10 Future Prospects of Molecular Markers in Plants

Reyazul R. Mir and Rajeev K. Varshney

Contents

Molecular Markers in Plants, First Edition. Edited by Robert J. Henry.

Abstract

The last ~30 years have witnessed a continuous evolution of new molecular marker systems from restriction fragment length polymorphisms, random amplified polymorphic DNAs, and amplified fragment length polymorphisms to present-day popular marker systems such as simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs), and diversity array technologies. Advent of low-cost and high-throughput sequencing technologies, commonly called next-generation sequencing (NGS) technologies have increased the speed of SSR and SNP discovery. NGS technologies in combination with restriction enzymes are now ready for detecting genome-wide polymorphism and new marker systems like RAD-tag sequencing, genotyping by sequencing are becoming popular. It seems that NGS-based marker systems will be dominating marker systems in future. These new emerging marker systems are expected to facilitate enhanced adoption of modern genetics and breeding approaches like genome-wide association studies and genome-wide selection that generally require markers at high-density in crop plants.

Keywords

Molecular markers, Next-generation sequencing technologies, Genome-wide association studies (GWAS), Genome-wide selection (GWS)

Introduction

Molecular markers refer to assays that allow detection of specific sequence differences between two or more individuals of an organism (Langridge and Chalmers, 2004). DNA-based molecular markers have revolutionized the genetics and molecular breeding of crops plants. They provide most powerful diagnostic tools for the detection of polymorphism at the level of specific loci and at the whole-genome level. During the last ~30 years, new molecular marker systems continuously evolved from low-throughput restriction fragment length polymorphisms (RFLPs) in 1980s to high-throughput array-based markers in 2000s and now sequencing-based marker systems in 2010s (Figure 10.1). This continuous evolution of molecular marker technologies was mostly attributed to (i) different needs of researchers working on plant genetics, genomics, and molecular breeding (for instance, the challenge of simultaneous whole-genome screening rather than screening for a single locus at a time); (ii) desire to cut down the cost of molecular marker genotyping to make their use routine in tracking loci and genomic regions in molecular breeding programs for crop

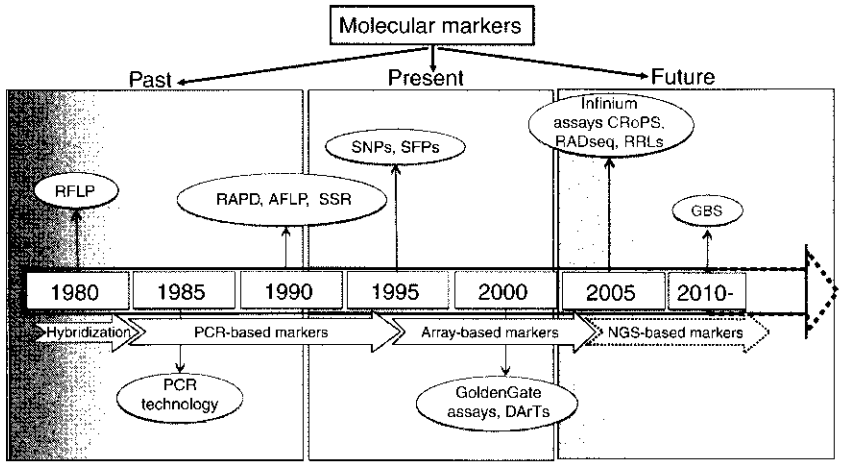

Figure 10.1 Paradigm shift in marker discovery: from hybridization-based RFLPs to NGS-based high-throughput markers. Markers have been classified into past, present, and future molecular markers. Markers highlighted with blue color are micro-array-based markers, while as those highlighted with red color are NGS-based markers. (For a color version of this figure, see color plate section.)

improvement; and (iii) continuous evolution in automation, robotics, and nanotechnology.

The improvements in screening techniques by molecular markers have been found important in facilitating the tracking of agronomically important genes (Langridge and Chalmers, 2004). However, the ultimate approach of study of polymorphism in any crop would be to sequence/resequence the entire genome (or a part of it) of a large number of accessions. This was unimaginable during 1980s and is even still not very cost-effective. Therefore, DNA-based molecular markers including RFLPs, random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), and microsatellite or simple sequence repeats (SSRs) were employed in the past for detecting and utilization genetic variation (Collard *et al.*, 2005; Gupta *et al.*, 2002; Gupta, Rustgi, and Mir, 2008). These molecular markers were developed from random genomic DNA libraries (RFLPs and SSRs), random PCR amplification of genomic DNA (RAPDs), or both (AFLP). Among these markers, SSR markers have become the markers of choice because of their various desirable attributes (Gupta and Varshney, 2000). Although, in recent years, single nucleotide polymorphisms (SNPs), whose discovery required sequence information, also became the popular markers (in addition to SSRs) due to their abundance and uniform distribution throughout the genomes. Advent of next-generation sequencing (NGS) technologies are making SNP discovery cheaper (Varshney, Graner, and Sorrells, 2005). These SNPs, once identified, can be assayed using low-, high-, or ultra-high-throughput genotyping platforms depending on the need of researchers (Deschamps and Campbell, 2009; Varshney *et al.*, 2009).

In addition to SNPs, some other recently emerged novel array-based marker systems, including diversity array technology (DArT) and single feature polymorphisms (SFPs), have also been developed, where array-based assays have been utilized to provide for the desired ultra-high-throughput and low cost. The development of high-throughput array-based markers (e.g., DArT) overcomes the problem of expensive and laborious scoring of marker panels across target populations in gel-based marker systems. SFPs, on the other hand, have been found very useful for detecting the functional polymorphism associated with traits.

Above-mentioned molecular markers have been extensively used in different areas of plant genetics research and breeding applications, for example, genetic characterization/DNA fingerprinting, genetic integrity, genetic mapping, trait mapping, marker-assisted selection (MAS), and molecular breeding. However, in recent years, some modern genetics and breeding approaches like genome-wide association studies (GWAS), genome-wide selection (GWS), and so on that generally need genome-wide or high-throughput marker screening of large populations have become available. This chapter deals with a brief discussion on molecular markers that were and are being used extensively in the past and present and then provides an overview on the emerging marker genotyping technologies such as genotyping by sequencing (GBS). In addition, two modern breeding approaches namely GWS and GWAS have also been discussed.

