP markers are described below.

The successful use and first time demonstration of the power of GWAS was through the identification of a putative gene associated with a QTL in maize [94]. In that study, a single locus with major effect on oleic acid was mapped to a 4 cM genetic interval by using SNP haplotypes at 8,590 loci. The authors identified a fatty acid desaturase gene, *fad2*, at ~2 kb from one of the associated markers, and this was considered a likely causative gene. With the discovery of millions of SNPs in maize and the availability of tools such as NAM populations, GWAS was effectively applied to dissect the genetic architecture of leaf traits and it was also shown that variations at the *liguleless* genes contributed to more upright leaf phenotype [95]. Utility of the GWAS approach was demonstrated in barley through the mapping of a QTL for spot blotch disease resistance [96]. Using the diversity array technology (DArT) and SNP markers, the authors identified several QTL, some of which were not identified for this trait earlier. Another variant of the association mapping method is the CGA, where the association between one or few gene candidate loci and the trait of interest is tested. Using this approach 24 gene candidates were analyzed for association with the field resistance to late blight disease in potato and plant maturity. Nine SNPs were identified to be associated with maturity corrected resistance, explaining 50% of the genetic variance of this trait [97]. Two SNPs at the *allene oxide synthase 2* (*StAOS2*) gene locus were associated with the largest effect on the trait of interest. A GWAS approach was also successfully applied to understand the genetic architecture of complex diseases such as northern and southern corn leaf blights [70, 98]. Although the number of papers dedicated to the application of GWAS to reveal the genetic basis of agronomic traits is growing, the practical utility of minor QTL in molecular breeding is yet to be shown. As GWAS requires large number of molecular markers, the utility of GWAS in dissection of molecular basis of traits in polyploid crops such as canola, wheat, and cotton has been fairly limited due to the insufficient number of polymorphic markers and the absence of reference genome. However, recently developed associative transcriptomics method has

a potential to overcome the above-mentioned shortages [99]. Harper et al. [99] leveraged differentially expressed transcriptome sequences to develop molecular markers in tetraploid crop *Brassica napus* and associated them with glucosinolate content variation in seeds. Due to the precision of this method, scientists were able to correlate specific deletions in canola genome with two QTL controlling the trait. Annotation of deleted regions revealed the orthologs of the transcription factor HAG1, which controlled aliphatic glucosinolate biosynthesis in *A. thaliana*. This research work gives an optimism on successful application of GWAS in polyploid crops.

## 5. Implementation of SNP Markers in Plant Breeding

Due to the availability of HTP SNP detection and validation technologies, the development of SNP markers becomes a routine process, especially in crops with reference genome. How has that influenced the application of SNP markers in plant breeding? In a review article, Xu and Crouch [100] indicated fairly low number of articles dedicated to the marker assisted selection for the 1986–2005 period. The combination of three key phrases (“marker-assisted selection” AND “SNP” AND “plant breeding”), indeed, shows only 637 articles at Google Scholar for that period. However, similar search for the period, spanning 2006 through 2012, demonstrates almost sevenfold (~4,560) increase in the number of articles indicating the application of SNPs in MAS. A vast majority of those publications are from public sector and primarily describe mapping QTL using SNPs and state the potential usefulness of those markers in MAS without any experimental support for that. For most of those research studies, QTL mapping is the final destination and further application of those markers in actual MAS leading to the development of varieties seldom happens. Fairly low impact of academic research in the MAS-based variety development can be explained by the lack of funding to complete the entire marker development pipeline (MDP), which can be long term and cost intensive. MDP includes several steps such as (1) population development, (2) initial QTL mapping, (3) QTL validation (testing in several locations and years and implementing fine mapping), and (4) marker validation (development of inexpensive but HTP and automation amenable assays) [101]. Every step of the development of markers linked to QTL is associated with numerous constraints, which may take several years and substantial funding to resolve. However, since 2006, there have been a few success stories about the development of varieties using SNPs in publications derived from academic research, including the development of submergence-tolerant rice cultivars [102], rice cultivars with improved eating, cooking, and sensory quality [103], leaf rust resistant wheat variety “Patwin” [104], and maize cultivar with low phytic acid [105]. Although the private sector does not normally release details of its breeding methodologies to the public, several papers published by Monsanto [106, 107], Pioneer Hi-bred [71], Syngenta [108], and Dow AgroSciences [109] indicate

that commercial organizations are the main drivers in the application of SNP markers in MAS [110].

