African Journal of Biotechnology Vol. 7 (4), pp. 331-341, 19 February, 2008
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2008 Academic Journals

Review

Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs): Emerging molecular marker tools for improving agronomic traits in plant biotechnology

Kwadwo Owusu Ayeh

Department of Plant and Environmental Sciences, The Norwegian University of Life Sciences (UMB), Postboks 5003, 1432, ÅS-NORWAY. E-mail: kwadwo.owusu.ayeh@umb.no. Tel: + 47-64965606.

Accepted 16 January, 2008

Expressed Sequence Tags (ESTs) and Single Nucleotide Polymorphisms (SNPs) are providing in depth knowledge in plant biology, breeding and biotechnology. The emergence of many novel molecular marker techniques are changing and accelerating the process of producing mutations in plant molecular biology research. This coupled with the availability of cheap sequencing techniques and access to a complete genome sequence has been shown to complement traditional marker –based approaches. Expressed Sequence Tags (ESTs) have provided an important source for the study of Single Nucleotide Polymorphisms (SNPs) in plants. SNP markers have become popular, partly because of their high density within the genome and also their ease with which they are characterized. This review also focuses on some methods used in genotyping SNPs.

Key words: Molecular markers, expressed sequence tags (EST), single nucleotide polymorphisms (SNPs).

INTRODUCTION

Mankind, in his quest to improve food quality, has been looking for variant forms of crops. Plant breeders and farmers have created new varieties through traditional plant breeding. In traditional plant breeding crosses between plants are performed. Sexual crossing of such nature are done in an uncontrolled manner and this often leads to a random combination of genes which results in new traits, some of which may be undesirable. Selection and careful evaluation of the offspring is therefore necessary.

Taditional plant breeding has gone through many phases, from the era of cross pollination between varieties of the same species, to hybridisation between different species and genus to overcome barriers imposed by combining species belonging to different families. However, traditional plant breeding is costly and time consuming and moreover, the selection and evaluation of the new varieties can take several years to achieve. With the discovery of the DNA, a new area of modern plant biotechnology begun. In plant breeding the development of molecular marker systems facilitated the selection and evaluation process greatly. These molecular tools have increased the speed and precision for achieving desired agronomic traits.

Restriction fragment length polymorphism (RFLPs) belongs to the first generation of hybridization-based markers developed in humans in the 1980s (Botstein et al., 1980; Demartinville et al., 1982) and thereafter used in plant research (Weber and Helentjaris, 1989). In RFLP, the variation (s) in the length of DNA fragments produced by a digestion of genomic DNAs and hybridisation to specific markers of two or more individuals of a species is compared (Kahl, 2001). RFLPs have been used extensively to compare genomes in the major cereal families such as rye, wheat, maize, sorghum, barley and rice (Bennetzen, 2000; Devos et al., 1993; Dubcovsky et al., 2001). Extensive studies have also been carried on genome relationships in the grass family using RFLP technique (Devos and Gale, 2000). The advantages of RFLPs include detecting unlimited number of loci, codominant and the use of probes from other species. However, RFLPs are expensive, time consuming and la-

bour intensive. PCR based marker systems are more rapid and requires less plant material. The first of PCR based marker was known as Rapid Amplified Polymorphic DNAs (RAPDs) and are produced by PCR using genomic DNA and arbitrary primers (Welsh and McClelland, 1990; Williams et al., 1990; Jacobson and Hedrén, 2007). However, the results from RAPDs may not be reproduced in different laboratories. RFLPs and RAPDs have been used to map or tag agronomically important genes including resistance genes against viruses, bacteria, fungi, nematodes and insects (Mohan et al., 1994; Mohan et al., 1997). Amplified Fragment Length Polymorphism (AFLPs) combines both PCR and RFLP and it is generated by digestion of PCR amplified fragments using restriction enzymes (Vos et al., 1995). For example, AFLPs have been used to assess the levels of genetic diversity within and between populations of Nordic meadow (*Festuca pratensis* Huds) (Fjellheim and Rognli, 2005). Recently, AFLP technique has been used in the analysis of pathogens which affect wheat kernels in Russia (Gannibal et al., 2007). AFLPs are highly reproducible and this enables rapid generation and high frequency of identifiable AFLPs, making it an attractive technique for identifying polymorphisms and for determining linkages by analysing individuals from a segregating population (Mohan, 1997).