Molecular Markers: The Past

Protein-based marker systems including isoenzymes are the molecular markers that were in wide use long before DNA markers became popular (Market and Moller, 1959; Bernatzky and Tanksley, 1986; Glaszmann, 1987; Ishikawa *et al.*, 1992). These markers require protein extraction, which is labor intensive, not adapted to automation, and high-throughput analysis for plant molecular breeding (McMillin, Allan, and Roberts, 1986; Winzeler, Winzeler, and Keller, 1995). Therefore, these markers were soon replaced by DNA-based molecular markers such as RFLPs (Botstein *et al.*, 1980; Tanksley *et al.*, 1989), AFLPs (Vos *et al.*, 1995), RAPDs (Williams *et al.*, 1990), and microsatellites or SSRs (Litt and Lutty, 1989).

RFLP markers represent one of the earliest types of DNA-based molecular marker systems detecting variation in restriction fragment length (Botstein *et al.*, 1980). The sequence variation detected by RFLPs can be either due to single nucleotide changes leading to the creation or removal of recognition site of a restriction endonuclease or due to insertions or deletions of several nucleotides that leads to detectable shift in fragment size. However, due to the time-consuming multistep protocol and the requirement of radioactivity for detection, RFLPs lost their importance in plant molecular breeding (Mohler and Schwarz, 2005).

With the discovery of PCR technology, a universal marker technology called RAPDs making use of single random oligonucleotide primer (~10-bp long) were discovered in 1990 for the simultaneous amplification of several discrete DNA fragments (Williams *et al.*, 1990). The advantages of RAPDs being the use of universal primers, thus enabling the cost-effective accomplishment of various genetic analyses in a short period of time (see Tingey and del Tufo, 1993). Subsequently, by taking the advantage of features of RFLP together with PCR technology, a multilocus fingerprinting technology called AFLPs was discovered that could be applied to DNA of any origin or complexity (Ehrlich, Gelfand, and Sninsky, 1991; Vos *et al.*, 1995). AFLP has been extensively used for detection of polymorphism at a larger number of loci simultaneously in germplasm collection, construction of high-density genetic map, as well as trait mapping in several studies (Becker *et al.*, 1995; Mackill *et al.*, 1996; Powell *et al.*, 1996; Mohler and Schwarz, 2005).

While RFLP, AFLP, and RAPD markers have also been used for trait mapping, it is not straightforward to use the associated marker/fragment in molecular breeding applications. To overcome some of these problems, the fragments of RFLP, AFLP, or RAPD marker systems associated with the trait were also converted into simple and robust PCR-based markers, for example, sequence tagged site (Olsen *et al.*, 1989) or sequence characterized amplified regions (Paran and Michelmore, 1993).

In addition, several variants were derived from the above marker systems. Some of these include allele-specific-associated primers, single-strand conformation polymorphism, arbitrarily primed PCR (Welsh and McClelland, 1990), DNA amplification fingerprinting (Caetano-Anolles, Bassam, and Gresshoff, 1991), and selectively amplified microsatellite polymorphic locus (Morgante and Vogel, 1994).

Molecular Markers: The Present

A number of molecular markers that became available for plant genotyping have its own advantages and disadvantages. No single marker system seems to be adequate to address all the concerns in plant genome analysis. Therefore, a choice in terms of selection of molecular marker has to be made by a researcher to fulfill his/her research aim because each marker type differs from each other in terms of information content, number of scorable polymorphic loci per reaction, and degree of automation. In addition, the choice of method often depends on the genetic resolution needed as well as on technological and financial constraints (Langridge and Chalmers, 2004; Mohler and Schwarz, 2005).

Some of the molecular markers that have been discovered in recent past but are still indispensable include SSRs, SNPs, and micro-array-based markers like SFPs, DArT markers, and NGS-based high-throughput

markers (Gupta *et al.*, 1999; Gupta and Varshney, 2000; Langridge and Chalmers, 2004; Mohler and Schwarz, 2005; Gupta, Rustgi, and Mir, 2008; Varshney *et al.*, 2009). Among these molecular markers, SSRs were considered as the markers of choice because of their abundance in eukaryotic and prokaryotic genomes, reproducibility, transferability, etc., and are still being considered indispensable owing to their tremendous potential in tracking genes in molecular breeding programs (Gupta and Varshney, 2000; Mohler and Schwarz, 2005). However, if genome-wide high-throughput markers are required to address a problem, then SNPs provide the solution. In addition, SNP markers possess several other desirable attributes including (i) their abundance, (ii) uniform distribution throughout the genomes, (iii) highest resolution to create haplotypes, (iv) study the association of heritable traits with underlying genetic variation, and (v) their stability over generations. Therefore, SNP markers became the markers of choice for whole-genome analysis or complex trait mapping (Deschamps and Campbell, 2009).

Several other markers making use of micro-arrays including DArTs and SFPs are also becoming popular at present (Gupta, Rustgi, and Mir, 2008). DArT markers have been proven useful in many plant species with a limited expense in terms of time and money for a variety of applications such as genetic diversity, population structure, construction of high-density maps, and quantitative trait locus (QTL) mapping (Jaccoud *et al.*, 2001; Wenzl, Carling, and Kudrna, 2004, 2007; Kilian *et al.*, 2005; Gupta, Rustgi, and Mir, 2008). SFPs on the other hand also became available for all the major crops including barley (Cui *et al.*, 2005; Rostoks *et al.*, 2005), rice (Kumar *et al.*, 2007), maize (Kirst *et al.*, 2006; Gore *et al.*, 2007), wheat (Banks, Jordan, and Somers, 2009), and pigeonpea (Saxena *et al.*, 2011) and have been used for (i) genetic mapping (Banks, Jordan, and Somers, 2009) and (ii) QTL interval mapping and association mapping leading to detection of main effect QTLs and eQTLs (Potokina *et al.*, 2008; Kim *et al.*, 2006, 2009). However, the problem with array-based markers is that they are specific to particular population in which they are developed; therefore, genotyping of new populations will be biased toward alleles present in the original survey, which can be a serious problem in studies involving wild or highly divergent populations (Davey *et al.*, 2011).