Current MAS strategies fit the breeding programs for the traits that are highly heritable and governed by a single gene or one major QTL that explains a large portion of the phenotypic variability. In reality, most of the agronomic traits such as yield, drought and heat tolerance, nitrogen and water use efficiency, and fiber quality in cotton have complex inheritance that is controlled by multiple QTL with minor effect. Use of one of those minor QTL in MAS will be inefficient because of its negligible effect on phenotype.

The MAS scheme using paternity testing has recently been proposed to address challenges associated with selection gains that can be achieved in outbred forage crops [111]. Paternity testing, a nonlinkage-based MAS scheme, improves selection gains by increasing parental control in the selection gain equation. The authors demonstrated paternity testing MAS in three red clover breeding populations by using permutation-based truncation selection for a biomass-persistence index trait and achieved paternity-based selection gains that were greater than double the selection gains based on maternity alone. The paternity was determined by using a small set (11) of SSR markers. SNP markers can also be used for paternity testing, but one would require a relatively larger number of SNP loci [112].

Meuwissen et al. [113] described a new methodology in plant breeding called genomic selection (GS) that was intended to solve problems related to MAS of complex traits. This methodology also applies molecular markers but in a different fashion in both diploid and polyploid crop species. Unlike MAS, in GS markers are not used for tracking a trait. In GS high-density marker coverage is needed to potentially have all QTL in LD with at least one marker. Then the comprehensive information on all possible loci, haplotypes, and marker effects across the entire genome is used to calculate genomic estimated breeding value (GEBV) of a particular line in the breeding population.

GS of superior lines can be carried out within any breeding population. In order to enable successful GS, the experimental population must be identified. The population should not be necessarily derived from bi-parental cross but must be representative of selection candidates in the breeding program to which GS will be applied [114]. The experimental population must be genotyped with a large number of markers. Taking into account the low cost of sequencing, the best choice is the GBS implementation, which will yield maximum number of polymorphisms. The sequence of the two events, that is, phenotypic and genotypic data collection, is arbitrary and can be done in parallel. When both phenotypic and genotypic data are ready, one can start “training” molecular markers [115]. In order to train the GS model, the effect of each marker is calculated computationally. The effect of a marker is represented by a number with a positive or negative sign that indicates the positive or negative effect, respectively, of a particular locus to the phenotype. When the effects of all markers are known, they are considered “trained” and ready to assess any breeding population different from the experimental one for the same trait. Availability of trained GS model does not

require the collection of phenotypic data from new breeding populations. The same set of “trained” markers will be used to genotype a new breeding population. Based on genotypic data, the known effects of each marker will be summed and GEBV of each line will be calculated. The higher the GEBV value of an individual line, the more the chances that this line will be selected and advanced in the breeding cycle. Thus, GS using high-density marker coverage has a potential to capture QTL with major and minor effects and eliminate the need to collect phenotypic data in all breeding cycles. Also, the application of GS was demonstrated to reduce the number of breeding cycles and increase the annual gain [114]. One of the problems of GS is the level of GEBV accuracy. Simulation studies based on simulated and empirical data demonstrated that GEBV accuracy could be within 0.62–0.85. Heffner et al. [114] used previously reported GEBV accuracy of 0.53 and reported three- and twofold annual gain in maize and winter barley, respectively. The obvious advantages of GS over traditional MAS have been successfully proven in animal breeding [116]. Rapid evolution of sequencing technologies and HTP SNP genotyping systems are enabling generation and validation of millions of markers, giving a “cautious optimism” for successful application of GS in breeding for complex traits [117–120].

## 6. Conclusion

SNP markers have become extremely popular in plant molecular genetics due to their genome-wide abundance and amenability for high- to ultra-high-throughput detection platforms. Unlike earlier marker systems, SNPs made it possible to create saturated, if not, supersaturated genetic maps, thereby enabling genome-wide tracking, fine mapping of target regions, rapid association of markers with a trait, and accelerated cloning of gene/QTL of interest. On the flip side, there are some challenges that need to be addressed or overcome while using SNPs. For example, the biallelic nature of SNPs needs to be compensated by discovering and using a larger number of SNPs to arrive at the same or higher power as that of earlier-generation molecular markers. This could be cost prohibitive depending on the crop and the sequence resources available for that genome. Working with polyploid crops is another challenge where useful SNPs are only a small percentage of the total available polymorphisms. Creative strategies need to be employed to generate a reasonable number of SNPs in those species. The use of SNP markers in MAB programs has been growing at a faster pace and so is the development of technologies and platforms for the discovery and HTP screening of SNPs in many crops. SNP chips are currently available for several crops; however, one disadvantage is that these readily available chips are made based on SNPs discovered from certain genotypes and, therefore, may not be ideal for projects utilizing unrelated genotypes. This necessitates creation of multiple chips or the usage of technologies that permit design flexibility but are economical. Although GBS creates great opportunities to discover a large number of SNPs at lower per sample cost within the genotypes of interest, the lack of adequate



computational capabilities such as reliable data imputation algorithms and powerful computers allowing quick processing and the storage of a large amount of sequencing data becomes a major bottleneck. Despite certain disadvantages or challenges, it is clear that SNP markers, in combination with genomics and other next-generation technologies, have been accelerating the pace and gains of plant breeding.