Another class of molecular markers which depends on the availability of short oligonucleotide repeat sequences in the genome of plants is the simple sequence repeat (SSR) polymorphism or microsatellites (Tautz and Rentz, 1984; Hearne et al., 1992). SSR markers are fairly cheap and no sequence information is required for their detection. SSR gives good polymorphism as well as requiring only a small quantity of DNA to start with. However, it suffers a similar disadvantage to that of RAPDs in that its reproducibility in different laboratories may be low as described earlier. Polymorphisms can be detected in sequences which occur between two SSRs. These set of markers are known as Inter Simple Sequence Repeats (ISSR) (Zietkiewicz et al., 1994; Yang et al., 1996).

Currently plant biologists are exploiting the use of Expressed Sequence Tags (ESTs) and Single Nucleotide Polymorphisms (SNPs) as markers in gene discovery research. ESTs are short DNA sequences corresponding to a fragment of a complimentary DNA (cDNA) molecule and which may be expressed in a cell at a particular given time. ESTs are currently used as a fast and efficient method of profiling genes expressed in various tissues, cell types or developmental stages (Adams et al., 1991). One of the many interesting applications of ESTs database (dbEST) is gene discovery where many new genes can be found by querying the dbEST with a protein or DNA sequence. On the other hand, a Single Nucleotide Polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide- A-T-C or G- in the genome differs between members of a species (or between paired chromosomes in an individual). SNPs may occur in the coding, non-coding and intergenic regions of the genome, thus enabling the discovery of genes as a result of the differences in the nucleotide sequences.

However, other markers such as DNA Amplification Fingerprinting (DAF), Cleaved Amplified Polymorphic Sequences (CAPS), Specific Amplicon Polymorphism, Marker Assisted Breeding and Sequence Tagged Sites (STS) are all in use in modern plant breeding.

The emergence of these molecular markers in plants has greatly broadened our knowledge in understanding the genetic basis of important agronomic traits which have resulted in the characterization of many genes using existing information from known crop species. This has resulted in the improvement of important agronomic traits and the creation of genetic variation in the field of plant biotechnology over the last few years. Moreover, it has led to the development of many new technologies with the aim to further learn more about the plant genomes and how gene functions can be obtained from them. The development of the herbicide-resistant and insect-resistant crops made a significant impact in the agricultural biotechnology industry. Since then, little success has been recorded probably due to our lack of knowledge of the genome sequences from plants. Currently, there is a wealth of information in genomics and together with the reduction in cost of analysing the sequence of genomes, new methods have emerged for identifying candidate genes for improvement in plants (Gutterson and Zhang, 2004). The general understanding of plant growth, development, differentiation and defence has greatly been improved through the study of genome structure and function in some model plants such as *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000). This has further been boosted by the study of other model plants including rice (*Oryza sativa)*, maize, (*Zea mays*) and *Lotus japonicus* (Yu et al., 2002; Martienssen et al., 2004; Handberg and Stougaard, 1992). In addition, the use of *Populus tremula x Populus tremuloides* as a model plant has made it possible to study woody species (Wullschleger et al., 2002).

The purpose of this review paper is to summarize and present some molecular biology concepts and techniques used in discovering genes in plants, the use of ESTs and SNPs in gene discovery research. Furthermore, this review paper also looks at some methods used in SNP genotyping and the potential application of functional genomics and single nucleotide polymorphisms in plant biotechnology.