While, at present, SSR markers have become available for almost all major crop plant species, SNP markers have already been developed in several plant species (Cho *et al.*, 1999; Rafalski, 2002; Zhao *et al.*, 2006; Jones *et al.*, 2007, 2009; Yu *et al.*, 2008). SNP discovery, although was expensive earlier because of involvement of Sanger sequencing, has become cost-effective with the use of NGS technologies (Varshney *et al.*, 2009; Metzker, 2010). The NGS technologies that have recently become available commercially included Roche/454 (http://www.454.com/), Solexa/Illumina (http://www.illumina.com/), and SOLiD/ABI

(http://www.appliedbiosystems.com/). These technologies are also referred as second-generation sequencing (SGS) technologies. Very recently, some other advanced sequencing technologies that are referred as third- or next-next-generation sequencing technologies (TGS or NNGS) are also ready to become commercially available. Some of these technologies include (i) HeliScope Single Molecule Sequencing by Helicos BioSciences (www.helicosbio.com), (ii) Single Molecule Real Time (SMRT™) Sequencing technology by Pacific Biosciences (www.pacificbio sciences.com), (iii) Ion Personal Genome Machine (PGM™) Sequencing (http://www.iontorrent.com/), and (iv) Polonator G.007 Sequencing (http://www.polonator.org/).

These NGS technologies have been found valuable for the discovery, validation, and assessment of genetic markers in populations (Varshney *et al.*, 2009; Davey *et al.*, 2011). These technologies have been used for SNP discovery in both types of species where reference genome is available as well as not available. In case reference genome sequence is available (e.g., in *Arabidopsis*, Medicago, maize, rice, poplar, grapevine, and papaya), resequencing of some genotypes/lines is followed by mapping of short sequence reads or tags on the reference genome. The alignment of short reads to a reference sequence allows the discovery of different types of sequence variations, including SNPs, short insertion/deletions (indels), and copy number variants. For instance, in case of maize, resequencing of six elite maize inbred lines leads to the discovery of 1 000 000 SNPs and 30 000 indels (Lai *et al.*, 2010). However, in case reference genome is not available, many crop communities have developed transcriptome assembly (TA) and the generated transcript reads from a number of individuals have been aligned with TA for identification of SNPs. This approach has been used in many plant species such as chickpea (Garg *et al.*, 2011; Hiremath *et al.*, 2011), pigeonpea (Dubey *et al.*, 2011), carrot (Iorizzo *et al.*, 2011), and lentil (Kaur *et al.*, 2001). Once these SNPs are identified, there is a range of genotyping platforms that can be used to assay SNPs in desirable numbers (Ragoussis, 2009). Some of these platforms include (i) Illiminas GoldenGate platform (Syvanen, 2005; Fan, Chee, and Gunderson, 2006), (ii) Illumina's BeadChip™ based Infinium platform (Steemers and Gunderson, 2007), (iii) TaqMan by Life Technologies (Livak, 2003), and (iv) Competitive Allele Specific PCR (KASPar) by KBiosciences (http://www.kbioscience.co.uk/index.html).

Molecular Markers: The Future

NGS platforms, as mentioned above, have revolutionized genomics approaches to biology and have drastically increased the speed at which DNA sequence can be acquired while reducing the costs by several orders of magnitude. NGS methods for genome-wide genetic marker development

and genotyping are now making use of restriction enzyme digestion of target genomes to reduce the complexity of the target. NGS of reduced representations is proving useful in the rapid and robust identification of SNPs and small indels in a range of plant species even with complex genomes (Deschamps and Campbell, 2009). The use of restriction enzymes for high-throughput genetic marker discovery and genotyping have several advantages (see below) and are becoming the methods of choice for marker discovery in near future. One of the advantages of these methods being their suitability for both model organisms with high-quality reference genome sequences and also to nonmodel species with no available genomic data (see Davey *et al.*, 2011). Therefore, it is anticipated that these emerging technologies will answer many complex biological questions with high accuracy. For instance, these methods may help us in identifying recombination breakpoints for linkage mapping or QTL mapping, locating differentially expressed genomic regions between populations for quantitative genetics studies, genotyping large number of progenies for MAS, or resolving the phylogeography of tens of wild populations.

The innovative and emerging methods of marker discovery making use of NGS and restriction enzymes involve the following key steps: (i) digestion of multiple samples of genomic DNA (from individuals or populations) with one or more restriction enzymes, (ii) selection or reduction of the resulting restriction fragments, and (iii) NGS of the final set of fragments suitable for current NGS platforms (<1 kb in size). Once these steps are through, the polymorphisms in the resulting sequenced fragments can be treated as genetic markers. All these emerging methods can be grouped into three broad classes: (i) reduced-representation sequencing, including reduced-representation libraries (RRLs) and complexity reduction of polymorphic sequences (CRoPS); (ii) restriction site-associated DNA (RAD)-seq; and (iii) low coverage genotyping, including multiplexed shotgun genotyping (MSG) and genotyping by sequencing (GBS). Depending on the need, one of the above class can be selected. For instance, for the study of crop plants, where no reference genome is yet available, a large number of markers need to be scored accurately in most individuals to ensure precise population parameters estimation, RAD-seq or reduced-representation methods are most appropriate. However, for genotyping applications in QTL mapping and MAS, where parental genotypes are well known and progenies with limited polymorphism are to be sequenced, low-coverage genotyping is considered sufficient for linkage to be inferred, provided that a reference genome is available. A brief account on these emerging methods is given below and a comparison has been provided in Table 10.1.

Reduced-Representation Sequencing (CRoPS and RRLs)

Although whole-genome resequencing of populations will soon become affordable, one can believe that reduced-representation methods will be still

Table 10.1 Comparison of various NGS-based high-throughput marker discovery methods.