## Acknowledgments

The authors would like to thank Drs. Shunxue Tang and Peizhong Zheng of the Trait Genetics and Technologies Department of Dow AgroSciences (DAS) and Raghav Ram of the IP Portfolio Development Department of DAS for careful review of the paper and the DAS Seeds and Traits R & D leaders Drs. David Meyer and Steve Thompson for general support and help.

## References

- [1] J. L. Weber and P. E. May, "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction," *American Journal of Human Genetics*, vol. 44, no. 3, pp. 388–396, 1989.
- [2] R. Ophir and D. Graur, "Patterns and rates of indel evolution in processed pseudogenes from humans and murids," *Gene*, vol. 205, no. 1–2, pp. 191–202, 1997.
- [3] D. G. Wang, J. B. Fan, C. J. Siao et al., "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome," *Science*, vol. 280, no. 5366, pp. 1077–1082, 1998.
- [4] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, "Construction of a genetic linkage map in man using restriction fragment length polymorphisms," *American Journal of Human Genetics*, vol. 32, no. 3, pp. 314–331, 1980.
- [5] P. K. Gupta, R. K. Varshney, P. C. Sharma, and B. Ramesh, "Molecular markers and their applications in wheat breeding," *Plant Breeding*, vol. 118, no. 5, pp. 369–390, 1999.
- [6] R. Bernardo, "Molecular markers and selection for complex traits in plants: learning from the last 20 years," *Crop Science*, vol. 48, no. 5, pp. 1649–1664, 2008.
- [7] J. Welsh and M. McClelland, "Fingerprinting genomes using PCR with arbitrary primers," *Nucleic Acids Research*, vol. 18, no. 24, pp. 7213–7218, 1990.
- [8] P. Vos, R. Hogers, M. Bleeker et al., "AFLP: a new technique for DNA fingerprinting," *Nucleic Acids Research*, vol. 23, no. 21, pp. 4407–4414, 1995.
- [9] H. J. Jacob, K. Lindpaintner, S. E. Lincoln et al., "Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat," *Cell*, vol. 67, no. 1, pp. 213–224, 1991.
- [10] E. S. Lander and S. Botstein, "Mapping mendelian factors underlying quantitative traits using RFLP linkage maps," *Genetics*, vol. 121, no. 1, p. 185, 1989.
- [11] J. G. K. Williams, A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey, "DNA polymorphisms amplified by arbitrary primers are useful as genetic markers," *Nucleic Acids Research*, vol. 18, no. 22, pp. 6531–6535, 1990.
- [12] Z. Zhang, X. Guo, B. Liu, L. Tang, and F. Chen, "Genetic diversity and genetic relationship of *Jatropha curcas* between China and Southeast Asian revealed by amplified fragment length polymorphisms," *African Journal of Biotechnology*, vol. 10, no. 15, pp. 2825–2832, 2011.
- [13] W. Powell, G. C. Machray, and J. Proven, "Polymorphism revealed by simple sequence repeats," *Trends in Plant Science*, vol. 1, no. 7, pp. 215–222, 1996.
- [14] S. Ghosh, P. Malhotra, P. V. Lalitha, S. Guha-Mukherjee, and V. S. Chauhan, "Novel genetic mapping tools in plants: SNPs and LD-based approaches," *Plant Science*, vol. 162, no. 3, pp. 329–333, 2002.
- [15] M. W. Ganal, T. Altmann, and M. S. Röder, "SNP identification in crop plants," *Current Opinion in Plant Biology*, vol. 12, no. 2, pp. 211–217, 2009.
- [16] B. C. Meyers, S. V. Tingey, and M. Morgante, "Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome," *Genome Research*, vol. 11, no. 10, pp. 1660–1676, 2001.
- [17] S. I. Wright, I. V. Bi, S. C. Schroeder et al., "Evolution: the effects of artificial selection on the maize genome," *Science*, vol. 308, no. 5726, pp. 1310–1314, 2005.
- [18] J. Batley, G. Barker, H. O'Sullivan, K. J. Edwards, and D. Edwards, "Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data," *Plant Physiology*, vol. 132, no. 1, pp. 84–91, 2003.
- [19] A. Pratap, S. Gupta, J. Kumar, and R. Solanki, "Soybean," *Technological Innovations in Major World Oil Crops*, vol. 1, pp. 293–321, 2012.
- [20] I. Y. Choi, D. L. Hyten, L. K. Matukumalli et al., "A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis," *Genetics*, vol. 176, no. 1, pp. 685–696, 2007.
- [21] E. R. Mardis, "The impact of next-generation sequencing technology on genetics," *Trends in Genetics*, vol. 24, no. 3, pp. 133–141, 2008.
- [22] O. Morozova and M. A. Marra, "Applications of next-generation sequencing technologies in functional genomics," *Genomics*, vol. 92, no. 5, pp. 255–264, 2008.
- [23] W. B. Barbazuk, S. J. Emrich, H. D. Chen, L. Li, and P. S. Schnable, "SNP discovery via 454 transcriptome sequencing," *The Plant Journal*, vol. 51, no. 5, pp. 910–918, 2007.
- [24] M. Trick, Y. Long, J. Meng, and I. Bancroft, "Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing," *Plant Biotechnology Journal*, vol. 7, no. 4, pp. 334–346, 2009.
- [25] E. Novaes, D. R. Drost, W. G. Farmerie et al., "High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome," *BMC Genomics*, vol. 9, article 312, 2008.
- [26] P. C. Bundock, F. G. Elliott, G. Ablett et al., "Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing," *Plant Biotechnology Journal*, vol. 7, no. 4, pp. 347–354, 2009.
- [27] T. L. Parchman, K. S. Geist, J. A. Grahnen, C. W. Benkman, and C. A. Buerkle, "Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery," *BMC Genomics*, vol. 11, no. 1, article 180, 2010.
- [28] K. Lai, C. Duran, P. J. Berkman et al., "Single nucleotide polymorphism discovery from wheat next-generation sequence data," *Plant Biotechnology Journal*, vol. 10, no. 6, pp. 743–749, 2012.
- [29] D. Kuhn, "Design of an Illumina Infinium 6k SNPchip for genotyping two large avocado mapping populations," in *Proceedings of the 20th Conference on Plant and Animal Genome*, San Diego, CA, January 2012.



