

REVIEW

Array-based high-throughput DNA markers for crop improvement

PK Gupta¹, S Rustgi² and RR Mir¹

¹*Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut, Uttar Pradesh, India*
and ²*Johnson Hall, Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA*

The last two decades have witnessed a remarkable activity in the development and use of molecular markers both in animal and plant systems. This activity started with low-throughput restriction fragment length polymorphisms and culminated in recent years with single nucleotide polymorphisms (SNPs), which are abundant and uniformly distributed. Although the latter became the markers of choice for many, their discovery needed previous sequence information. However, with the availability of microarrays, SNP platforms have been developed, which allow genotyping of thousands of markers in parallel. Besides SNPs, some other novel marker systems, including single feature polymorphisms, diversity array technology and restriction site-associated DNA markers, have also been developed, where array-

based assays have been utilized to provide for the desired ultra-high throughput and low cost. These microarray-based markers are the markers of choice for the future and are already being used for construction of high-density maps, quantitative trait loci (QTL) mapping (including expression QTLs) and genetic diversity analysis with a limited expense in terms of time and money. In this study, we briefly describe the characteristics of these array-based marker systems and review the work that has already been done involving development and use of these markers, not only in simple eukaryotes like yeast, but also in a variety of seed plants with simple or complex genomes.

Heredity (2008) **101**, 5–18; doi:10.1038/hdy.2008.35; published online 7 May 2008

Keywords: SNPs; SFPs; DArT; microarray; RAD markers

Introduction

A study of DNA polymorphism has become an active area of research in all important crops and several model plant species like *Arabidopsis thaliana* and *Brachypodium distachyon*. This involves development and use of molecular markers, which have proved useful not only for marker-assisted selection during plant breeding, but also for understanding crop domestication and plant evolution. To resolve the pattern of DNA polymorphism in any crop, the ultimate approach would be to sequence/resequence the entire genome (or a part of it) in a large number of accessions. This was, however, unimaginable during the 1980s and still remains cost ineffective, therefore DNA-based molecular markers (for example, restriction fragment length polymorphisms, random amplified polymorphic DNAs, simple sequence repeat (SSRs) and amplified fragment length polymorphisms (AFLPs)) have largely been employed for the study of DNA polymorphism (Collard *et al.*, 2005). Most of these molecular markers are based on the use of restriction digestion of genomic DNA, followed by hybridization of electrophoresed DNA, and/or visualization of the products of PCR carried out using suitably

designed PCR primers. More recently, however, single nucleotide polymorphisms (SNPs), whose discovery was largely based on sequence information, became the markers of choice due to their abundance and uniform distribution throughout a genome. Once discovered, SNP genotyping can be done using any of the dozens of available methods. For SSR/SNP genotyping, some efforts in the past were made to provide for the desired high throughput and cost effectiveness through the use of PCR tetrad machines (handling 384 PCR reactions at a time), multiplexing, multiple loading of the gels and the use of automatic sequencers. However, this appeared inadequate and there has been an increasing demand to develop ultra-high-throughput low-cost assays for a variety of novel marker systems including SNPs. These new methods will allow ultra-high-throughput genotyping of either one or few individuals for hundreds of thousands of markers, or that of thousands of individuals for one or few markers. High-density oligonucleotide arrays, which are now becoming available in several crops, provide a means for achieving this goal of low-cost ultra-high-throughput genotyping. These arrays may also be custom made according to specific needs and, therefore, also allowed for the development of novel marker systems like single feature polymorphisms (SFPs) (including gene-specific hybridization polymorphisms and gene expression markers), diversity array technology (DArT) and restriction site-associated DNA (RAD) markers, which have now become the markers of choice. Technologies have also been developed, which make use of tag arrays for detection of the products of genotyping

Correspondence: Professor PK Gupta, Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut, Uttar Pradesh 250004, India.

E-mail: pkgupta36@vsnl.com

Received 7 January 2008; revised 23 March 2008; accepted 26 March 2008; published online 7 May 2008

reactions. These novel array- or chip-based markers are useful for a variety of purposes including genome-wide association studies, population studies, bulk segregant analysis, quantitative trait loci (QTL) interval mapping, whole genome profiling and background screening and so on (Steinmetz *et al.*, 2002; Winzeler *et al.*, 2003; Wenzl *et al.*, 2004, 2007a; Hazen *et al.*, 2005; Kim *et al.*, 2006). A brief account of the development and use of these high-throughput array-based molecular markers in plants is presented in this study.

Single nucleotide polymorphisms

Single nucleotide polymorphisms refer to DNA polymorphisms at the level of individual base pairs and constitute ~90% of genetic variation in any organism, therefore DNA-based markers are often classified into SNPs and non-SNP markers. The SNPs are generally discovered *in silico* from genomic or expressed sequence tag (EST) sequences available in the databases, or through sequencing or resequencing of candidate genes/PCR products/whole genomes in more than one genotypes (see later). Once SNPs are discovered, genotyping for these markers can be done using any one of more than 30 different available methods, although only a few of these methods are microarray based, providing the desired ultra-high throughput (for reviews, see Khlestkina and Salina, 2006; Kim and Misra, 2007). The choice of genotyping method largely depends upon the nature of study. For instance, whereas in some cases, we need to scan one or more individuals for SNPs ranging in number from dozens to several thousands, in other cases, we may need to allele-type thousands of individuals for a specific locus. In either case, ultra-high throughput and low-cost techniques are needed.

High-density platforms for SNP genotyping

Several high-density platforms are now available for genotyping one or more genomic DNA samples for dozens to thousands of SNPs in parallel (Table 1). These platforms have been widely discussed in several earlier reviews (Svänen, 2005; Fan *et al.*, 2006), so their detailed discussion in this study will be repetitive. Therefore, only an overview of these platforms is included.

GoldenGate genotyping: Illumina's GoldenGate assay, based on their BeadArray/BeadChip technology, is one of the most popular platforms providing for a cost-effective assay (per genotype cost \$0.03) for genotyping a fairly large collection of samples for a customized pool of SNP markers in parallel (a pool may range from 96 to 1536 SNPs). Each assay involves a multiplexed SNP genotyping reaction, involving use of two allele-specific oligonucleotides and a locus-specific oligonucleotide for each SNP, the locus-specific oligonucleotide carrying an anti-tag sequence for detection by the BeadArray. At a particular level of multiplexing (pooling), 96 or 384 BeadArrays (each bead carrying a tag oligo for detection of the product of SNP genotyping reaction) are arranged in a matrix, called Sentrix Arrays Matrix, so that up to 384 samples can be processed in one reaction, thus permitting genotyping of each of these 384 samples for as many as 1536 SNPs, simultaneously. Multiple pools can also be used to increase the number of SNPs further (for details with illustration, consult Fan *et al.*, 2003; Hyten *et al.*, 2008). However, beyond a certain level of multiplexing, Infinium assay that is discussed later in this review provides a better alternative.

GoldenGate technology is now being used for several crops. For instance, at Southern California Genotyping Consortium, Oligonucleotide Pool Assays are being developed for several plant systems including *Arabidopsis*, barley, wheat and maize, which will be used in the future for high-throughput SNP genotyping in these plant systems. In barley, a pilot oligonucleotide pool assay for detection of SNPs in 1524 unigenes was initially developed at SCRI, UK, which is being extended to an oligonucleotide pool assay for SNPs in as many as ~3000 unigenes. This platform is being used by US barley Coordinated Agricultural Project to allele-type 3840 genotypes and by Association Genetics of UK Elite Barley (AGOUEB) to allele-type ~1500 genotypes. The genotypes will also be assessed for about 40 traits that are pertinent for barley breeding, therefore the information may be used for a study of haplotype-trait associations (Hayes and Szücs, 2006). Also in maize, genome-wide SNP genotyping has been initiated with a large collaborative effort, which has already made use of the Illumina's GoldenGate assay platform to genotype

Table 1 Micro-array-based high-throughput SNP genotyping systems

Features	Illumina		Affymetrix		Beckman coulter
	GoldenGate	Infinium	MIP	GeneChip or oligonucleotide arrays	SNPstream
Array type	Tag array on beads	Specific probe primers on beads	Tag array on glass	Oligonucleotide array on glass	Tag array on glass
Reaction	ASPE	ASPE	SBE	Allele-specific hybridization	SBE
Labelling and detection	Two-colour fluorescence	Biotin-avidin, single-colour fluorescence	Two- or four-colour fluorescence	Biotin-avidin, single-color fluorescence	Two-colour fluorescence
Multiplexing	Multiplexed from 384–1536 (and multiples thereof)	Multiplexed from 10 000 to hundreds of thousands	12 000 SNPs	—	Multiplexed from 12–48 SNPs in 384 samples/array
SNP, sample size	1536 SNPs per array (~110 000 SNPs/SAM)	Up to 500 000 SNPs	10 000 SNPs	Up to 500 000 SNPs	Tens of SNPs, hundreds of samples per plate

Abbreviations: ASPE, allele-specific primer extension; SAM, Sentrix Arrays Matrix; SBE, single-base extension.