Parameter	CRoPS	RRLs	RAD-seq	GBS	MGS
DNA required	300 ng/sample	25 μg pooled	300 ng/sample	100 ng/sample	10 ng/sample
Restriction enzyme used	Frequent cutter	Frequent cutter	Both frequent and rare cutter	Frequent cutter	Frequent cutter
PCR needed	–	–	√	√	–
SNP discovery in populations without reference genome	√	√	√	Yes but challenging	Yes but challenging
Suitability for large repetitive genome fractions or higher ploidy levels	×	×	√	√	√
Polymorphism detection	***	***	**	**	**
Suitability for QTL mapping and MAS	*	*	**	***	***

***, high; **, moderate; *, low; √, yes; ×, not suitable; –, no information.

preferable as many biological problems can be answered with a small set of markers and thus do not require every base of the genome to be sequenced. This approach will also save cost and time and has been earlier used to reduce the sequencing work by methylation filtration or Cot-fractionation in maize (Emberton *et al.*, 2005).

RRLs and CRoPS are two such methods, where only a subset of genomic regions instead of whole genome is sequenced (see Davey *et al.*, 2011). Both these methods are suitable for populations with low levels of polymorphism, since they involve use of an enzyme with a higher cutting frequency to produce sufficient polymorphic markers. One of the disadvantages of these methods is that they are not suitable for genomes with a large repetitive genome fractions or high ploidy levels, e.g., wheat. When a high-quality reference genome is available, the reads from RRL sequencing can be mapped to the reference genome and SNPs can be called for whole-genome resequencing projects (Li *et al.*, 2009; Nielsen, Albrechtsen, and Song, 2011). However, if a reference genome is not available, long reads from the Roche Genome Sequencer platform or reads from both ends of the library fragments from any NGS platform (paired-end reads) can be used to assemble the fragments *de novo* before calling SNPs. Paired-end reads also facilitate the calling of structural variations in RRLs (Kerstens *et al.*, 2011).

RRLs were initially used for preparation of an SNP map of the human genome using capillary sequencing (Altshuler *et al.*, 2000). In plants, RRLs

approach in combination of NGS has been used recently for identification of thousands to millions of candidate SNPs in maize (Gore *et al.*, 2009), soybean (Hyten *et al.*, 2010a), common bean (Hyten *et al.*, 2010b), and jointed goatgrass (You *et al.*, 2011). RRLs have usually been used to sequence pools of DNA samples from multiple individuals, thus allowing the detection of polymorphisms within a population but not for each individual.

In contrast to RRLs, CRoPS can be employed to identify polymorphisms in individual samples by incorporating short barcode identifier sequences [designated as multiplex identifier sequences (MIDs) on the Roche Genome Sequencer platform] into the ligated adaptors and using an adaptor containing a unique barcode for each DNA sample. The barcodes can be used to separate sequence reads for different samples computationally, and enable population studies to be carried out using NGS sequencing platform(s). CRoPS has also been used to discover more than 1000 SNPs in maize (van Orsouw *et al.*, 2007; Mammadov *et al.*, 2010).

RAD-Tag Sequencing (RAD-seq)

RAD-seq method provides a reliable means for genome complexity reduction (Miller *et al.*, 2007) and is based on obtaining the sequence adjacent to a set of particular restriction enzyme recognition sites. The application of high-throughput sequencing technology has allowed significant progress in developing a RAD genotyping platform. The value of sequencing restriction site-associated genomic DNA (i.e., RAD tags) for high-density SNP discovery and genotyping was first demonstrated by Baird *et al.* (2008). This involves digestion of genomic DNA with a six to eight base-cutter restriction enzyme, and a barcoded adapter is ligated to compatible sticky ends. Before sequencing, DNA samples each with a different barcode are pooled, randomly sheared to a length suitable for the sequencing platform (300–700 bp), and a second adapter is ligated after polishing and filling ends (Elshire *et al.*, 2011).

RAD-tag sequencing has been found very effective for the rapid and large discovery of molecular markers, even in a species with low polymorphism. For instance, in case of eggplant (*Solanum melongena*), RAD-tag sequencing has resulted in the development of >10 000 SNPs, 1600 indels and 1800 putative SSRs (Barchi *et al.*, 2011). These markers will prove useful for rapid saturation of the best available intraspecific genetic map in eggplant and for the study of comparative genomic analyses within the *Solanaceae* family. In addition, RAD-seq has been used very recently for the construction of linkage maps in barley (Chutimanitsakun *et al.*, 2011) and ryegrass (Pfender *et al.*, 2011).

Low-Coverage Sequencing for Genotyping (GBS and MSG)

The methods described above reduce the proportion of the genome targeted for sequencing so that each marker can be sequenced at high

coverage across many individuals at low cost and high accuracy. However, another alternative to this approach is to sequence many target markers at low coverage per individual, and the decision about the number of markers, coverage, and number of individuals genotyped depend on the goal(s) of study (Davey *et al.*, 2011). When the high-quality reference sequence is available, genomic DNA of different individuals can be sequenced at low coverage. Challenges arise when the reference genome sequence is not available, and if available is poorly assembled, or derived from a distantly related taxa or a species with large and repetitive genome. However, the method can be successfully implemented in genotyping recombinant populations in which the parental genotypes are either known or can be assigned probabilities. This approach has been used to construct genetic maps for rice based on low-coverage whole-genome resequencing of hundreds of recombinant inbred lines (Huang *et al.*, 2009; Xie *et al.*, 2010) and to generate a haplotype map of maize based on 3.3 million SNPs, using low-coverage sequencing of three RRLs that were cut with a range of different restriction enzymes (Gore *et al.*, 2009).

Genotyping by Sequencing

The value of reducing genome complexity with restriction enzymes coupled with multiplex NGS for high-density SNP discovery and genotyping was originally demonstrated with RAD tags (Baird *et al.*, 2008). GBS also involves the digestion of genomic DNA with a frequent cutter and next-generation high-throughput sequencing of all resulting restriction fragments (Elshire *et al.*, 2011). In case of maize and barley, methylation-sensitive enzyme "*Ape*KI" was used to reduce the complexity and to select the hypomethylated regions of genome. One of the advantages of using GBS is that it can be applied to any crop species at a low per-sample cost and is fairly straightforward for small genomes. However, for species with complex genomes, target enrichment or reduction of genome complexity must be employed to ensure sufficient overlap in sequence coverage. The barcoding strategy used in GBS is same to RAD-seq but with fewer sequence phasing errors. Compared with the RAD-seq method, the procedure described here is substantially less complicated; generation of restriction fragments with appropriate adapters is more straightforward; single-well digestion of genomic DNA and adapter ligation results in reduced sample handling; there are fewer DNA purification steps; and fragments are not size selected.