FUNCTIONAL GENOMICS

Functional genomics may be explained as understanding the function of genes and other parts of the genome. The ultimate goal in functional genomics is to discover new

genes and their function. This will lead to the creation of mutant databases, cDNA and genomic libraries and over expressing or silence lines (Chory et al., 2000). Some of the methods used in gene discovery include Global Transcript Profiling, Reverse Genetics, Map Based Cloning and Proteomics. These components of functional genomics mentioned above, together with the sequencing of the genome of plants, play an important role in gene discovery and comparative genomics (Rudd, 2003).

Global transcription profiling

For a gene to be functional, it must be transcribed, processed and translated into a functional protein. Expression analyses of genes are important part of gene function studies. Global transcript profiling involves the analysis of transcripts using hybridisation-based techniques such as Northern blotting, RNA dot blotting or microarrays. These methods have become popular in recent years in allowing the evaluation and elucidation of gene expression. A wide range of biological processes, including hormone and stress responses have been addressed using transcript profiling. Changes in gene expression during cell division, expansion, secondary wall formation, lignification and cell death have been demonstrated in poplar (Hertzberg et al., 2001). Gene expres-sion upon pathogen attack, abiotic stress and hormonal responses in *arabidopsis* has been examined through the use of transcriptional profiling (Cheong et al., 2000). Analyses through Northern blot hybridization revealed how gene expression differed in flood-tolerant FR13A rice compared to the flood-sensitive PBI rice type (Agarwal and Grover, 2005). In rice, serial analysis gene expres-sion (SAGE) showed that most of the highly expressed genes in the seedlings were housekeeping genes (Mstumura et al., 1999). By transcription profiling, genes expressing enzymes involved in the shikimate pathway, leading to the biosynthesis of phenylalanine have been shown to be transcriptionally upregulated in primary stems in *arabidopsis* (Ehlting et al., 2005). Advances in PCR technology have made the exact quantification of messenger RNA products possible. Quantitative real-time PCR (qRT-PCR) relies on the ability to progressively quantify fluorescence emitted from specific double-stranded DNA binding dyes or fluorophore-labelled probes that bind to the DNA thereby enabling quantification (Song et al., 2002). qRT-PCR has been used to show that there is an increase in amplified DNA transcripts of NaCl dependent transcription factors in *arabidopsis* roots (Yuanqing and Deyholos, 2006), determination of transgene copy number and mRNA levels (Chiang et al., 1996; Ingham et al., 2001; Leutenegger et al., 1999).

Reverse genetics

This technique involves the isolation and detection of a gene without a prior knowledge of the protein or nucleotide sequence. This technique leads to the establishment of a map position of the genes being isolated. Geneticists have used knockout mutation as a reverse genetic method to study gene function. However, knockout methods, which may be carried out using T-DNA or transposable element insertions, chemical or radioactively induced mutation, provide very few clues of the function of the gene in question (Bouche and Bouchez, 2001). In view of the problems that are encountered in knock out mutation techniques, over expression offers an alternative and complementary strategy to knockout analysis. This is because over expression methods are less affected by the situation where it becomes difficult to observe the expression of a gene being looked for. This difficulty in observing the expression of a gene in knockout mutation is referred to as functional redundancy (Zhang, 2003). Reverse genetics has led to a new molecular marker technique that takes advantage of the availability of DNA sequencing. This new method is called Targeting Induced Local Lesions IN Genomes (TILLING). This novel method has enabled researchers to investigate the functions of genes in plants. TILLING is thus a powerful reverse genetic method that allows the generation of an allelic variation for any target gene, including essential genes (Slade and Knauf, 2005). The process combines random chemical mutagenesis, which is readily used on plant and animals (Kodym and Afza, 2003), with PCR-based screening (Figure 1) of genome regions of interest (McCallum et al., 2000). TILLING has enabled the detection of thousands of mutations in hundreds of *Arabidopsis* genes (Till et al., 2003). Unlike conventional mutation breeding, in which the mutation frequency is unknown or estimated only from mutations carrying a visible phenotype, TILLING provides a direct measure of mutations induced (Slade et al., 2005).