7200 RILs for 1300 SNPs using a customized Illumina BeadArray matrix (Yu and Buckler, 2006; Jones *et al.*, 2007; Buckler, personal communication). In soybean, recently a custom-made 384-SNP GoldenGate assay was successfully designed for genotyping of three RIL mapping populations; the above 384 SNPs were discovered through resequencing of five diverse accessions (involved as parents of the above three RIL mapping populations) and were selected such that each of these SNPs segregated in at least one mapping population (Hyten *et al.*, 2008). Similarly, in wheat, a programme on whole genome SNP genotyping using Illumina's GoldenGate and ABI's SNaPshot is being executed at the University of California (<http://wheat.pw.usda.gov/SNP/new/index.shtml>). The above genotyping activity may certainly be extended further through the use of Solexa's ultra-high throughput and low-cost resequencing technology (for reviews, on ultra-high-throughput DNA sequencing, see Gupta, 2008; Mardis, 2008).

Infinium assay for whole genome genotyping: Illumina's BeadChip-based Infinium assay, which is considered to be a more global approach for genotyping, also utilizes BeadArray technology and is a direct approach allowing parallel detection of most SNPs in a genome. More importantly, it eliminates the multiplexing bottleneck in sample preparation (needed in GoldenGate), making assay scalability mainly dependent on array-feature density (Fan *et al.*, 2006). The randomly amplified fragments representing total genomic sequence are hybridized to these BeadArrays available in the form of a BeadChip on a microscope slide with 12 sections/stripes (each section containing 1.1 million beads, carrying decoded oligonucleotides). The 12 stripes may be used either for loading 12 different bead pools for 720 000 assays for a single sample, or alternatively one can load a single bead pool 12 times for 60 000 assays across each of 12 different samples. Also, as many as 24, 48 or 96 BeadChips can be used simultaneously in a temperature-controlled chamber rack to allow robotic-assisted automated assay of multiple genomic DNA samples simultaneously, thus permitting genotyping of hundreds of thousands of SNPs in dozens of genotypes simultaneously (Sylvänen, 2005; Gunderson *et al.*, 2006; Steemers and Gunderson, 2007).

GenomeLab SNPstream genotyping system: Like GoldenGate assay, this genotyping system of Beckman Coulter combines solid-phase primer extension assay and universal tags for SNP genotyping. The instrument designed for this system allows processing of 4600–3 000 000 genotypes per day. DNA samples are used for either 12 or 48 multiplex PCR in a 384 plate using tagged extension primers that are extended using single fluorescence-labelled nucleotide terminator reactions. The PCR-amplified fragments are resolved by hybridization to the complementary tags available on SNPware Tag Array plate having tag arrays in 384-well microplate format, each well with 16 or 52 unique tags that are complementary to the tags of the 12 or 48 extension primers, plus four controls to ensure accuracy. An individual SNP associated with a PCR-amplified fragment is identified by the position of the hybridizing tag in the well. This allows genotyping of 384 samples for either 12 or 48 SNPs per array, as against genotyping of

relatively fewer samples for one thousand to more than a million SNPs in other high-throughput microarray-based SNP genotyping systems (Figure 1). This genotyping system has already been used for human blood/saliva samples for forensic purposes and in several plant systems including corn, canola, cotton (Shah *et al.*, 2003) and poplar (Meirmans *et al.*, 2006).

MegAllele genotyping system (molecular inversion probe or MIP technology): MegAllele genotyping system of Affymetrix is based on ParAllele's Molecular Inversion Probe (MIP) Technology. It allows tens of thousands of genotyping reactions in each of four reaction tubes that are used for each assay. Possible addition of a single specific nucleotide is allowed in each of the four reaction tubes by adding in each tube only one of the four dNTPs, which differ for the four tubes. In contrast to this, GoldenGate uses allele-specific primer extension to score SNPs. Furthermore, MIP uses a single circularizable probe (called padlock probe), whereas in GoldenGate assay, both upstream and downstream probes are separate (Nilsson *et al.*, 1994; Sylvänen, 2005; Fan *et al.*, 2006). However, MIP resembles GoldenGate assay in using tags (available in the form of glass GenFlex Tag Array in MIP and BeadArrays in case of GoldenGate assay) that are used for detection of the products of SNP genotyping reactions. A technology making use of modified padlock probes like those used in MegAllele system has been recently utilized in bread wheat (Reid *et al.*, 2007), and in the future, it may be utilized in other crops also.

GeneChip and allele-specific oligonucleotide tiling arrays: Tiling arrays developed by Affymetrix GeneChip platform on the basis of known sequences in several organisms have also been used for SNP discovery and detection. These tiling arrays may be either designed for resequencing (sequencing by hybridization or SBH) of specific genomic regions for SNP genotyping or may be designed for interrogating every individual nucleotide in a template genomic sequence by multiple probes available on the array. In the latter case, probes

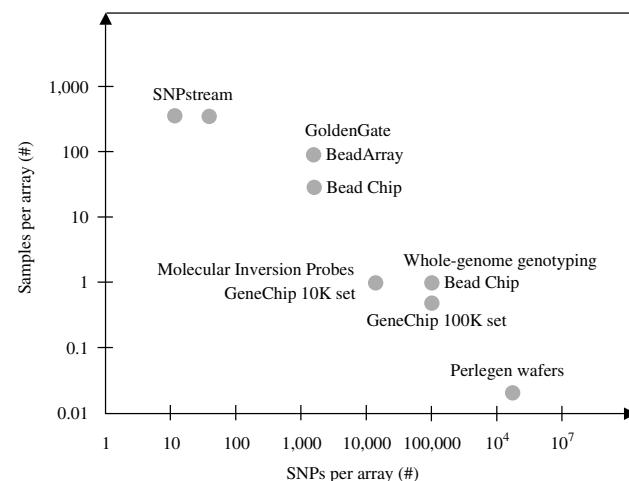


Figure 1 Comparison of SNP multiplexing levels and number of samples analyzed per array in microarray-based SNP genotyping systems (reproduced with permission from ref. Sylvänen, 2005).

are available in the form of probesets, so that for each SNP allele, there is one probeset with multiple probe pairs (each pair with a perfect match and a mismatch), the probes of the two alleles at a locus differing only at a specified position. In Affymetrix assay for SNP genotyping, a genomic representation generated through complexity reduction is labelled and hybridized to the above tiling array, which may contain up to 900 000 different oligos, each present in millions of copies (<http://keck.med.yale.edu/affymetrix/technology.htm>). The hybridization patterns are used for inferring SNP genotypes.

Tiling arrays have already been used in some crops for genome-wide discovery and detection of SNPs. In rice, where genome sequence of one genotype (Nipponbare) is already available, SNP discovery in 100 Mb of the rice genome has been undertaken at International Rice Research Institute using tiling microarrays that are based on allele-specific oligonucleotides from the non-repetitive regions of the genome. This approach allowed identification of 260 000 non-redundant SNPs by Perlegen's model-based (MB) algorithms (McNally *et al.*, 2006; Collard *et al.*, 2008). These SNPs are being currently validated (Collard, personal communication), and the collection of these rice SNPs is being extended through use of machine-learning (ML)-based techniques. The ML approach was used earlier in *Arabidopsis*, where in one study 20 diverse strains of this weed were genotyped for more than a million non-redundant SNPs for analyzing the patterns of DNA polymorphism (Clark *et al.*, 2007), and in another study, a genotyping array based on a subset of 250 000 'tag' SNPs was used to study genome-wide pattern of LD (Kim *et al.*, 2007). An 'InDel array' representing 240 unique InDel markers was also recently developed and used in *Arabidopsis* to genotype InDels of one or more nucleotides in an ultra-high-throughput manner (Salathia *et al.*, 2007).

SNP genotyping/allelotyping for a specific locus

During marker-assisted selection in plant breeding, one may be interested in microarray-based genotyping of thousands of plants for a specific gene of interest. This can be done by arraying PCR products from all segregating individual plants on a glass, followed by hybridization of this array with labelled probes representing alternative alleles of the gene. The utility of this technique described as tagged microarray marker approach has been demonstrated for humans and pea (Flavell *et al.*, 2003; Ji *et al.*, 2004). However, Affymetrix GeneChips can also be used for SNP genotyping of a number of samples for one or more genes of interest as done in *Arabidopsis*, where the array AT412 was used for the study of variation in several genes (for example, *Eds16* (Cho *et al.*, 1999), *Rsf1* (Spiegelman *et al.*, 2000) and *FRI* (Nordborg *et al.*, 2002)).

The ultra-high-throughput resequencing technologies like 454/Roche, Solexa, AB SOLiD and HeliScope single molecule sequencer that have recently become available, and the SEQUENOM's MassArray genotyping system involving matrix-assisted laser desorption/ionization-time of flight mass spectroscopy can also be used for SNP discovery and detection. These technologies, however, do not fall within the scope of the present discussion on microarray-based markers. Efforts are also being made for microarray-based genome-wide capturing of exons for selective resequencing that may allow SNP genotyping (Hedges *et al.*, 2007).