GBS is a technically simple, highly multiplexed approach that may lead to the discovery of ~25 000 SNP markers in one experiment and may be suitable for population studies, germplasm characterization, breeding, and trait mapping in diverse organisms (Elshire *et al.*, 2011). The sequence tag in GBS can be treated as dominant markers for kinship analysis in absence of a reference genome. In addition, plant breeders may conduct genomic selection on a novel germplasm or species without first having to develop

any prior molecular tools, or conservation biologists to determine population structure without prior knowledge of the genome or diversity in the species. These exciting new avenues for applying GBS to breeding, conservation, and global species and population surveys are now poised to become an indispensable component of future biology.

Multiplexed Shotgun Genotyping

Multiplexed shotgun genotyping follows a similar approach as in the case of GBS, except that only a barcoded adaptor is used that is ligated to both ends of each fragment, and fragments are size selected before sequencing. This approach has been used to identify recombination breakpoints in a large number of individuals simultaneously at a resolution sufficient for most mapping purposes including mapping of QTLs and induced mutations (Andolfatto *et al.*, 2011). This does not require genotyping of every marker for every individual, but it does require that markers are mapped to a relatively well-assembled reference genome (with a median scaffold size of >100 kb).

Novel Approaches or Platforms for Plant Breeding

The advent of NGS technologies and high-throughput marker genotyping platforms offer the possibility to generate high-density genome-wide marker profiles in low-cost and high-throughput approach manner. It is also important to note that there are several genotyping and sequencing centers around the world that offer utilizing the sequencing and genotyping facilities. Therefore, it is possible for geneticists and breeders from developed as well as developing countries to have access to the high-density and genome-wide marker profiling. As a result, the use of not only commonly used genetics and breeding approaches such as linkage mapping, marker assisted backcrossing, marker assisted recurrent selection, and advanced backcross QTL analysis will be accelerated, the adoption of new approaches such as GWAS mainly used in human disease studies and GWS mainly used in cattle breeding will also be facilitated in plant genetics and breeding applications. A brief account on these two approaches has been presented as following.

Genome-Wide Association Studies

Most of the traits in plants are complex quantitative in nature, and for the genetic dissection of these traits, two most important approaches including linkage analysis and association mapping have been proposed. Linkage mapping has been extensively used in the past and has the potential to localize major genes within 10–20 cM interval using as few as

200 SSR markers. However, this approach may not be effective and powerful enough to detect large number of small effect genes/QTLs (interacting in a complicated manner) controlling complex quantitative traits like that of drought tolerance. In this context, GWAS has potential to overcome some of the disadvantages associated with linkage mapping and is therefore a preferred approach (see Myles *et al.*, 2009; Chamarthi *et al.*, 2011; Mir *et al.*, 2011). GWAS has been initially applied to map diseases or genetic disorder in human (Ozaki *et al.*, 2002; Altshuler, Daly, and Lander, 2008; Donnelly, 2008).

In short, GWAS is the genotyping with enough markers distributed throughout the genome of an organism so as to assure that the functional alleles will likely be in linkage disequilibrium (LD) with at least one of the genotyped markers (Myles *et al.*, 2009). The different steps involved in GWAS include (i) discovery of large number of SNPs segregating in a small panel of genotypes, (ii) development of suitable genotyping assays, and (iii) genotyping of suitable germplasm/core/mini-core collections for which extensive phenotypic data on the targeted traits are already available. The number of the SNPs and their density required for genotyping the germplasm collection will, of course, depend on genome size and LD decay in the species and the germplasm collections. Therefore, the number of markers required for undertaking GWAS varied across different species; for instance, 140 000 markers for *Arabidopsis* genome (Kim *et al.*, 2007), more than two million markers for grapevine, and 10–15 million for diverse maize varieties (Myles *et al.*, 2009). Genotyping of the germplasm collections with such a high-density markers was unimaginable in earlier times, the new marker-genotyping platforms like Infinium assays or NGS-based marker systems (e.g., RRLs, CroPS, GBS, MSG) mentioned in the article can offer such a possibility. For instance, Infinium assays have become available in some crops like soybean (Haun *et al.*, 2011), maize (Martin Ganal, personal communication), and loblolly pine (Eckert *et al.*, 2010), and GBS approach is being optimized in maize and barley (Elshire *et al.*, 2011).

In recent years, several GWAS reports have become available in plant species like *Arabidopsis* (Aranzana *et al.*, 2005; Nordborg and Weigel, 2008), maize (Kump *et al.*, 2011; Tian *et al.*, 2011). It is anticipated that with the pace of advances being made in the area of genomics and bioinformatics, the next few years may be an exciting time to see GWAS getting underway in majority of the major crop species. While deploying the GWAS, the plant communities need to aware with challenges associated with such studies that include (i) design and data analysis, (ii) choice of SNPs and/or sequencing platforms for high-density genotyping, (iii) SNP × SNP interactions in a whole-genome scan, and (iv) genotyping errors (Thomas, 2006). Nevertheless, with the help of GWAS, it will be possible to uncover all the genes/QTLs responsible for quantitative and complex traits that are of interest to breeders and then use them in molecular breeding for crop improvement.

Genome-Wide Selection

GWS or genomic selection is one of the recently emerged molecular breeding approaches for improving quantitative traits in large plant breeding populations using genome-wide marker profiles (Meuwissen, Hayes, and Goddard, 2001; Bernardo and Yu, 2007; Jannink, Lorenz, and Iwata, 2010). In GWS approach, genomic predictions are made for a possible performance of an individual in a given population, which basically rely on LD between genetic markers and QTL. GWS uses two types of datasets: a training population and a validation population (Rutkoski, Heffner, and Sorrells, 2010). The training set is the reference population comprising the breeding lines used in the breeding program. In general, one of the following information is available or generated on this training population: (i) phenotypic data over a range of environmental conditions; (ii) genome-wide marker profiling data; or (iii) pedigree information or kinship. These datasets are used with certain statistical methods to incorporate this information. Subsequently, based only on the marker effects, the genetic values of new genotypes, popularly called the genomic estimated breeding values (GEBVs), are predicted. The validation set contains the selection candidates (derived from the reference population) that have been genotyped (but not phenotyped) and selected based on GEBVs in the training set. Subsequently, selected candidates are used for the crossing and the desirable progenies can be selected further by using the same model mentioned above (Jannink *et al.*, 2010).