Map based cloning

Map based cloning is a functional genomic approach for gene detection. It is also called positional cloning or map-assisted cloning (MAC). Map based cloning involves the cloning of a specific gene in the absence of a transcript or protein product (Gunther, 2001). The technique uses closely linked molecular markers with the desired gene. In MAC, knowledge about the chromosomal location of the gene, construction of a physical map and together with a genomic library is required (Young, 1990). The identification and sequencing of the clones in the genomic library that contain the genomic fragment corresponding to the area of interest on the physical map is then carried out. This process has been described as ‘chromosome walking’ and this strategy relies on an accurate position of the target gene on a genetic map. The tightly linked markers are used to isolate the clones containing these markers from a large-insert genomic

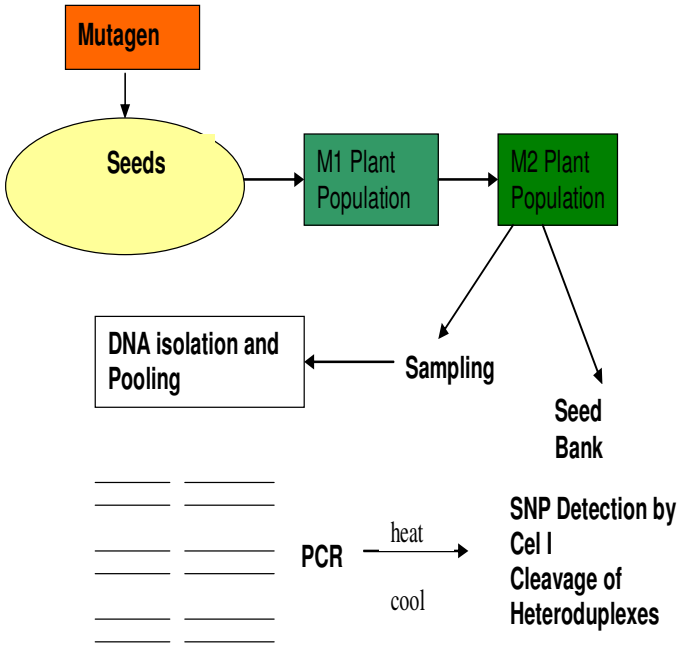

Figure 1. Seeds are induced to produce genetic variants by a chemical mutagen. Seeds are sown to produce plants up to the M2 population. M2 population seeds are sampled and DNA isolated from them and then pooled for PCR analysis. PCR products denatured are allowed to form heteroduplex and subsequently cleaved by Cel I and then analysed. (Source: Slade and Knauf, 2005).

library. The area of the genomic library that corresponds to the region of interest in the physical map is further subcloned, sequenced and analysed. The disadvantage in MAC is the time-consuming and manual work involved. Map-based cloning has led to the definition and isolation of Quantitative Trait Loci (QTL) as genetic loci where functionally different alleles segregate and cause significant effects on a quantitative trait (Salvi and Tuberosa, 2005). The use of MAC with a QTL approach has enabled the detection of a locus associated with NaCl tolerance in *Arabidopsis* (Quesada et al. 2002). Positional cloning is thus a promising method for identification of novel genes for which only the locus phenotype is known (Mindrinos et al., 1994: Grant et al., 1995).