Single feature polymorphisms

When labelled genomic DNA from two or more genotypes is separately hybridized to the same high-density oligonucleotide array, SFPs are detected as significant differences in hybridization signals among

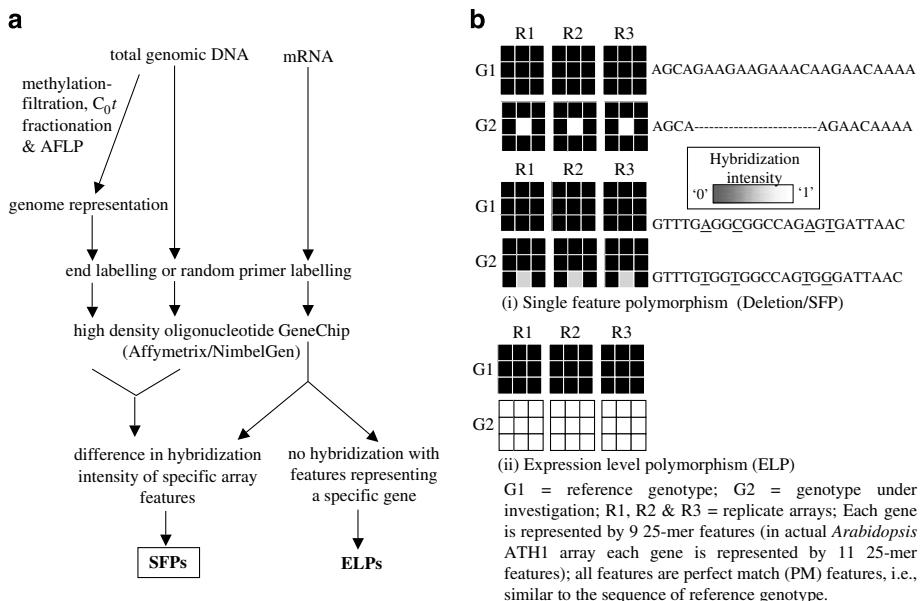


Figure 2 Schematic representation of (a) the procedure used for detecting SFPs and ELPs using conventional oligonucleotide expression GeneChip, and (b) difference in (i) hybridization intensities of the reference and the genotype under investigation due to deletion (upper panel), SFP (lower panel) and (ii) pattern of hybridization between the reference and the genotype under investigation due to ELP.

the genotypes used (Figure 2). In this assay, the oligonucleotides, available on the array, function as probes and are described as features, hence the term 'single feature polymorphism'. For visualizing the signal, whole genome DNA is either Dnase-1 treated and then end labelled with γ -p³² dCTP or else is directly labelled with biotin-14-dCTP. Sometimes, genomic DNA with reduced complexity or complementary RNA (cRNA) is also used to improve the power of SFP detection, although the use of cRNA also has some limitations (see later). The quantity of DNA needed for successful hybridization depends on the size of genome, therefore more DNA is needed for complex genomes (for example, 300 ng for *Arabidopsis*; 420 ng for rice; 6 μ g for wheat and so on). SFPs detected in this manner represent allelic variations ranging from SNPs to large deletions, although in the majority of studies, most SFPs have been found to be SNPs.

Designing of microarrays and their use for detection of SFPs/expression level polymorphisms

Expression oligonucleotide arrays that are often used for SFP technology include Affymetrix (<http://www.affymetrix.com>) GeneChips or Nimblegen (<http://www.nimblegen.com>) arrays, which could be either the catalogue microarray meant for a variety of uses or may be custom made for the intended use. However, microarray with only small probes (25 bp) are used for SFP technology, as these detect sequence polymorphisms with high level of specificity; probes that are 200–1000 bp long are not suitable, as they tolerate mismatches in molecular hybridization (Zhu and Salmeron, 2007). Also to ensure that only a single locus will hybridize to each feature, sequences that are likely to hybridize to multiple loci are eliminated (although repeat sequences and multicopy genes have also been used for SFP detection).

The oligonucleotide probes (features), needed for SFP analysis, can be designed either from genomic sequences (as done in *Arabidopsis*) or from cDNA/EST sequences (as done in barley). In an Affymetrix GeneChip often used for SFP analysis, each of a large number of genes is represented by a variable number of probes covering the entire gene sequence, thus constituting a probeset (the entire GeneChip having a number of probesets, which is equal to the number of genes to be sampled). Such a strategy provides an opportunity to assay multiple loci within each gene for detecting polymorphisms (that is, SFPs). However, when cRNA is used as a surrogate for genomic DNA, we may have to sample multiple tissues to take care of spatial/temporal expression of genes. Besides this, while using cRNA, we need to distinguish between SFPs and expression level polymorphisms (ELPs), as two genotypes may not have sequence variation (SFP) in a particular gene, but may still differ for its ELP. While in an ELP, all probes representing a probeset will give same hybridization affinity with a single cRNA sample and the hybridization intensity will differ only with different cRNA samples (probeset level polymorphism), but in case of SFPs, the affinity of a particular probe will differ from that of all other probes in a probeset for the same cRNA sample (probe level polymorphism; see Figure 2).

SFPs in simple genome of yeast

In the very first study involving SFPs, Winzeler *et al.* (1998) initially identified 3714 SFPs between two different strains of budding yeast. Subsequently, SFPs in yeast were put to a variety of uses including fine mapping and positional cloning of a QTL for high-temperature growth (Steinmetz *et al.*, 2002) and also for assessment of genealogical relationship among 14 strains of budding yeast (Winzeler *et al.*, 2003). In place of genomic DNA, cDNA/cRNA was also used to study SFPs in yeast, which allowed parallel identification of ELPs and SFPs (Brem *et al.*, 2002; Ronald *et al.*, 2005).

SFPs in complex genomes of seed plants

In the initial phase of SFP development, it was thought that the technique is suitable for only small genomes, as an increase in genome size leads to significant reduction in signal-to-noise ratio. However, it was later recognized that even in complex genomes, SFPs may be detected with reasonable accuracy, provided genome complexity is reduced during sample preparation, and replicating arrays involving multiple tissues/development stages are used in cases of hybridization with cRNAs (see later for details). Consequently, the technique was later used in a number of seed plants including those with moderately complex genomes (for example, *Arabidopsis* and rice) and also those with large and highly complex genomes (for example, maize, soybean, tomato, lettuce, barley and wheat). The results of SFP analysis in these plant species with complex genomes are summarized in Table 2. However, there are some limitations that make this technology less competitive with some of the recently developed ultra-high-throughput SNP technologies discussed earlier in this review (see later for details).

As evident from the information presented in Table 2, SFPs have been used for a variety of purposes including detection of marker-trait associations, which sometimes involved construction of a molecular map followed by QTL interval mapping. For instance, in *Arabidopsis*, bulk segregant analysis has been used for mapping of circadian and developmental genes (Hazen *et al.*, 2005), ion accumulation genes (Gong *et al.*, 2004) and light-responsive QTLs (Wolyn *et al.*, 2004). Similarly, in tomato, 17 SFPs were identified, which were tightly linked to a disease resistance locus (T Zhu, unpublished data; reported by Zhu and Salmeron, 2007). Other important examples of the utility of SFPs are the development of a map of 34 000 SFP loci representing 11 000 genes in maize (Zhu *et al.*, 2006) and that of 8500 SFP loci from 6000 genes in tomato (Salmeron and Zhu, 2007).

Strengths and limitations of SFP technology: It has been argued that for crop plants with large genomes carrying high proportion of repetitive DNA, SNP genotyping for association studies at the whole genome level becomes prohibitive. For instance, in maize, an estimated one million SNP markers are needed for genome-wide association studies (Gore *et al.*, 2007), although this number may be substantially reduced, if tagSNPs are used. This seems unnecessary because large proportion of these SNPs would belong to the non-coding regions. Under these circumstances, an attractive alternative is provided by SFP technology, which also allows coupling of genotyping with gene expression analysis. However,