- [30] J. R. Russell, M. Bayer, C. Booth et al., "Identification, utilisation and mapping of novel transcriptome-based markers from blackcurrant (*Ribes nigrum*)," *BMC Plant Biology*, vol. 11, article 147, 2011.
- [31] E. Hodges, Z. Xuan, V. Balija et al., "Genome-wide in situ exon capture for selective resequencing," *Nature Genetics*, vol. 39, no. 12, pp. 1522–1527, 2007.
- [32] N. M. Springer, K. Ying, Y. Fu et al., "Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content," *PLoS Genetics*, vol. 5, no. 11, Article ID e1000734, 2009.
- [33] R. K. Varshney, "Gene-based marker systems in plants: high throughput approaches for marker discovery and genotyping," in *Molecular Techniques in Crop Improvement*, S. M. Jain and D. S. Brar, Eds., pp. 119–142, 2009.
- [34] A. Dean, "On a chromosome far, far away: LCRs and gene expression," *Trends in Genetics*, vol. 22, no. 1, pp. 38–45, 2006.
- [35] Y. Yuan, P. J. SanMiguel, and J. L. Bennetzen, "High-Cot sequence analysis of the maize genome," *The Plant Journal*, vol. 34, no. 2, pp. 249–255, 2003.
- [36] J. Emberton, J. Ma, Y. Yuan, P. SanMiguel, and J. L. Bennetzen, "Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries," *Genome Research*, vol. 15, no. 10, pp. 1441–1446, 2005.
- [37] D. T. Okou, K. M. Steinberg, C. Middle, D. J. Cutler, T. J. Albert, and M. E. Zwick, "Microarray-based genomic selection for high-throughput resequencing," *Nature Methods*, vol. 4, no. 11, pp. 907–909, 2007.
- [38] N. J. van Orsouw, R. C. J. Hogers, A. Janssen et al., "Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes," *PLoS ONE*, vol. 2, no. 11, Article ID e1172, 2007.
- [39] N. A. Baird, P. D. Etter, T. S. Atwood et al., "Rapid SNP discovery and genetic mapping using sequenced RAD markers," *PLoS ONE*, vol. 3, no. 10, Article ID e3376, 2008.
- [40] J. A. Mammadov, W. Chen, R. Ren et al., "Development of highly polymorphic SNP markers from the complexity reduced portion of maize (*Zea mays* L.) genome for use in marker-assisted breeding," *Theoretical and Applied Genetics*, vol. 121, no. 3, pp. 577–588, 2010.
- [41] Y. Chutimanitsakun, R. W. Nipper, A. Cuesta-Marcos et al., "Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley," *BMC Genomics*, vol. 12, article 4, 2011.
- [42] H. Yu, W. Xie, J. Wang et al., "Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers," *PLoS ONE*, vol. 6, no. 3, Article ID e17595, 2011.
- [43] A. Bus, J. Hecht, B. Huettel, R. Reinhardt, and B. Stich, "High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing," *BMC Genomics*, vol. 13, no. 1, p. 281, 2012.
- [44] J. Tang, J. A. M. Leunissen, R. E. Voorrips, C. G. van der Linden, and B. Vosman, "HaploSNPer: a web-based allele and SNP detection tool," *BMC Genetics*, vol. 9, article 23, 2008.
- [45] A. Narechania, M. A. Gore, E. S. Buckler et al., "Large-scale discovery of gene-enriched SNPs," *The Plant Genome*, vol. 2, no. 2, pp. 121–133, 2009.
- [46] J. C. Nelson, S. Wang, Y. Wu et al., "Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum," *BMC Genomics*, vol. 12, article 352, 2011.