Proteomics

This is one of the emerging and fast growing areas of research and though debatable; it is expected to provide a better understanding of regulation and function in biological systems than information revealed by genome sequencing projects. In addition to its definition as characterisation of proteins, proteomics extends to the studies of other protein properties such as expression levels, posttranslational modification and interaction with other proteins (Aebersold and Mann, 2003; Barbeir-Brygoo and Joyard, 2004). It consists of three main steps which include: 1) identification of proteins, their isoforms and their prevalence in each tissue, 2) characterising the biochemical and cellular functions of each protein and 3) analysis of protein regulation and its relation to other regulatory networks (Bertone and Synder, 2005). An important and useful technique in quantifying proteins is the High-Performance Liquid Chromatography (HPLC). HPLC is a fast and sensitive technique that provides the high resolution required for the detection and quantification of compounds in complex biological samples (Goulard et al., 2001). HPLC has been used to characterise and quantify the betalin pigments in the plant species of eight genera in the Amaranthaceae (Yizhong et al., 2005), analysis of polyamines in plants (Walter and Geuns, 1987) and the analysis of DNA methylation in plants (Johnston et al., 2005).

EXPRESSED SEQUENCE TAGS (ESTs)

Functional genomic approaches may provide powerful tools for identifying expressed genes. The discovery of novel genes and its possible utilization in modern plant breeding continue to engage the attention of most plant biologists. ESTs are short DNA molecules (300 - 500 bp) reverse-transcribed from a cellular mRNA population (MacIntosh et al., 2001). They are generated by large-scale single-pass sequencing of randomly picked cDNA clones and have proven to be efficient and rapid means to identify novel genes (Adams et al., 1991). ESTs thus represent informative source of expressed genes and provide a sequence resource that can be exploited for large-scale gene discovery (Whitefield et al., 2002).

By using comparative genomic approaches, the putative functions for some of these new cDNA clones may be found (Velculescu et al., 1995) and thereby constitute an important tool for a better understanding of plant genome structure, gene expression and function (Lopez at al., 2005). A large number of ESTs have been 2003; Nishiyama et al., 2003). Ji et al. (2006) studied and generated from various plant species including both mosses and cycads (Brenner et al., 2003; Kirst et al., 2003) model and crop plants like *A. thaliana*, rice, wheat and maize as well as other species such as gymnosperms, produced ESTs clones assembled into 375 contigs and 696 clusters when *Glycine soya* was subjected to saline conditions with the objective of mining salt tolerance genes. A number of ESTs have been generated and produced by studying genes involved in stress adaptation in the mangrove plant *Acanthus ebracteatus* Vahl (Huang and Madan, 1999; Nguyen et al., 2006), studying the genome of *Panax ginseng* C.A Meyer (Choi et al., 2005; Kim et al., 2006). Coles et al. (2005)

Table 1. Some specific plant EST databases with their websites.

Plant EST database	Website
B-EST barley database	http; //:pgrc.ipk-gatersleben.de/est/est/login
Chlamydomonas resource centre	http: //www.biology.duke.edu/chlamy.genome
Kazusa EST database	http: //www.kazusa.or.jp/en/plant/database
NCBI Unigenes	http://www.ncbi.nlm.nih.gov/UniGene/
PlantGDB	http://www.zmdb.iastate.edu/PlantGDB/
Solanaceae genomics network	http://sgn.cornell.edu/
TIGR Plant Gene Indices	http://www.tigr.org/tdb/tgi/plant.shtml
University Minnesota	http://www.ccgb.umn.edu/
Pscroph database	http://www.pscroph.ucdavis.edu
dbEST	http://www.ncbi.nlm.nih.gov/dbEST/index.html
REDB (Rice EST Database)	http://www.ncpgr.cn
Mendel-GFDb and Mendel-ESTS:	http://www.mendel.ac.uk/
US Mirror:	http://genome.cornell.edu/
EGENES	http://www.genome.jp/kegg-bin/create_kegg_
Sputnik	http://www.mips.gsf.de/proj/sputni
ESTree db	http://www.itb.cnr.it/estree
TbestDB	http://www.tbestdb.bcm.umontreal.ca