Table 2 List of SFP studies conducted in various seed plants

Plant species	Resources used ^a	Salient features	References	
Wheat	GeneChip genome array and cRNA	SFP analysis used for mapping translocation breakpoints on wheat chromosome arm 1BS	Bhat et al. (2007)	
	Customized wheat oligonucleotide array and cRNA	44 SFPs for 7E (<i>Thinopyrum</i> and <i>Lophopyrum</i>) identified using alien substitution/addition lines	Buescher et al. (2007)	
	GeneChip genome array and cRNA	SFPs identified in 948 genes using two wheat varieties ('Eltan' and 'Oregon feed wheat')	Ling et al. (2006)	
	GeneChip genome array	Identified 100's of SFP markers and integrated them into an existing SSR map using DH population	Somers et al. (2008)	
	GeneChip genome array and cRNA	5376 SFPs in 'LaGrue' (<i>japonica</i>), and 25325 SFPs in 'RT0034' (<i>indica</i>), when compared with 'Cypess' (<i>japonica</i>)	Kumar et al. (2007)	
	GeneChip genome array and cRNA	64/46 SFPs detected from shoot/root datasets, respectively, when comparisons were made among 'Golden Promise' and 'Maythorpe'	Walia et al. (2007)	
	GeneChip genome array and cRNA	924 of 1257 genes assigned to chromosomes with the help of SFPs detected using CS-Betzes ditelo additions	Bilgic et al. (2007)	
	GeneChip genome array and cRNA	Mapped > 2000 transcript derived markers (TDMs; including both SFPs and GEMs) and 23738 expression QTLs using a DH population (Steptoe × Morex)	Potokina et al. (2008)	
	GeneChip Genome Array and cRNA	> 4000 SFPs identified among 'Steptoe' and 'Morex', and segregation studied using DH population (Steptoe × Morex)	Luo et al. (2007)	
	GeneChip genome array and cRNA	10 504 SFPs identified among 'Golden Promise' and 'Morex', also compared with known SNPs suggesting 67% sensitivity to the assay	Rostoks et al. (2005)	
Rice	GeneChip genome array and cRNA	2007 SFPs identified among 'Steptoe' vs 'Morex', 'Morex' vs 'Barke', and Oregon Wolfe Barley Dominant vs Recessive, and 80% were confirmed by direct sequencing	Cui et al. (2005)	
	GeneChip genome array and cRNA/methylation filtered/ C_0t filtrated DNA	Assessment of various target preparation and hybridization methodologies (for example, cRNA, methylation filtration, high C_0t and AFLP) using three diverse maize inbred lines	Gore et al. (2007)	
	GeneChip genome array	34 034 SFPs identified and mapped using an intervarietal mapping population; mapped loci validated through sequencing	Zhu et al. (2006)	
	CornChip0 (a customized array) and cRNA	36 196 SFPs identified among 'B73' (reference genotype) and three US maize lines: Mo17, Wf9-BG and W23	Kirst et al. (2006)	
	ATH1 genome array and gDNA	77 420 SFPs detected using 23 wild accessions of <i>A. thaliana</i> and reference strain, Col	Borevitz et al. (2007)	
	ATH1 genome array and cDNA	An integrated map prepared using 540 SFPs and 38 SSRs on 211 RILs ('Bay-0' × 'Sha'), was used to detect 36,871 expression QTLs	West et al. (2007)	
	ATH1 genome array and gDNA	16 000 SFPs detected among Col and Ler, also allowed assignment of 676 SFPs on the genetic map	Singer et al. (2006)	
	ATH1 genome array and cRNA	1–11 SFPs identified in 6728 probesets among 'Col' and 'Ler'	DeCook et al. (2006)	
	AtGenome1 and gDNA	~ 4000 SFPs identification among 'Col' and 'Ler'. Comparison of partial sequences from Ler with whole genome sequence of Col suggested >96% sensitivity	(Borevitz et al. 2003)	
	ATH1 Genome Array and gDNA	> 8000 SFPs detected using various accession (Ler, Kas-1, Lz-0, Bur-0, Nd-1 and Col) in different studies	Wolyn et al. (2004); Hazen et al. (2005); Werner et al. (2005a, b); West et al. (2006)	
Arabidopsis	ATH1 genome array and cRNA	1257 informative SFPs detected and ~ 600 SFPs could be mapped genetically	Kliebenstein et al. (2006)	
	ATH1 genome array and cRNA	Demonstrated that SFPs are not a significant source of variation and differences among accessions were predominantly due to ELPs	Gong et al. (2004)	
	ATH1 Genome Array and gDNA	SFPs detected in four probesets allowed identification of a 523 bp deletion within the second exon of <i>AtHKT1</i> using wild-type and mutant 'FN1148' DNA		
	Lettuce	GeneChip genome array and cDNA/gDNA/high- C_0t DNA	Assessment of various target preparation and hybridization methodologies (e.g., cDNA, total gDNA and high C_0t) are underway	Caldwell et al. (2006, 2007)
	Soybean	GeneChip genome array and gDNA	1500 putative SFPs identified using 'Dwight' (susceptible to aphids) and 'Dowling' (resistant, due to <i>Rag1</i>), and ~ 30 SFPs identified when two NILs for <i>Rag1</i> were compared with the recurrent parent 'Dwight'	Kaczorowski et al. (2007)
	Tomato	GeneChip genome array and gDNA	17 SFPs tightly linked to <i>FRI</i> disease resistance locus were identified using bulk segregant analysis approach	Zhu and Salmeron (2007)
		Custom-designed Affymetrix array and complexity-reduced gDNA	Used AFLP to reduce complexity for detection of > 2000 markers (30% were scored codominantly) using F_3 population	Oeveren et al. (2008)

Abbreviation: gDNA, genomic DNA.

^aVarious terminologies for Affymetrix GeneChips were used in the literature; to avoid confusion, a standard nomenclature provided by the manufacturers was used in this table.

to make SFP technology useful in crops with complex genomes, a complexity reduction and gene enrichment method needs to be developed for the preparation of target DNA that is used for hybridization. Unfortunately, initial results involving four different gene-enrichment methods in maize are not very encouraging, and there seems to be a need for further work for developing not only the improved gene-enrichment methods, but also the custom-designed arrays for SFP detection/genotyping (Gore *et al.*, 2007).

Several other limitations of using SFP technology in crops with complex genomes that have been recognized during the use of this technology include the following: First, while using cRNA as surrogate for genomic DNA, extensive replications with samples from multiple tissues are needed, thus increasing the cost per data point. Second, for achieving reasonable SFP detection power in a crop like maize, 20% or higher false discovery rate (FDR) needs to be allowed (Gore *et al.*, 2007), although for several inbred plant species with smaller genomes, a much lower FDR may be used without loss of power. Third, in an array-hybridization experiment in maize, the detection of probeset polymorphism (ELPs) was found to be more effective than the detection of probe polymorphism (SFP), thus limiting the resolution to gene level rather than to nucleotide level (Gore *et al.*, 2007; Zhu and Salmeron, 2007); even in *Arabidopsis*, it was shown that ELPs rather than SFPs are the major source of variation in SFP technology (Kliebenstein *et al.*, 2006). Fourth, SFP technology often fails to detect polymorphisms due to SNPs occurring at the edges of the oligonucleotide probes and mainly detects only those polymorphisms that are due to internal SNPs (at positions 6–15, as observed in maize and barley).

Efforts to improve SFP technology: As mentioned above, complexity reduction and gene enrichment of the target DNA is one approach to make SFP technology suitable for complex genomes. This has often been achieved through the use of cRNAs, as done in *Arabidopsis* and several crop plants, including barley, rice, maize and wheat. This allows simultaneous acquisition of data for SFP genotyping and expression analysis, and thus facilitates development of two different types of markers, SFPs and gene expression markers; the latter recorded as large differences in transcript levels between the parents of a mapping population. In *Arabidopsis*, both these types of markers were used for construction of a high-density SFP/gene expression marker map and for mapping of expression QTLs or ELPs (West *et al.*, 2006).

The other complexity reduction and gene-enrichment methods include methylation filtration, C₀t filtration and AFLP, but these methods offered only a modest improvement in power to detect SFPs, when used with maize genome (cf. Gore *et al.*, 2007). Recently, Salmeron and Zhu (2007) proposed a new enzyme-mediated genome complexity reduction method to detect what are described as 'gene-specific hybridization polymorphisms'. In this method, separate libraries of fragments produced using methylation sensitive/insensitive enzymes were used for hybridization and led to the detection of gene-specific hybridization polymorphism markers. This method has not yet been widely tested, but

its utility in maize suggested that it may also prove useful for other crops with large and complex genomes.

Single feature polymorphism detection/genotyping has also been improved through the use of improved statistical tools. For instance, the use of robustified projection pursuit involved differentiation of signal intensities between two genotypes, first at the level of probeset and then at the level of individual probes. This should help in removing copy number effects and should also allow distinction between SFPs and ELPs. Several other statistical tools and softwares are available, which would lead to improvement in SFP detection.

Methods have also been suggested to reduce the number of false positives and false negatives observed during SFP studies. The false positives are believed to result due to alternative splicing or polyadenylation, gene duplications, chance alignments with RNA from another region, gene expression markers (resulting from polymorphism(s) at *trans*-acting regulators and secondary structures in target DNA) or SNPs that occur immediately adjacent to the position of a 25mer probe (cf. Luo *et al.*, 2007; Zhu and Salmeron, 2007; Potokina *et al.*, 2008). The number of these false positives can be reduced by (i) sampling more than one tissues per genotype (Rostoks *et al.*, 2005; Luo *et al.*, 2007); (ii) excluding sequence duplications during array design, (iii) studying segregation pattern in the progeny of the cross made from genotypes under question (Luo *et al.*, 2007), (iv) adjusting stringency parameters (Luo *et al.*, 2007) and through (v) the use of replication arrays. Similarly, false negatives can be attributed to position (in nucleotides) of known SNP(s) on the corresponding array feature, and can be reduced by increasing feature density per gene (Rostoks *et al.*, 2005).