- [47] R. J. Elshire, J. C. Glaubitz, Q. Sun et al., "A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species," *PLoS ONE*, vol. 6, no. 5, Article ID e19379, 2011.
- [48] S. R. Browning and B. L. Browning, "Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering," *American Journal of Human Genetics*, vol. 81, no. 5, pp. 1084–1097, 2007.
- [49] B. N. Howie, P. Donnelly, and J. Marchini, "A flexible and accurate genotype imputation method for the next generation of genome-wide association studies," *PLoS Genetics*, vol. 5, no. 6, Article ID e1000529, 2009.
- [50] X. Huang, X. Wei, T. Sang et al., "Genome-wide association studies of 14 agronomic traits in rice landraces," *Nature Genetics*, vol. 42, no. 11, pp. 961–967, 2010.
- [51] J. Marchini and B. Howie, "Genotype imputation for genome-wide association studies," *Nature Reviews Genetics*, vol. 11, no. 7, pp. 499–511, 2010.
- [52] J. B. Fan, A. Oliphant, R. Shen et al., "Highly parallel SNP genotyping," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 68, pp. 69–78, 2003.
- [53] F. J. Steemers and K. L. Gunderson, "Whole genome genotyping technologies on the BeadArray platform," *Biotechnology Journal*, vol. 2, no. 1, pp. 41–49, 2007.
- [54] K. J. Livak, S. J. A. Flood, J. Marmaro, W. Giusti, and K. Deetz, "Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization," *Genome Research*, vol. 4, no. 6, pp. 357–362, 1995.
- [55] S. P. Kumpatla, R. Buyyarapu, I. Y. Abdurakhmonov, and J. A. Mammadov, "Genomics-assisted plant breeding in the 21st century: technological advances and progress," in *Plant Breeding*, I. Y. Abdurakhmonov, Ed., pp. 131–184.
- [56] R. Buyyarapu, R. Ren, S. Kumpatla et al., "In silico discovery and validation of SNP markers for molecular breeding in cotton," in *Proceedings of the 19th Conference on Plant & Animal Genome*, San Diego, Calif, USA, January 2011.
- [57] S. D. Tanksley, "Mapping polygenes," *Annual Review of Genetics*, vol. 27, pp. 205–233, 1993.
- [58] D. Bhatramakki, M. Dolan, M. Hanafey et al., "Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers," *Plant Molecular Biology*, vol. 48, no. 5-6, pp. 539–547, 2002.
- [59] E. S. Jones, H. Sullivan, D. Bhatramakki, and J. S. C. Smith, "A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.)," *Theoretical and Applied Genetics*, vol. 115, no. 3, pp. 361–371, 2007.
- [60] S. Konishi, T. Izawa, S. Y. Lin et al., "An SNP caused loss of seed shattering during rice domestication," *Science*, vol. 312, no. 5778, pp. 1392–1396, 2006.
- [61] A. S. Iyer and S. R. McCouch, "The rice bacterial blight resistance gene xa5 encodes a novel form of disease resistance," *Molecular Plant-Microbe Interactions*, vol. 17, no. 12, pp. 1348–1354, 2004.
- [62] E. Drenkard, B. G. Richter, S. Rozen et al., "A simple procedure for the analysis of single nucleotide polymorphism facilitates map-based cloning in Arabidopsis," *Plant Physiology*, vol. 124, no. 4, pp. 1483–1492, 2000.
- [63] K. Garg, P. Green, and D. A. Nickerson, "Identification of candidate coding region single nucleotide polymorphisms in