In brief, GWS combines powerful statistical methods with new marker-genotyping approach, as mentioned in this article, to select untested germplasm lines based on predicted performance. It reduces the expense and years involved in field testing, thereby greatly cutting the time needed to complete plant breeding cycles and bring new varieties to market. In addition, plant breeders can select for the ability of particular varieties to thrive under other agronomic stresses faced by smallholder farmers, like drought or nitrogen-depleted soil.

GWS differs from the traditional breeding (TB) and MAS approaches. For instance, TB programs rely mainly on phenotypes being evaluated in several environments; selection and recombination are based solely on the resulting data plus pedigree information, when available. MAS approach uses molecular markers in LD with QTL and a progeny line is selected based on the marker allele associated with the trait of interest. In case of GWS, prediction of a breeding line is made after combining genome-wide marker profile data with phenotypic and pedigree data (when available). As a result, GWS increases the accuracy of the prediction of breeding and genotypic values. Furthermore, in a TB program, the crop breeding cycle is about 5–7 years that can be reduced to about 3 years by using MAS. In contrast, GWS shortens it to as little as 1 year.

Although the potential of GWS has been demonstrated in the case of Hybro-broilers (Euribrid; http://www.thepigsite.com/swinenews/

12912/first-use-of-commercial-genomic-selection) in chicken, its potential is yet to be seen in the case of the crop breeding. As required tools and technologies for implementing GWS are now available in at least some crops such as maize and wheat, some efforts to deploy them have been initiated in these crops (Jean-Luc Jannink, personal communication) An important challenge in implementing GWS in crop species is difficulty in calculating the GEBVs based on phenotyping data on different set of the populations. In such cases, GBVs could be predicted with multiple within-population evaluations or with one across-population evaluation in which the training set comprises individuals from all populations. Combining populations in a training set may be advantageous because the effects of the markers can be estimated from a larger number of phenotypes. This is particularly of interest when the training set for one of the populations is too small for a proper within-population evaluation. On the other hand, it is expected that some markers may be in high LD with a QTL in one population but not in the other population, especially when these markers are more distant from the QTL or when the populations have diverged for many generations (Andreescu *et al.*, 2007; Gautier *et al.*, 2007).

In summary, besides accelerating the selection cycles, genomic selection offers the opportunity to increase the selection gains per unit of time. Therefore, it is believed that alternating progeny field testing with selection based only on markers should increase the genetic gains per unit of time. However, unresolved questions such as how much (if any) genetic diversity will be diminished by this combination of phenotypic and GWS remains. As mentioned in this article, generating genome-wide marker profiling data has become cheaper as compared with undertaking phenotyping on larger populations; GWS holds good potential to be used in breeding programs in coming future.

Conclusions

A continuous evolution in molecular marker technologies has resulted in the development of ultra-high-throughput genotyping platforms. However, low-throughput molecular markers such as SSRs are still indispensable for tracking specific genomic regions in molecular breeding programs. SNP markers are most preferable for development of high-throughput genotypic platforms for genome-wide marker screening. It seems that the recently emerged NGS-based molecular marker system may replace the array-based high-throughput marker systems in coming future, especially when costs is decreasing and throughput is increasing for the NGS technologies. These future marker systems may prove very useful for enhancing deployment of modern genetics and breeding approaches such as GWAS and GWS that are still in infancy in plant systems for crop improvement.

Acknowledgments

Authors are thankful to Generation Challenge Programme (GCP) of CGIAR, Department of Biotechnology (DBT) and Indian Council of Agricultural Research (ICAR) of Government of India for funding the research of authors. Thanks are also due to Mahendar Thudi for extending his help in writing the manuscript.

References

Altshuler, D., Daly, M.J. and Lander, E.S. (2008) Genetic mapping in human disease. *Science*, **322**, 881–888.

Altshuler, D., Pollara, V.J., Cowles, C.R. *et al.* (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Andolfatto, P., Davison, D., Erezyilmaz, D. *et al.* (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Research*, **21**, 610–617.

Andreescu, C., Avendano, S., Brown, S.R. *et al.* (2007) Linkage disequilibrium in related breeding lines of chickens. *Genetics*, **177**, 2161–2169.

Aranzana, M.J., Kim, S., Zhao, K. *et al.* (2005) Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, **1**, e60.

Baird, N.A., Etter, P.D., Atwood, T.S. *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, **3**, e3376.

Banks, T.W., Jordan, M.C. and Somers, D.J. (2009) Single-feature polymorphism mapping in bread wheat. *Plant Genome*, **2**, 167–178.

Barchi, L., Lanteri, S., Portis, E. *et al.* (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, **12**, 304.

Becker, J., Vos, P., Kupier, M. *et al.* (1995) Combined mapping of AFLP and RFLP markers in barley. *Molecular Genetics and Genomics*, **249**, 65–73.

Bernardo, R. and Yu, J. (2007) Prospects for genome wide selection for quantitative traits in maize. *Crop Science*, **47**, 1082–1090.

Bernatzky, R. and Tanksley, S.D. (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics*, **112**, 887–898.

Botstein, D., White, R.L., Skolnick, M. *et al.* (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32**, 314–331.

Caetano-Anolles, G., Bassam, B.J. and Gresshoff, P. (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnology*, **9**, 553–557.

Chamarthi, S.K., Kumar, A., Vuong, T. *et al.* (2011) Trait mapping and molecular breeding in legumes: concepts and examples in soybean, common bean and chickpea, in *Biology and Breeding of Food Legumes* (eds A. Pratap and J. Kumar), CABI International, Oxfordshire.

Cho, R.J., Mindrinos, M., Richards, D.R. *et al.* (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics*, **23**, 203–207.

Chutimanitsakun, Y., Nipper, R.W., Cuesta-Marcos, A.L. *et al.* (2011) Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. *BMC Genomics*, **12**, 4.

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

Cui, A., Xu, J., Asghar, R. *et al.* (2005) Detecting single-feature polymorphisms using oligonucleotide arrays and robustified projection pursuit. *Bioinformatics*, **21**, 3852–3858.