Diversity array technology

Diversity array technology is a high-throughput microarray hybridization-based technique that allows simultaneous typing of several hundred polymorphic loci spread over a genome without any previous sequence information about these loci (Jaccoud *et al.*, 2001; Wenzl *et al.*, 2004). The technique has also been shown to be reproducible and cost effective. Generally, 50–100 ng of genomic DNA is used for genotyping nearly 5000–8000 genomic loci in parallel in a single-reaction assay to discover polymorphic markers. The same platform is used for both discovery and scoring of markers, so that no specific assay for genotyping needs to be developed after marker discovery, except an initial assembly of all polymorphic markers (detected in a metagenome) into a single genotypic array. The genotyping array with only polymorphic markers thus developed is routinely used for genotyping (Huttner *et al.*, 2005).

How are DArT markers developed and scored?

Diversity array technology markers are polymorphic segments of genomic DNA that are present in a particular genomic representation and are identified through differential hybridization on a diversity 'genotyping array' that is developed specifically for this purpose. These markers are biallelic and dominant (presence vs absence) or co-dominant (two doses vs one dose vs absent). DArT technology involves initial development of a 'discovery array', which is then used to

identify polymorphic DArT markers that are assembled into a 'genotyping array'. The 'discovery array' is developed from a metagenome (pool of genomes representing the germplasm of interest) that is subjected to complexity reduction to reduce the level of repetitive DNA, as repetitive sequences interfere with DArT assays and do not contribute to polymorphic clones (Kilian *et al.*, 2005). Individual clones from genomic representation are amplified and spotted onto glass slides to give the desired 'discovery array' (www.DiversityArrays.com). Labelled genomic representations of individual genomes that were earlier included in the metagenome pool are then hybridized to this discovery array, and polymorphic clones (called DArT markers) thus detected are assembled into a 'genotyping array' for routine genotyping work (Figure 3). The software DArTsoft is used for analysis of hybridization intensities. The efficiency of identification of polymorphic DArT markers depends on the level of genetic diversity available within the metagenome pool that is used for developing the discovery array. For instance, only 5–10% of wheat and barley DArT clones and 25–30% of cassava DArT clones were found to be polymorphic (Huttner *et al.*, 2005).

Molecular basis of DArT polymorphisms

Diversity arrays generally detect polymorphisms due to single base-pair changes (SNPs) at the restriction sites of endonucleases, and InDels/rearrangements within restriction fragments (Jaccoud *et al.*, 2001). The complexity reduction method applied to DNA samples to be used for DArT analysis will determine the type of polymorphisms detected by DArT markers. For instance, if we use a methylation-sensitive restriction enzyme like *Pst*I, it will

presumably identify polymorphisms due to both sequence variation (SNP and InDel) and DNA methylation (Kilian *et al.*, 2005). Although the dynamic nature of methylation states may suggest instability of some of these markers, in barley, the majority (97%) of DArT markers from a *Pst*I/*Bst*NI representation were found to be stable (Wenzl *et al.*, 2004). Similarly, in *Arabidopsis*, while comparing 107 genome sequences of Ler strain with those of a control Col strain, ~90% of DArT markers from a representation generated with methylation-sensitive *Pst*I detected SNP variation and only the remaining <10% were attributed to methylation polymorphism (Wittenberg *et al.*, 2005).

Present status of DArT development in different crop plants

Diversity array technology markers have already been developed for a fairly large number of plant species, including some orphan crops, for which no molecular information is available (Huttner *et al.*, 2006; for details, see Table 3). The initial proof of concept was provided by using rice, which is one of the major cereal crops having relatively simple genome (Jaccoud *et al.*, 2001). This study was supplemented by a study of about 30 000 rice genomic DNA clones involving 14 different complexity reduction methods, each involving a different frequent cutter restriction enzyme along with methylation-sensitive *Pst*I, which is a rare cutter (Kilian *et al.*, 2005). Later, barley and other crops having more complex genomes were also used (Wenzl *et al.*, 2004; also see Table 3). As evident from Table 3, DArT markers have been developed now on large scale (for instance, in wheat >3000 markers) and were extensively utilized for the study of genetic diversity, preparation of integrated framework linkage maps and association mapping (White *et al.*, 2008). DArT markers have also been used for QTL mapping for *Fusarium* head blight in wheat and for leaf pubescence in barley (Rheault *et al.*, 2007; Wenzl *et al.*, 2007a).

RAD markers

More recently, a variety of microarrays (including tiling/cDNA/oligonucleotide arrays) have also been used to develop the so-called RAD markers for study of genome-wide variations associated with restriction sites for individual restriction enzymes. For this purpose, first a genome-wide library of RAD tags is developed from genomic DNA, which is then used for hybridization on to the chosen microarray to detect all restriction-site-associated variations in a single assay. The development of RAD tags involves the following steps: (i) digestion of genomic DNA with a specific restriction enzyme; (ii) ligation of biotinylated linkers to the digested DNA; (iii) random shearing of ligated DNA into fragments smaller than the average distance between restriction sites, leaving small fragments with restriction sites attached to the biotinylated linkers; (iv) immobilization of these fragments on streptavidin-coated beads; and (v) release of DNA tags from the beads by digestion at the original restriction sites. This process specifically isolates DNA tags directly flanking the restriction sites of a particular restriction enzyme throughout the genome. The RAD tags from each of a number of samples, when hybridized

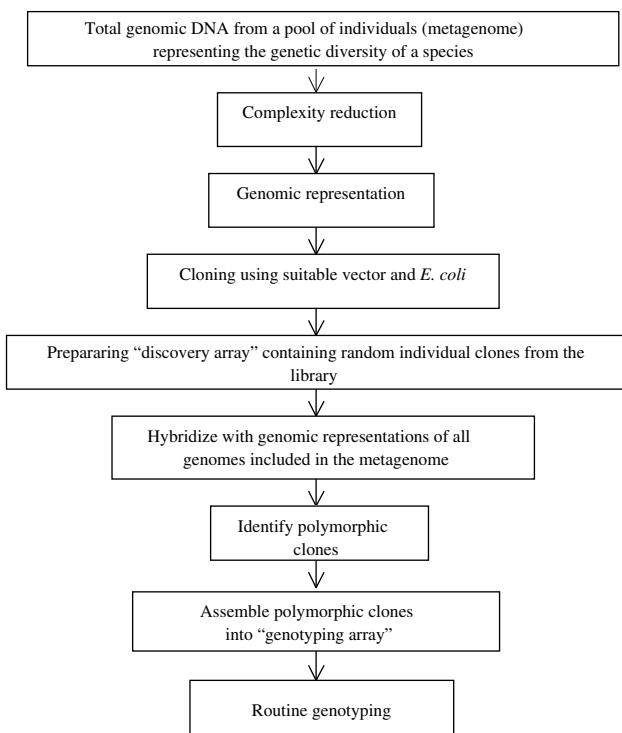


Figure 3 Schematic representation of steps involved in the development of 'genotyping array' for DArT technology.