developed and characterised an EST database for quinoa (*Chenopodium quinoa* Willd) and demonstrated the usefulness of EST libraries as a starting point for detecting DNA sequence polymorphisms (SNPs). They compared cDNA sequenc-es of quinoa with sequences in the TIGR *A. thaliana* and GeneBank protein database. 67% of the quinoa proteins showed homology to *Arabidopsis* proteins with putative function, 18% had no significant matches, 9% had significant homology to *Arabidopsis* proteins with no known function and 6% sharing significant homology with plant proteins from species other than *Arabidopsis.* According to the dbEST release (September, 2007), there are currently over 46 million ESTs belonging to both plants and animals. Many of these dbESTs have their websites where they can be assessed (Table 1). Although there is no real substitute for a complete genome sequence, EST sequencing certainly avoids the biggest problems associated with genome size and the accompanying retrotransposon repetitiveness (Tang et al., 2003).

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs)

In the past, molecular markers were mainly based on genomic DNA (Gupta et al., 2002). The DNA could be-long to transcribed or non- transcribed regions of the genome (Gupta and Rustgi, 2004). However, there is a rapid accumulation of new markers, obtained from a large number of cDNA clones in a variety of plants and the accumulation of a large number of expressed sequence tags (ESTs) in the public databases. These markers are gradually gaining popularity compared to the older DNA-based markers. The genomic-based markers such as RFLP have their limitations. RFLPs are usually labour intensive and fairly expensive. In addition, RFLPs require large quantities of DNA and often polymorphism may be very low (Williams et al., 1991). RAPDs have a low reproducibility both between and within laboratories, mak-ing them less attractive. AFLPs require only a small amount of DNA; however, they are dominant and the technique itself can be challenging. Many of these markers mentioned above require the use of gel electrophoresis, are time consuming and expensive. The emerging new molecular markers such as Single Nucleo-tide Polymorphism (SNPs) techniques do not always need these electrophoresis-based assays. SNPs are excellent markers for association mapping of genes controlling complex traits and provide the highest map resolution (Botstein and Risch, 2003; Brookes, 1999; Bhattramakki et al., 2002). SNPs are robust in usage and polymorphisms are identifiable and there are several methods that can be used to detect them. SNPs are the most frequent type of variation found in DNA (Brookes, 1999; Cho et al., 1999) and their discovery together with insertions/deletions has formed the basis of most differences between alleles. SNPs can thus be explained as any polymorphism between two genomes that is based on a single nucleotide exchange. In plants, studies on the occurrence and nature of SNPs are beginning to receive considerable attention, particularly in *Arabidopsis* where over 37, 000 SNPs have been identified through the comparison of two accessions (Jander et al., 2002). It has been reported in maize that there occurs a frequency of one non-coding SNP per 31 bp and 1 coding SNP per 124 bp in 18 maize genes assayed in 36 inbred lines (Ching et al., 2002).

In a related transcriptome-based molecular marker

technique, ESTs has been used to detect both length and sequence polymorphisms (Brown et al., 2001). Development of new SNPs include re-sequencing of PCR amplicons with or without pre-screening, electronic SNP (eSNP) discovery in shotgun genomic libraries and eSNP discovery in expressed sequence tag (EST) libraries (Rafalski, 2002). A number of EST collections have been used to describe and detect SNPs in maize (*Zea mays* L.) (Ching et al., 2002) and Soybean (Glycine max L. Merr.) (Zhu et al., 2003). In another studies using cassava (*Manihot esculenta* Crantz), Lopez et al. (2005) have developed strategies for detecting SNPs from ESTs of cassava. Lopez and co-workers adopted a two-way approach using bioinformatics-based analysis of the available ESTs and by PCR based analysis of non-coding sequences. They arranged a total of 11, 954 ESTs from 5 cassava cultivars into 1,875 contigs and 3,175 singletons using the StackPack software (Lopez et al., 2005). Using a PCR approach, EST sequence from the 5'-ends of genes was used. The estimated frequency for intra-cultivar SNPs was one per 905 bp and one per 1,032 bp for the inter-cultivar SNPs. For SNPs detection derived from 3' ESTs, they detected a total of 136 SNPs and the frequency of one per 66 bp. They concluded that the number of SNPs in the 5'-EST, where there is a high proportion of coding regions, was low. SNPs outside the coding regions were high in the 3'-end of selected ESTs and in BAC end sequences. The latter two methods allowed the evaluation of the presence of non-coding SNPs in several cultivars. They further observed that the non-coding regions can accumulate a greater number of polymorphisms and that not all genes accumulate SNPs at the same time. They observed two groups of genes: 1) those containing a relatively high number of SNPs (more than 6) and those with few or no SNPs.