- 165 human genes using assembled expressed sequence tags," *Genome Research*, vol. 9, no. 11, pp. 1087–1092, 1999.
- [64] S. Nasu, J. Suzuki, R. Ohta et al., "Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers," *DNA Research*, vol. 9, no. 5, pp. 163–171, 2002.
- [65] K. Hayashi, N. Hashimoto, M. Daigen, and I. Ashikawa, "Development of PCR-based SNP markers for rice blast resistance genes at the *Piz* locus," *Theoretical and Applied Genetics*, vol. 108, no. 7, pp. 1212–1220, 2004.
- [66] M. Ashikari and M. Matsuoka, "Identification, isolation and pyramiding of quantitative trait loci for rice breeding," *Trends in Plant Science*, vol. 11, no. 7, pp. 344–350, 2006.
- [67] K. K. Jena and D. J. Mackill, "Molecular markers and their use in marker-assisted selection in rice," *Crop Science*, vol. 48, no. 4, pp. 1266–1276, 2008.
- [68] R. K. Varshney, D. A. Hoisington, and A. K. Tyagi, "Advances in cereal genomics and applications in crop breeding," *Trends in Biotechnology*, vol. 24, no. 11, pp. 490–499, 2006.
- [69] E. S. Buckler, J. B. Holland, P. J. Bradbury et al., "The genetic architecture of maize flowering time," *Science*, vol. 325, no. 5941, pp. 714–718, 2009.
- [70] J. A. Poland, P. J. Bradbury, E. S. Buckler, and R. J. Nelson, "Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 17, pp. 6893–6898, 2011.
- [71] P. Zheng, W. B. Allen, K. Roesler et al., "A phenylalanine in DGAT is a key determinant of oil content and composition in maize," *Nature Genetics*, vol. 40, no. 3, pp. 367–372, 2008.
- [72] E. Akhunov, C. Nicolet, and J. Dvorak, "Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay," *Theoretical and Applied Genetics*, vol. 119, no. 3, pp. 507–517, 2009.
- [73] A. M. Allen, G. L. Barker, S. T. Berry et al., "Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.)," *Plant Biotechnology Journal*, vol. 9, no. 9, pp. 1086–1099, 2011.
- [74] A. Bérard, M. C. Le Paslier, M. Dardevet et al., "High-throughput single nucleotide polymorphism genotyping in wheat (*Triticum* spp.)," *Plant Biotechnology Journal*, vol. 7, no. 4, pp. 364–374, 2009.
- [75] M. O. Winfield, P. A. Wilkinson, A. M. Allen et al., "Targeted re-sequencing of the allohexaploid wheat exome," *Plant Biotechnology Journal*, vol. 10, no. 6, pp. 733–742, 2012.
- [76] E. S. Lagudah, S. G. Krattinger, S. Herrera-Foessel et al., "Gene-specific markers for the wheat gene *Lr34/Yr18/Pm38* which confers resistance to multiple fungal pathogens," *Theoretical and Applied Genetics*, vol. 119, no. 5, pp. 889–898, 2009.
- [77] H. Buerstmayr, T. Ban, and J. A. Anderson, "QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review," *Plant Breeding*, vol. 128, no. 1, pp. 1–26, 2009.
- [78] A. N. Bernardo, H. Ma, D. Zhang, and G. Bai, "Single nucleotide polymorphism in wheat chromosome region harboring *Fhb1* for Fusarium head blight resistance," *Molecular Breeding*, vol. 29, no. 2, pp. 477–488, 2012.
- [79] P. K. Gupta, P. Langridge, and R. R. Mir, "Marker-assisted wheat breeding: present status and future possibilities," *Molecular Breeding*, vol. 26, no. 2, pp. 145–161, 2010.