Davey, J.W., Hohenlohe, P.A., Etter, P.D. *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Genetics*, **12**, 499–510.

Deschamps, S. and Campbell, M. (2009) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Molecular Breeding*, **25**, 553–570.

Donnelly, P. (2008) Progress and challenges in genome-wide association studies in humans. *Nature*, **456**, 728–731.

Dubey, A., Farmer, A., Schluetter, J. *et al.* (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). *DNA Research*, **18**, 153–164.

Eckert, A.J., van Heerwaarden, J., Wegrzyn, J.L. *et al.* (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, **185**, 969–982.

Ehrlich, H.A., Gelfand, D.H. and Sninsky, J.J. (1991) Recent advances in the polymerase chain reaction. *Science*, **252**, 1643–1651.

Elshire, R.J., Glaubitz, J.C., Sun, Q. *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

Emberton, J., Ma, J., Yuan, Y. *et al.* (2005) Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. *Genome Research*, **15**, 1441–1446.

Fan. J.B., Chee, M.S. and Gunderson, K.L. (2006) Highly parallel genomic assays. *Nature Review Genetics*, **7**, 632-644.

Garg, R., Patel, R.K., Tyagi, A.K. *et al.* (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*, **18**, 53–63.

Gautier, M., Faraut, T., Moazami-Goudarzi, K. *et al.* (2007) Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*, **177**, 1059–1070.

Glaszmann, J.C. (1987) Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics*, **74**, 21–30

Gore, M., Bradbury, P., Hogers, R. *et al.* (2007) Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. *Crop Science*, **47**, 135–148.

Gore, M.A., Chia, J.M., Elshire, R.J. *et al.* (2009) A first-generation haplotype map of maize. *Science*, **326**, 1115–1117.

Gupta, P.K., Balyan, H.S., Edwards, K.J. *et al.* (2002) Genetic mapping of 66 new microsatellite (SSR) in bread wheat. *Theoretical and Applied Genetics*, **105**, 413–422.

Gupta, P.K., Rustgi, S. and Mir, R.R. (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity*, **101**, 5–18.

Gupta, P.K. and Varshney, R.K. (2000) The development and use of microsatellite markers for genetics and plant breeding with emphasis on bread wheat. *Euphytica*, **113**, 163–185.

Gupta, P.K., Varshney, R.K., Sharma, P.C. *et al.* (1999) Molecular markers and their applications in wheat breeding. *Plant Breeding*, **118**, 369–390.

Haun, W.J., Hyten, D.L., Xu, W.W. *et al.* (2011) The composition and origin of genomic variation among individuals of the soybean reference cultivar Williams 82. *Plant Physiology*, **155**, 645–655.

Hiremath, P.J., Farmer, A., Cannon, S.B. *et al.* (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnology Journal*, **9**, 922–931.

Huang, S., Li, R., Zhang, Z. *et al.* (2009) The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics*, **41**, 1275–1281.

Hyten, D.L., Cannon, S.B., Song, Q. *et al.* (2010a) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics*, **11**, 38.

Hyten, D.L., Song, Q., Fickus, E.W. *et al.* (2010b) High-throughput SNP discovery and assay development in common bean. *BMC Genomics*, **11**, 475.

Ishikawa, R., Harada, T., Niizeki, M. *et al.* (1992) Reconstruction of linkage map with isozyme, morphological and physiological markers in rice chromosome-12. *Japanese Journal of Breeding*, **42**, 235–244.

Iorizzo, M., Senalik, DA., Grzebelus, D. *et al.* (2011) *De novo* assembly and characterization of the carrot transcriptom reveals novel genes, new markers, and genetic diversity. *BMC Genomics*, **12**, 389.

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.

Jannink, J.-L., Lorenz, A.J. and Iwata, H. (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, **9**, 166–177.

Jones, E., Chu, W.-C., Ayele, M. *et al.* (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. *Molecular Breeding*, **24**, 165–176.

Jones, E., Sullivan, H., Bhattramakki, D. *et al.* (2007) 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*, **115**, 361–371.

Kaur, S., Cogan, N.O.I., Pembleton, L.W. *et al.* (2001) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. *BMC Genomics*, **12**, 265.

Kerstens, H.H.D., Crooijmans, R.P.M.A., Dibbits, B.W. *et al.* (2011) Structural variation in the chicken genome identified by paired-end next-generation DNA sequencing of reduced representation libraries. *BMC Genomics*, **12**, 94.

Kilian, A., Huttner, E., Wenzl, P. *et al.* (2005) The fast and the cheap: SNP and DArT-based whole genome profiling for crop improvement. In *Proceedings of the International Congress in the Wake of the Double Helix: From the Green Revolution to the Gene Revolution* (eds R. Tuberosa, R.L. Phillips and M. Gale), May 27–31, 2003, Avenue Media, Bologna, pp. 443–461.

Kim, S., Plagnol, V., Hu, T.T. *et al.* (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, **39**, 1151–1155.

Kim, S., Zhao, K., Jiang, R. *et al.* (2006) Association mapping with single-feature polymorphisms. *Genetics*, **173**, 1125–1133.

Kim, S-H., Bhat, P.R., Cui, X. *et al.* (2009) Detection and validation of single feature polymorphisms using RNA expression data from a rice genome array. *BMC Plant Biology*, **9**, 65.

Kirst, M., Caldo, R., Casati, P. *et al.* (2006) Genetic diversity contribution to errors in short oligonucleotide microarray analysis. *Plant Biotechnology Journal*, **4**, 489–498.

Kumar, R., Qiu, J., Joshi, T. *et al.* (2007) Single feature polymorphism discovery in rice. *PLoS One*, **2**, e284.

Kump, K.L., Bradbury, P.J., Buckler, E.S. *et al.* (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics*, **43**, 163–168.

Lai, J., Li, R., Xu, X. *et al.* (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics*, **42**, 1027–1030.

Langridge, P. and Chalmers, K. (2004) The principle: identification and application of molecular markers. in *Biotechnology in Agriculture and Forestry, Molecular Marker Systems in Plant Breeding and Crop Improvement* (eds H. Lörz and G. Wenzel), Springer-Verlag, Berlin, 49, 129–149.