Table 3 A list of DArT studies conducted in crop plants

Crop	DArT markers available	Salient features	Reference
Rice	>500	Initial proof of concept and validation of DArT using AFLP like complexity reduction methods to generate genomic representations from nine rice cultivars Evaluated genetic diversity in a general purpose rice gene pool and validated DArT for rice genotyping	Jaccoud et al. (2001) Xie et al. (2006)
Barley	>1000	Selected two complexity reduction methods (<i>PstI/TaqI</i> and <i>PstI/BstNI</i>) that generated the most polymorphic genomic representations and validated DArT markers. Constructed a genetic map comprising ~385 unique DArT markers and spanning 1137 cm, using barley cultivars Steptoe and Morex Constructed high-density consensus linkage map with ~3000 loci from the combined data set of 10 populations including 2085 DArT loci with SSR, RFLP and STS loci Used DArT and SSR markers for QTL mapping of <i>Fusarium</i> head blight (FHB) Mapped >600 DArT markers spanning ~2000 cm; identified 15 clustered loci for multiple resistance Mapped <i>Rsp1</i> on 3H and <i>Rsp2</i> , and <i>Rsp3</i> on 1H Constructed high-density genetic map with 558 SSR and 442 DArT markers. Revealed a high level of co-linearity and telomeric coverage of DArT markers on all seven chromosomes Tested suitability of DArT for bulk segregant analysis in barley; validated an aluminium (Al) tolerance locus on chromosome 4H Detected associations of DArT markers with resistance to wheat stem rust in <i>Aegilops sharonensis</i> and wild barley. Two significant marker associations were detected for resistance to the wheat stem rust race MCCF in the same bin as the <i>rpg4/Rpg5</i> complex on chromosome 7(5 H)	Wenzl et al. (2004) Wenzl et al. (2006) Rheault et al. (2007) Aslop et al. (2007) Lee and Neate (2007) Hearnden et al. (2007) Wenzl et al. (2007a) Steffenson et al. (2007)
Wheat	>3000	First successful report of DArT in bread wheat (<i>T. aestivum</i>). Mapped DArT, AFLP and SSR markers using 90 DH lines Developed complexity reduction methods to generate large frequency of polymorphic clones to produce large genotyping arrays. Constructed a framework linkage map of 93 DH lines using DArT, AFLP and SSR markers. Inferred origin of DArT markers from gene-rich regions Used >1000 DArT markers and developed a linkage map with 90 SSR and 543 DArT (278 bread wheat-derived probes and 265 durum wheat probes) markers using 176 RILs 242 DArT markers were used to study association with resistance against stem rust, leaf rust, yellow rust, powdery mildew, yield and yield-contributing traits Studied genetic diversity of UK, US and Australian wheat varieties using DArT markers	Akbari et al. (2006) Semagn et al. (2006) Wenzl et al. (2007b) Crossa et al. (2007) White et al. (2008)
Cassava	734	Tested three complexity reduction methods to select the two that generated largest frequency of polymorphic clones (<i>PstI/TaqI</i> : 14.6%, <i>PstI/BstNI</i> : 17.2%) to produce large genotyping arrays. Detected ~1000 polymorphic clones on two arrays; <i>PstI/TaqI</i> array was validated using 38 accessions	Xia et al. (2005)
<i>Arabidopsis</i>	—	Developed a genetic linkage map with 107 DArT markers using an F ₂ population derived from a Col × Ler cross. The quality and molecular basis of DArT markers were validated; SNPs accounted for nearly all DArT markers	Wittenberg et al. (2005)
Pigeonpea	—	Used 96 accessions for diversity analysis using DArT markers showing genetic relationships. Among eight complexity-reduction methods tested, <i>PstI/HaeIII</i> genomic representation generated largest frequency of DArT markers. Nearly 700 polymorphic clones were identified	Yang et al. (2006)
Sugarcane	>300	Using most effective complexity reduction method (<i>PstI/TaqI</i>) more than 300 markers were developed and used for diversity study and mapping	Heller-Uszynska et al. (2007); Uszynska et al. (2007)

Table 3 Continued

Crop	DArT markers available	Salient features	Reference
Sorghum	—	Genotyping of two mapping populations of sorghum 90562 \times ISCV745 and IS8525 \times 31945-2	Mace et al. (2007)
		Genome-wide scan with 935 DArT markers using 190 core collections; mapped 330 non-redundant DArT markers covering whole genome with \sim 1 marker each 3 cm; associations were found due to physical linkage	Bouchet et al. (2007)
		Performed whole genome scan using DArT markers; identified genomic regions associated with yield and adaptation	Jordan et al. (2007)
Banana	—	Used DArT markers for genetic and physical mapping	Kilian (2007)
		Used DArT markers for diversity analysis of 182 accessions	Hippolyte et al. (2007)
		Developed two arrays of 6000 clones each; markers distinguished diversity between and within A and B genomes	Huttner et al. (2007)

on to a microarray, allows high-throughput identification and/or typing of differential hybridization patterns. These markers have clear advantage over the existing marker systems (for example, restriction fragment length polymorphisms, AFLPs and DArT markers) that could assay only a subset of SNPs that disrupt restriction sites. RAD markers were successfully developed in a number of organisms including fruit fly, zebrafish, threespine stickleback, *Neurospora* (Lewis et al., 2007; Miller et al., 2007a,b) and will certainly find their way in most of the laboratories working on higher plants.

Statistical tools for microarray-based markers

The data generated due to hybridization of DNA on microarrays that are used for microarray-based DNA markers are invariably subjected to detailed statistical analysis. Therefore, a brief discussion of statistical tools used for this purpose is in order. For instance, before accepting an oligonucleotide probe as a marker during SNP/SFP/DArT genotyping assays, one would need to find out whether or not the difference between hybridization signals obtained due to two genotypes is significant. Different methods have been used for this purpose. For instance, methods are available for the estimation of expected hybridization intensities (\hat{I}), which can be compared with observed intensities (I), so that a significant difference (as shown by *t*-test) between the means of ratio \hat{I}/I for two genotypes involving the same probe will suggest that the probe is a polymorphic marker. Computer programs have been developed for this purpose (Ronald et al., 2005).

The most commonly used test statistics used for the analysis of data for array-based markers is the FDR threshold that is used to control type I error (Benjamini and Hochberg, 1995, 2000) and to minimize the number of false positives during SNP/SFP genotyping. This is achieved by first ranking the test statistics based on its *P*-value/*Q*-value, followed by checking/eliminating hypotheses on the basis of chosen FDR values (FDRs are calculated following Benjamini and Hochberg, 1995). This provides more leverage and less stringency relative to other methods, but it also reduces the number of false negatives significantly.

The utility of FDR as a statistical tool also depends on whether MB or ML method is used for genotyping. For instance, during SNP genotyping in *A. thaliana*, it was shown that at the same FDR, ML algorithm identifies

significantly more true SNPs than MB methods (Clark et al., 2007). It was inferred, therefore, that ML method can be used to complement and extend SNP predictions made through MB approach. Furthermore, ML algorithm helps us to detect polymorphic regions containing InDels and variation hotspots, where MB SNP detection algorithms generally fail to identify individual SNPs.

Other statistical methods that are specifically dedicated to SFP/SNP detection include 'robustified projection pursuit' (Cui et al., 2005), GeSNP (Greenhall et al., 2007) and two other methods developed by Luo et al. (2007). In a recent study, three methods were compared by testing their relative efficiencies to detect SFPs associated with known barley SNPs (already mapped using same population). It was concluded that different methods have their own advantages/disadvantages, and each is useful only under specific circumstances (Luo et al., 2007). For instance, the method developed by Winzeler et al. (1998) is appropriate for SFP prediction from genomic DNA microarray data only, because it assumes that DNA molecules uniformly hybridize onto a microarray chip across all genes. Similarly, the method developed by Ronald et al. (2005) can, however, utilize cRNA microarray data, but does not take into account the possible large variation in the abundance of transcripts belonging to different genes. Also, the method developed by Cui et al. (2005) relies solely on the data obtained from perfect match probes for SFP detection. It may, therefore, be important to take into account the assumptions associated with each method, while selecting for an analytical method to predict SFPs. Two methods recently developed by Luo et al. (2007) address most of these issues and can be used for SFP analysis in seed plants. A web-based program 'GeSNP', which was initially designed to detect SNPs from microarray data, was recently tested for its suitability to predict SFPs in mice, humans and chimpanzees (Greenhall et al., 2007), but its use for plants has yet to be tried.

Perspectives and conclusions

The use of microarrays for the development of DNA-based markers, similar to SNP, SFP, DArT and RAD markers, has provided technology platforms for medium- to ultra-high-throughput genotyping at a low cost (Table 4). These technologies have also made it possible to have access to polymorphic regions of a genome at

Table 4 A summary of differences among different array-based techniques for detecting DNA polymorphisms

Parameter	SNP	SFP	DArT	RAD
Sequence information	Required	Required	Not required	Not required
Resolution	High	High	Moderate	Moderate
PCR reaction	Required in some assays like MIP and GoldenGate	Not required	Required, at the time of development and hybridization	Not required
Cost	Cost effective	Cost effective	Cost effective	Cost effective
Number of markers per assay	High	High	Moderate	Moderate
Data type	Presence/absence	Presence/absence	Presence/absence	Presence/absence
Array used	Tag array on beads/glass, oligonucleotide array/GeneChip	High-density oligonucleotide array/GeneChip	Glass-spotted DNA microarray	Tiling microarray or GeneChip

genome-wide scale at a low cost, and have been shown to be particularly useful for genomes, where the level of polymorphism is low and large-scale genome sequencing is still time consuming and expensive (for example, soybean, tomato and bread wheat). In the future, these array-based marker technologies and other non-array-based high-throughput genotyping platforms (for example, SNplex and MassArray) will be used for a variety of studies including the development of high-density molecular maps, which may then be used for QTL interval mapping and LD-based association studies for functional and evolutionary studies.

It should also be recognized that most SFPs, DArT and RAD markers are actually SNPs, therefore these marker systems have all the merits of SNPs without the requirement of sequence-based discovery of SNPs. For several crops, GeneChips for SFP genotyping and the 'diversity microarrays' for DArT markers have already been developed making their subsequent use cost effective. Currently, the major competitors for microarray-based genotyping are Illumina and Affymetrix, and the genotyping platforms offered by them have their own merits and demerits, which have been briefly outlined in this review.

Genome-wide SNP genotyping has also been improved through the development of SNPchips, where genotyping of only tagSNPs instead of all known SNPs is done (a tagSNP is a representative SNP in a genomic region with high LD, therefore it allows identification of genetic variation without genotyping every SNP). It is also recognized that in many studies, one would like to examine many samples for a few SNPs rather than examining few samples for thousands of SNPs, unless the primary objective of such a study is to examine genome-wide pattern of nucleotide diversity. This will also bring down the cost of genotyping, where specific regions or candidate genes are examined for allelotyping.