- [80] K. S. Kim, S. Bellendir, K. A. Hudson et al., "Fine mapping the soybean aphid resistance gene *Rag1* in soybean," *Theoretical and Applied Genetics*, vol. 120, no. 5, pp. 1063–1071, 2010.
- [81] K. S. Kim, C. B. Hill, G. L. Hartman, D. L. Hyten, M. E. Hudson, and B. W. Diers, "Fine mapping of the soybean aphid-resistance gene *Rag2* in soybean PI 200538," *Theoretical and Applied Genetics*, vol. 121, no. 3, pp. 599–610, 2010.
- [82] B. K. Ha, R. S. Hussey, and H. R. Boerma, "Development of SNP assays for marker-assisted selection of two southern root-knot nematode resistance QTL in soybean," *Crop Science*, vol. 47, no. 2, pp. S73–S82, 2007.
- [83] X. Hu, M. Sullivan-Gilbert, M. Gupta, and S. A. Thompson, "Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.)," *Theoretical and Applied Genetics*, vol. 113, no. 3, pp. 497–507, 2006.
- [84] A. Lehmsiek, M. W. Sutherland, and R. B. McNamara, "The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley," *Theoretical and Applied Genetics*, vol. 117, no. 5, pp. 721–728, 2008.
- [85] M. K. Grimmer, S. Trybush, S. Hanley, S. A. Francis, A. Karp, and M. J. C. Asher, "An anchored linkage map for sugar beet based on AFLP, SNP and RAPD markers and QTL mapping of a new source of resistance to Beet necrotic yellow vein virus," *Theoretical and Applied Genetics*, vol. 114, no. 7, pp. 1151–1160, 2007.
- [86] M. K. Grimmer, T. Kraft, S. A. Francis, and M. J. C. Asher, "QTL mapping of BNYVV resistance from the WB258 source in sugar beet," *Plant Breeding*, vol. 127, no. 6, pp. 650–652, 2008.
- [87] W. Muchero, N. N. Diop, P. R. Bhat et al., "A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 43, pp. 18159–18164, 2009.
- [88] G. Jander, S. R. Norris, S. D. Rounsley, D. F. Bush, I. M. Levin, and R. L. Last, "Arabidopsis map-based cloning in the post-genome era," *Plant Physiology*, vol. 129, no. 2, pp. 440–450, 2002.
- [89] I. Y. Abdurakhmonov and A. Abdurakimov, "Application of association mapping to understanding the genetic diversity of plant germplasm resources," *International Journal of Plant Genomics*, vol. 2008, Article ID 574927, 2008.
- [90] D. Hall, C. Tegström, and P. K. Ingvarsson, "Using association mapping to dissect the genetic basis of complex traits in plants," *Briefings in Functional Genomics and Proteomics*, vol. 9, no. 2, pp. 157–165, 2010.
- [91] S. Myles, J. Peiffer, P. J. Brown et al., "Association mapping: critical considerations shift from genotyping to experimental design," *Plant Cell*, vol. 21, no. 8, pp. 2194–2202, 2009.
- [92] M. Gore, E. S. Buckler, J. Yu, and C. Zhu, "Status and prospects of association mapping in plants," *The Plant Genome*, vol. 1, no. 1, pp. 5–20, 2008.
- [93] J. A. Rafalski, "Association genetics in crop improvement," *Current Opinion in Plant Biology*, vol. 13, no. 2, pp. 174–180, 2010.
- [94] A. Beló, P. Zheng, S. Luck et al., "Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize," *Molecular Genetics and Genomics*, vol. 279, no. 1, pp. 1–10, 2008.
- [95] F. Tian, P. J. Bradbury, P. J. Brown et al., "Genome-wide association study of leaf architecture in the maize nested