Li, R., Fan, W., Tian, G. *et al.* (2009) The sequence and de novo assembly of the giant panda genome. *Nature*, **463**, 311–317.

Litt, M. and Lutty, J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, **44**, 397–401.

Livak, K.J. (2003) SNP genotyping by the 5′-nuclease reaction. *Methods in Molecular Biology*, **212**, 129–147.

Mackill, D.J., Zhang, Z., Redona, E.D. *et al.* (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome*, **39**, 969–977.

Mammadov, J.A., Chen, W., Ren, R. *et al.* (2010) 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*, **121**, 577–588.

Market, C.L. and Moller, F. (1959) Multiple forms of enzymes, tissue, ontogenetic and species specific pattern. *Proceedings of the National Academic Sciences of the United States of America*, **45**, 753–763.

McMillin, D.E., Allan, R.E. and Roberts, D.E. (1986) Association of an isozyme locus and strawbreaker foot rot resistance derived from *Aegilops ventricosa* in wheat. *Theoretical and Applied Genetics*, **72**, 743–747.

Metzker, M.L. (2010) Sequencing technologies—the next generation. *Nature Review Genetics*, **1**, 31–46.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.

Miller, M.R., Dunham, J.P., Amores, A. *et al.* (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240–248.

Mir, R.R., Zaman-Allah, M., Sreenivasulu, N. *et al.* (2011) Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theoretical and Applied Genetics*. doi: 10.1007/s00122-012-1904-9.

Mohler, V. and Schwarz, G. (2005) Genotyping tools in plant breeding: from restriction fragment length polymorphisms to single nucleotide polymorphisms, in *Biotechnology in Agriculture and Forestry* (eds H. Lorz and G. Wenzel), Springer, Berlin, pp. 23–38

Morgante, M. and Vogel, J. (1994) Compound micro satellite primers for the detection of genetic polymorphisms. U.S. Patent Application, 08/326456.

Myles, S., Peiffer, J., Brown, P.J. *et al.* (2009) Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell*, **21**, 2194–2202.

Nielsen, P.J.S., Albrechtsen, A. and Song, Y.S. (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Review Genetics*, **12**, 443–451.

Nordborg, M. and Weigel, D. (2008) Next-generation genetics in plants. *Nature*, **456**, 720–723.

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

Ozaki, K., Ohnishi, Y., Iida, A. *et al.* (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nature Genetics*, **32**, 650–654.

Paran, I. and Michelmore, R.W. (1993) Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics*, **85**, 985–993.

Pfender, W.F., Saha, M.C., Johnson, E.A. *et al.* (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theoretical and Applied Genetics*, **122**, 1467–1480.

Potokina, E., Druka, A., Luo, Z. *et al.* (2008) Gene expression quantitative trait locus analysis of 16000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J*, **53**, 90–101.

Powell, W., Morgante, M., Andre, C. *et al.* (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, **2**, 225–238.

Rafalski, A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, **5**, 94–100.

Ragoussis, J. (2009) Genotyping technologies for genetic research. *Annual Review of Genomics and Human Genetics*, **10**, 117–133.

Rostoks, N., Borevitz, J.O., Hedley, P.E. *et al.* (2005) Single-feature polymorphism discovery in the barley transcriptome. *Genome Biology*, **6**, R54.

Rutkoski, J.E., Heffner, E.L. and Sorrells, M.E. (2010) Genomic selection for durable stem rust resistance in wheat. *Euphytica*, **179**, 161–173.

Saxena, R.K., Cui, X., Thakur, V. *et al.* (2011) Single feature polymorphisms (SFPs) for drought tolerance in pigeonpea (*Cajanus* spp.). *Functional and Integrative Genomics*, **11**, 651–657.

Steemers, F.J. and Gunderson, K.L. (2007) Whole genome genotyping technologies on the Bead Array platform. *Biotechnology Journal*, **2**, 41–49.

Syvanen, A.C. (2005) Toward genome-wide SNP genotyping. *Nature Genetics*, **37**, S5–S10.

Tanksley, S.D., Young, N.D., Paterson, A.H. *et al.* (1989) RFLP mapping in plant breeding: new tools for an old science. *Nature Biotechnology*, 7, 257–264.

Thomas, D.C. (2006) Are we ready for genome-wide association studies? *Cancer Epidemiol Biomarkers Prevention*, **15**, 595–598.

Tian, F., Bradbury, P.J., Brown, P.J. *et al.* (2011) Genome-wide association study of maize identifies genes affecting leaf architecture. *Nature Genetics*, **43**, 159–162.

Tingey, S.V. and del Tufo, J.P. (1993) Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiology*, **101**, 349–352.

van Orsouw, N.J., Hogers, R.C.J., Janson, A. *et al.* (2007) Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One*, **2**, e1172.

Varshney, R.K., Graner, A. and Sorrells, M.E. (2005) Genomics-assisted breeding for crop improvement. *Trends in Plant Science*, **10**, 621–630.

Varshney, R.K., Nayak, S.N. and May, G.D. *et al.* (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology*, **27**, 522–530.

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

Welsh, J. and McClelland, M. (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research*, **18**, 7213–7218.

Wenzl, P., Carling, J. and Kudrna, D. (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 9915–9920.

Wenzl, P., Catizone, I., Thomson, B. *et al.* (2007) A DArT platform for high-throughput profiling of durum wheat. In *Plant and Animal Genome XV Conference*, Town & Country Convention Center, San Diego, CA, p. 263.

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

Winzeler, M., Winzeler, H. and Keller, B. (1995) Endopeptidase polymorphism and linkage of the *Ep-D1c* null allele with the *Lr19* leaf rust resistance gene in hexaploid wheat. *Plant Breeding*, **114**, 24–28.

Xie, W., Feng, Q., Yu, H. *et al.* (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing.

Proceedings of the National Academy of Sciences of the United States of America, **107**, 10578–10583.

You, F.M., Huo, N., Deal1, K.R. *et al.* (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics*, **12**, 59.

Yu, J., Holland, J.B., McMullen, M. *et al.* (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics*, **178**, 539–551.

Zhao, W., Canaran, P., Jurkuta, R. *et al.* (2006) Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucleic Acids Research*, **34**, D725–D757.