It has also been realized by many institutions and establishments that instead of having a specific high-throughput genotyping system installed, one may like to get the genotyping done either at a regional or national facility created in the public sector or at a commercial undertaking providing this service. This is desirable because an individual institution would not have enough genotyping work for optimum utilization of the instrument and will also lose the option of choosing any one of the several available genotyping platforms. Also, not many institutions, which would like to have genotyping facility, can afford to have the ultra-high-throughput

genotyping platforms installed at site, although they have the funds for getting the genotyping done on contract.

Acknowledgements

Thanks are due to Indian National Science Academy (INSA) for award of an INSA Honorary Scientist Position to PKG, Council of Scientific and Industrial Research (CSIR) and Department of Biotechnology (DBT) for providing financial support to SR and RRM and to Head, Department of Genetics and Plant Breeding, CCS University, Meerut for providing facilities. Thanks are also due to anonymous reviewer for constructive suggestions, which lead to improvement of manuscript.

References

- Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S et al. (2006). Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Genet* **113**: 1409–1420.
- Aslop BP, Kilian A, Carling J, Pickering RA, Steffenson BJ (2007). DArT marker-based linkage analysis and inheritance of multiple disease resistance in a wild \times cultivated barley population. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P333.
- Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57**: 289–300.
- Benjamini Y, Hochberg Y (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J Educ Behav Stat* **25**: 60–83.
- Bilgic H, Cho S, Garvin DF, Muehlbauer GJ (2007). Mapping barley genes to chromosome arms by transcript profiling of wheat-barley ditelosomic chromosome addition lines. *Genome* **50**: 898–906.
- Bhat PR, Lukaszewski A, Cui X, Xu J, Svensson JT, Wanamaker S et al. (2007). Mapping translocation breakpoints using a wheat microarray. *Nucl Acids Res* **35**: 2936–2943.
- Borevitz JO, Hazen SP, Michael TP, Morris GP, Baxter IR, Hu TT (2007). Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **104**: 12057–12062.
- Borevitz JO, Liang D, Plouffe D, Chang H-S, Zhu T, Weigel D et al. (2003). Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res* **13**: 513–523.
- Bouchet S, Billot C, Deu M, Rami JF, Xia L, Kilian A et al. (2007). Whole genome scan and linkage disequilibrium evaluation on a sorghum core collection. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P365.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**: 752–755.

Buescher E, Cui X, Anderson JM (2007). Detecting single-feature polymorphisms on the 7e *Thinopyrum* chromosome using the wheat oligonucleotide array. In: *Plant & Animal Genomes XV Conference*. Town & Country Convention Center: San Diego, CA, P185.

Caldwell DG, Kozik A, Chen F, Truco MJ, Michelmore RW, Van Deynze A (2006). Development of a lettuce (*Lactuca sativa*) 6.6 million feature affymetrix GeneChip for massively parallel genotyping and gene expression analysis. In: *Plant & Animal Genomes XIV Conference*. Town & Country Convention Center: San Diego, CA, P183.

Caldwell DG, van Leeuwen H, Stoffel K, Chen F, Kozik A, Truco MJ et al. (2007). Development and exploitation of a lettuce (*Lactuca sativa*) 6.6 million feature affymetrix GeneChip for massively parallel genotyping and gene expression analysis. In: *Plant & Animal Genomes XV Conference*. Town & Country Convention Center: San Diego, CA, P174.

Cho RJ, Mindrinos M, Richards DR, Sapolisky RJ, Anderson M, Drenkard E et al. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat Genet* **23**: 203–207.

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P et al. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**: 338–342.

Collard BCY, Cruz CMV, McNally KL, Virk PS, Mackill DJ (2008). Rice molecular breeding laboratories in the genomics era: current status and future considerations. *Int J Plant Genomics* (in press).

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**: 169–196.

Crossa J, Burgueño J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M et al. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* **177**: 1889–1913.

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

DeCook R, Lall S, Nettleton D, Howell SH (2006). Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* **172**: 1155–1164.

Fan JB, Chee MS, Gunderson KL (2006). Highly parallel genomic assays. *Nat Rev Genet* **7**: 632–644.

Fan JB, Oiphant A, Shen R, Kermani BG, Garcia F, Gunderson KL et al. (2003). Highly parallel SNP genotyping. *Cold Spring Harb Symp on Quant Biol* **LXVII**: 69–78.

Flavell AJ, Bolshakov VN, Booth A, Jing R, Russel J, Ellis THN et al. (2003). A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach. *Nucl Acids Res* **31**: e115.

Gong JM, Waner DA, Horie T, Li SL, Horie R, Abid KB et al. (2004). Microarray-based rapid cloning of an ion accumulation deletion mutant in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* **101**: 15404–15409.

Gore M, Bradbury P, Hogers R, Kirst M, Verstege E, van Oeveren J et al. (2007). Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. *Crop Sci* **47**: 135–148.

Greenhall JA, Zapala MA, Cáceres M, Libiger O, Barlow C, Schork NJ et al. (2007). Detecting genetic variation in microarray expression data. *Genome Res* **17**: 1228–1235.

Gunderson KL, Kuhn KM, Steemers FJ, Ng P, Murray SS, Shen R (2006). Whole-genome genotyping of haplotype tag single nucleotide polymorphisms. *Pharmacogenomics* **7**: 641–648.

Gupta PK (2008). Ultrafast and low-cost DNA sequencing methods for applied genomics research. *Proc Natl Acad Sci India* (in press).

Hayes P, Szücs P (2006). Disequilibrium and association in barley: thinking outside the glass. *Proc Natl Acad Sci USA* **103**: 18385–18386.

Hazen SP, Borevitz JO, Harmon FG, Pruneda-Paz JL, Schultz TF, Yanovsky MJ et al. (2005). Rapid array mapping of circadian clock and developmental mutations in *Arabidopsis*. *Plant Physiol* **138**: 990–997.

Hearnden PR, Eckermann PJ, McMichael GL, Hayden MJ, Eglington JK, Chalmers KJ (2007). A genetic map of 1000 SSR and DArT markers in a wide barley cross. *Theor Appl Genet* **115**: 383–391.

Heller-Uszynska K, Carling J, Evers M, Piperidis G, Gilmour R, Aitken K et al. (2007). Diversity Arrays Technology (DArT) for high throughput, whole-genome molecular analysis in sugarcane. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P188.

Hippolyte I, Bakry F, Carreel F, Jenny C, Arnaud E, Roux N et al. (2007). Progress of *Musa* (banana and plantain) genetic diversity and genome analysis at Agropolis-CIRAD. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, W35.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW et al. (2007). Genome-wide *in situ* exon capture for selective resequencing. *Nat Genet* **39**: 1522–1527.

Huttner E, Caig V, Carling J, Evers M, Howes N, Uzynski G et al. (2006). New plant breeding strategies using an affordable and effective whole-genome profiling method. *BioVision Alexandria*, 26–29, April 2006, p73.

Huttner E, Risterucci AM, Hippolyte I, Caig V, Carling J, Evers M et al. (2007). Establishment of diversity arrays technology for whole-genome profiling of banana. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, W34.

Huttner E, Wenzl P, Akbari M, Caig V, Carling J, Cayla C et al. (2005). Diversity arrays technology: a novel tool for harnessing the genetic potential of orphan crops. In: Serageldin I, Persley GJ (eds). *Discovery to Delivery: BioVision Alexandria 2004, Proceedings of the 2004 Conference of The World Biological Forum*. CABI Publishing: UK, pp 145–155.

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK et al. (2008). High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theor Appl Genet*, doi 10.1007/s00122-008-0726-2.

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

Jones E, Ayele M, Ho J, Yourstone K, Chu W-C, Smith S (2007). The development of a set of maize SNP markers for the Illumina platform. *49th Annual Maize Genetics Conference*. Pheasant Run St Charles: Illinois, P162.

Jordan DR, Hammer GL, Rodgers D, Butler DG, Hunt CH, Collard B et al. (2007). Multi-population to mapping to increase genetic diversity and grain yield in sorghum. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P398.

Ji M, Hou P, Li S, He N, Lu Z (2004). Microarray-based method for genotyping of functional single nucleotide polymorphisms using dual-color fluorescence hybridization. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **548**: 97–105.

Kaczorowski K, Kim KS, Diers B, Hudson M (2007). Mapping of the *Rag1* aphid resistance locus of soybean with microarray-generated SNP markers. In: *Botany & Plant Biology 2007 Joint Congress*. Hilton Chicago, Chicago, Illinois, id 2072.

Khlestikina EK, Salina EA (2006). SNP markers: methods of analysis, ways of development, and comparison on an example of common wheat. *Russ J Genet* **42**: 585–594.

Kilian A (2007). Towards effective deployment of Diversity Arrays Technology (DArT) in banana genomics and sequen-

cing. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, W36.