- association mapping population,” *Nature Genetics*, vol. 43, no. 2, pp. 159–162, 2011.
- [96] J. K. Roy, K. P. Smith, G. J. Muehlbauer, S. Chao, T. J. Close, and B. J. Steffenson, “Association mapping of spot blotch resistance in wild barley,” *Molecular Breeding*, vol. 26, no. 2, pp. 243–256, 2010.
- [97] K. Pajerowska-Mukhtar, B. Stich, U. Achenbach et al., “Single nucleotide polymorphisms in the Allene Oxide Synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars,” *Genetics*, vol. 181, no. 3, pp. 1115–1127, 2009.
- [98] K. L. Kump, P. J. Bradbury, R. J. Wissner et al., “Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population,” *Nature Genetics*, vol. 43, no. 2, pp. 163–168, 2011.
- [99] A. L. Harper, M. Trick, J. Higgins et al., “Associative transcriptomics of traits in the polyploid crop species *Brassica napus*,” *Nature Biotechnology*, vol. 30, no. 8, pp. 798–802, 2012.
- [100] Y. Xu and J. H. Crouch, “Marker-assisted selection in plant breeding: from publications to practice,” *Crop Science*, vol. 48, no. 2, pp. 391–407, 2008.
- [101] B. C. Y. Collard and D. J. Mackill, “Marker-assisted selection: an approach for precision plant breeding in the twenty-first century,” *Philosophical Transactions of the Royal Society B*, vol. 363, no. 1491, pp. 557–572, 2008.
- [102] E. M. Septiningsih, A. M. Pamplona, D. L. Sanchez et al., “Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond,” *Annals of Botany*, vol. 103, no. 2, pp. 151–160, 2009.
- [103] L. Jin, Y. Lu, Y. Shao et al., “Molecular marker assisted selection for improvement of the eating, cooking and sensory quality of rice (*Oryza sativa* L.),” *Journal of Cereal Science*, vol. 51, no. 1, pp. 159–164, 2010.
- [104] M. Asif, T. Shaheen, N. Tabbasam, Y. Zafar, and A. H. Paterson, “Marker-assisted breeding in higher plants,” *Alternative Farming Systems, Biotechnology, Drought Stress and Ecological Fertilisation*, vol. 6, pp. 39–76, 2011.
- [105] R. Naidoo, G. M. F. Watson, J. Derera, P. Tongoon, and M. Laing, “Marker-assisted selection for low phytic acid (*lpa1-1*) with single nucleotide polymorphism marker and amplified fragment length polymorphisms for background selection in a maize backcross breeding programme,” *Molecular Breeding*, vol. 30, pp. 1207–1217, 2012.
- [106] S. R. Eathington, T. M. Crosbie, M. D. Edwards, R. S. Reiter, and J. K. Bull, “Molecular markers in a commercial breeding program,” *Crop Science*, vol. 47, supplement 3, pp. S154–S163, 2007.
- [107] M. L. Rosso, S. A. Burleson, L. M. Maupin, and K. M. Rainey, “Development of breeder-friendly markers for selection of *MIPS1* mutations in soybean,” *Molecular Breeding*, vol. 28, no. 1, pp. 127–132, 2011.
- [108] J. M. Ribaut and M. Ragot, “Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives,” *Journal of Experimental Botany*, vol. 58, no. 2, pp. 351–360, 2007.
- [109] R. Ren, B. A. Nagel, S. P. Kumpatla et al., “Maize Cytoplasmic Male Sterility (Cms) C-Type Restorer Rf4 Gene, Molecular Markers And Their Use,” Google Patents, 2011.
- [110] M. Ragot, M. Lee, E. Guimarães et al., “Marker-assisted selection in maize: current status, potential, limitations and perspectives from the private and public sectors,” *Marker-Assisted Selection, Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*, pp. 117–150, 2007.
- [111] H. Riday, “Paternity testing: a non-linkage based marker-assisted selection scheme for outbred forage species,” *Crop Science*, vol. 51, no. 2, pp. 631–641, 2011.
- [112] D. W. Gjertson, C. H. Brenner, M. P. Baur et al., “ISFG: recommendations on biostatistics in paternity testing,” *Forensic Science International*, vol. 1, no. 3–4, pp. 223–231, 2007.
- [113] T. H. E. Meuwissen, B. J. Hayes, and M. E. Goddard, “Prediction of total genetic value using genome-wide dense marker maps,” *Genetics*, vol. 157, no. 4, pp. 1819–1829, 2001.
- [114] E. L. Heffner, M. E. Sorrells, and J. L. Jannink, “Genomic selection for crop improvement,” *Crop Science*, vol. 49, no. 1, pp. 1–12, 2009.
- [115] Z. Shengqiang, J. C. M. Dekkers, R. L. Fernando, and J. L. Jannink, “Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study,” *Genetics*, vol. 182, no. 1, pp. 355–364, 2009.
- [116] B. Hayes and M. Goddard, “Genome-wide association and genomic selection in animal breeding,” *Genome*, vol. 53, no. 11, pp. 876–883, 2010.
- [117] J. L. Jannink, A. J. Lorenz, and H. Iwata, “Genomic selection in plant breeding: from theory to practice,” *Briefings in Functional Genomics and Proteomics*, vol. 9, no. 2, pp. 166–177, 2010.
- [118] A. M. Mastrangelo, E. Mazzucotelli, D. Guerra, P. Vita, and L. Cattivelli, “Improvement of drought resistance in crops: from conventional breeding to genomic selection,” *Crop Stress and Its Management*, pp. 225–259, 2012.
- [119] M. D. V. Resende, M. F. R. Resende Jr., C. P. Sansaloni et al., “Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees,” *New Phytologist*, vol. 194, no. 1, pp. 116–128, 2012.
- [120] Y. Zhao, M. Gowda, W. Liu et al., “Accuracy of genomic selection in European maize elite breeding populations,” *Theoretical and Applied Genetics*, vol. 124, no. 4, pp. 769–776, 2012.