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

Kim S, Misra A (2007). SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng* **9**: 289–320.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S et al. (2007). Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* **39**: 1151–1155.

Kim S, Zhao K, Jiang R, Molitor J, Borevitz JO, Nordborg M et al. (2006). Association mapping with single-feature polymorphisms. *Genetics* **173**: 1125–1133.

Kirst M, Caldo R, Casati P, Tanimoto G, Walbot V, Wise RP et al. (2006). Genetic diversity contribution to errors in short oligonucleotide microarray analysis. *Plant Biotechnol J* **4**: 489–498.

Kliebenstein DJ, West MAL, van Leeuwen H, Kim K, Doerge RW, Michelmore RW et al. (2006). Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics* **172**: 1179–1189.

Kumar R, Qiu J, Joshi T, Valliyodan B, Xu D, Nguyen HT (2007). Single feature polymorphism discovery in rice. *PLoS ONE* **2**: e284.

Lee SH, Neate SM (2007). Molecular mapping of *Rsp1*, *Rsp2*, and *Rsp3* genes conferring resistance to *Septoria* speckled leaf blotch in barley. *Phytopathology* **97**: 155–161.

Lewis JA, Shiver AL, Stiffler N, Miller MR, Johnson EA, Selker EU (2007). High density detection of restriction site associated DNA (RAD) markers for rapid mapping of mutated loci in *Neurospora*. *Genetics* **177**: 1163–1171.

Ling P, Campbell KG, Little LM, Skinner DZ (2006). Service and research for molecular markers development in USDA-ARS western-regional small grain genotyping laboratory. In: *Plant & Animal Genomes XIV Conference*. Town & Country Convention Center: San Diego, CA, P203.

Luo ZW, Potokina E, Druka A, Wise R, Waugh R, Kearsey MJ (2007). SFP genotyping from affymetrix arrays is robust but largely detects *cis*-acting expression regulators. *Genetics* **176**: 789–800.

Mace ES, Kilian E, Halloran K, Xia L, Collard B, Jordan DR (2007). Application of Diversity Arrays Technology (DArT) for sorghum mapping, diversity analysis and breeding. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P366.

Mardis ER (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet* **24**: 133–141.

McNally KL, Bruskiewich R, Mackill D, Buell CR, Leach JE, Leung H (2006). Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol* **141**: 26–31.

Meirmans PG, Lamothe M, Gros-Louis MC, Perinet P, Isabel N (2006). Using SNPstream technology to distinguish poplar species and their hybrids. In: *Plant and Animal Genomes XIV Conference*. San Diego, CA, P237.

Miller MR, Atwood TS, Eames BF, Eberhart JK, Yan YL, Postlethwait JH et al. (2007a). RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol* **8**: R105.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007b). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* **17**: 240–248.

Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Choudhary BP, Landegren U (1994). Padlock probes: circularising oligonucleotides for localized DNA detection. *Science* **265**: 2085–2088.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J et al. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* **30**: 190–193.

Oeveren JV, Hogers R, Verstege E, Wittenberg A, Leeuw L, Witsenboer H et al. (2008). Codominant SFP genotyping in tomato. In: *Plant and Animal Genomes XVI Conference*. San Diego, CA, P106.

Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M (2008). Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J* **53**: 90–101.

Reid AL, O'Sullivan H, Barker GLA, Berry S, Edwards KJ (2007). Towards high throughput genotyping in wheat using mapped SNPs: the second step. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P83.

Rheault ME, Dallaire C, Marchand S, Zhang L, Lacroix M, Belzile F (2007). Using DArT and SSR markers for QTL mapping of *Fusarium* head blight resistance in six-row barley. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, P335.

Ronald J, Akey JM, Whittle J, Smith EN, Yvert G, Kruglyak L (2005). Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. *Genome Res* **15**: 284–291.

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J et al. (2005). Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol* **6**: R54.

Salathia N, Lee HN, Sangster TA, Morneau K, Landry CR, Schellenberg K et al. (2007). Indel arrays: an affordable alternative for genotyping. *Plant J* **51**: 727–737.

Salmeron J, Zhu T (2007). Syngenta Participations AG. Methods for screening for gene specific hybridization polymorphisms (GSHPs) and their use in genetic mapping and marker development, WO 2007/005305 A1.

Semagn K, Bjornstad A, Skinnes H, Maroy AG, Tarkegne Y, William M (2006). Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. *Genome* **49**: 545–555.

Shah MR, Kumpatla SP, Mody M, Gupta M, Thompson SA (2003). The SNP-IT platform: an efficient single nucleotide polymorphism detection technology for plant molecular genetics. In: *Plant and Animal Genomes XI Conference*. San Diego, CA, P237.

Singer T, Fan Y, Chang H-S, Zhu T, Hazen SP, Briggs SP (2006). A high-resolution map of *Arabidopsis* recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet* **2**: e144.

Somers DJ, Jordan MC, Banks TW (2008). Single feature polymorphism discovery using the affymetrix wheat GeneChip. In: *Plant and Animal Genomes XVI Conference*. San Diego, CA, P272.

Spiegelman JL, Mindrinos MN, Fankhauser C, Richards D, Lutes J, Chory J et al. (2000). Cloning of the *Arabidopsis RSF1* gene by using a mapping strategy based on high-density DNA arrays and denaturing high-performance liquid chromatography. *Plant Physiol* **12**: 2485–2498.

Steffenson BJ, Olivera P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ (2007). A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. *Aus J Agric Res* **58**: 1–13.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JL, Oefner PJ, McCusker JH et al. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**: 326–330.

Steemers FJ, Gunderson KL (2007). Whole genome genotyping technologies on the BeadArray™ platform. *Biotechnol J* **2**: 41–49.

Syvänen AC (2005). Toward genome-wide SNP genotyping. *Nat Genet* **37** (Suppl.): S5–S10.

Uszynska KH, Carling J, Evers M, Piperidis G, Gilmour R, Aitken K et al. (2007). Diversity Arrays Technology (DArT) for high throughput, whole-genome molecular analysis in

sugarcane. In: *Plant and Animal Genomes XV Conference*. San Diego, CA, W188.

Walia H, Wilson C, Condamine P, Ismail AM, Xu J, Cui X et al. (2007). Array-based genotyping and expression analysis of barley cv. Maythorpe and Golden Promise. *BMC Genomics* **8**: 87.

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A et al. (2004). Diversity arrays technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* **101**: 9915–9920.

Wenzl P, Catizone I, Thomson B, Huttner E, Mantovani P, Maccaferri M et al. (2007b). A DArT platform for high-throughput profiling of durum wheat. *Plant and Animal Genome XV Conference*, Town & Country Convention Center: San Diego, CA, P263.

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E et al. (2006). A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genomics* **7**: 206.

Wenzl P, Raman H, Wang J, Zhou M, Huttner E, Kilian A (2007a). A DArT platform for quantitative bulked segregant analysis. *BMC Genomics* **8**: 196.

Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D (2005a). FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* **170**: 1197–1207.

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J et al. (2005b). Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc Natl Acad Sci USA* **102**: 2460–2465.

West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW et al. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* **175**: 1441–1450.

West MAL, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA et al. (2006). High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Res* **16**: 787–795.

White J, Law JR, MacKay I, Chalmers KJ, Smith JSC, Kilian A et al. (2008). The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. *Theor Appl Genet* **116**: 439–453.

Winzeler EA, Castillo-Davis CI, Oshiro G, Liang D, Richards DR, Zhou Y et al. (2003). Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**: 79–89.

Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ et al. (1998). Direct allelic variation scanning of the yeast genome. *Science* **281**: 1194–1197.

Wittenberg AHJ, van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ (2005). Validation of the high throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Mol Genet Genomics* **274**: 30–39.

Wolyn D, Borevitz JO, Loudet O, Schwartz C, Maloof J, Ecker JR et al. (2004). Light response QTL identified with composite interval and extreme array mapping in *Arabidopsis thaliana*. *Genetics* **167**: 907–917.

Xia L, Peng K, Yang S, Wenzl P, De Vicente MC, Fregene M et al. (2005). DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. *Theor Appl Genet* **110**: 1092–1098.

Xie Y, McNally K, Li C-Y, Leung H, Zhu Y-Y (2006). A high-throughput genomic tool: Diversity Array Technology complementary for rice genotyping. *J Int Plant Biol* **48**: 1069–1076.

Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P et al. (2006). Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by Diversity Arrays Technology (DArT). *Theor Appl Genet* **113**: 585–595.

Yu J, Buckler ES (2006). Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* **17**: 155–160.

Zhu T, Salmeron J (2007). High-definition genome profiling for genetic marker discovery. *Trends Plant Sci* **12**: 196–202.

Zhu T, Xia Y, Chilcott C, Dunn M, Dace G, Sessions A et al. (2006). Maize ultra high-density gene map for genome-assisted breeding. In: *48th Annual Maize Genetics Conference*, March 9–12, Asilomar Conference Grounds, Pacific Grove, CA, P181.