SNP haplotypes (the specific pattern and order of alleles on a chromosome) may be detected in genomes and this may give us information as to whether they are in Linkage Disequilibrium (LD), that is if alleles at one locus are not randomly assorted with alleles at another locus (Borecki and Suarez, 2001). This phenomenon can be seen in populations that have experienced a bottleneck and inbreeding during domestication and may ultimately exhibit extensive LD (Hudson, 2001).

SNPs genotyping

There are several methods that are used in genotyping SNPs. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) (Griffin and Smith, 2000) and the detection of heteroduplex through sequences by denaturing high performance liquid chromatography (DHPLC) (Underhill et al., 1996) and DNA microarrays (Hacia and Collins, 1999) are some of the effective methods for the detection of SNPs. The principle of MALDI-TOF is briefly described here. The MALDI-TOF method utilises the principle of primer extension for the generation of products of SNP because it is easy to design and robust. The principle is shown in Figure 2 below. According to Sauer (2006), the principle involves the reaction of a PCR product with an extension primer that is chosen upstream of the SNP under investigation, ddNTPs and a thermostable DNA polymerase that will eventually yield allele specific products. The polymerase extends the 3'–end of the primer by specifically incorporating nucleotides that are complementary to the DNA template. The extension reaction terminates at the first nucleobase in the template where a nucleotide occurs that is complementary to one of the ddNTPs in the reaction mix. A thermostable DNA polymerase employed in a temperature-cycled reaction leads to linear amplification of the extended primers. The products are then incorporated into the MALDI-TOF column for SNPs genotyping.

However, these may be expensive to use and unavailable to laboratories in the under and developing world. A less expensive method that can be used for development of new SNPs could be the Cleaved-Amplified Polymorphic Sequence (CAPS) marker (Maeda et al., 1990). In CAPS, a set of oligonucleotide primers complementary to a sequence, which is known within a locus, are synthesized. Then using the PCR technique, the primers are used to amplify part of a locus in say two different organisms. The amplification products are subjected to restriction endonucleases to identify RFLPs among the different organisms (Konieczny and Ausubel, 1993; Lyamichev et al., 1993). The use of CAPS markers have also been exploited for the detection of single nucleotide polymorphisms using polymearse chain reaction (PCR) and restriction endonuclease (Jarvis et al., 1994; Michaelis and Amasino, 1998). With CAPS, Morales et al. (2004) found 11 SNPs (32%) in the coding regions and remaining 23 (68%) located in the non-coding regions, either in 3'UTR or sequences that were seen in the introns. Both synonymous and non-synonymous changes were observed in the coding regions. In the former, 7 SNPs (64%) in which 6 of them had the SNP located in the third nucleotide position of the codon were detected. The writers were, however of the view that their data could not be used as a basis in determining SNP frequency because only coding sequences, the neighbouring 3'UTR and intron regions were examined. The implication here is that the actual SNP frequency might have been underestimated. It is interesting to note, however that, the results of Morales et al. (2004) compared favourably with those obtained in analysing synonymous and non-synonymous proportions in maize. They further compared their work with that of Zhu et al. (2003), whose work on soybean identified more SNPs in non-coding introns and 3'UTRs. They thus submitted that the discovery of SNPs should be targeted

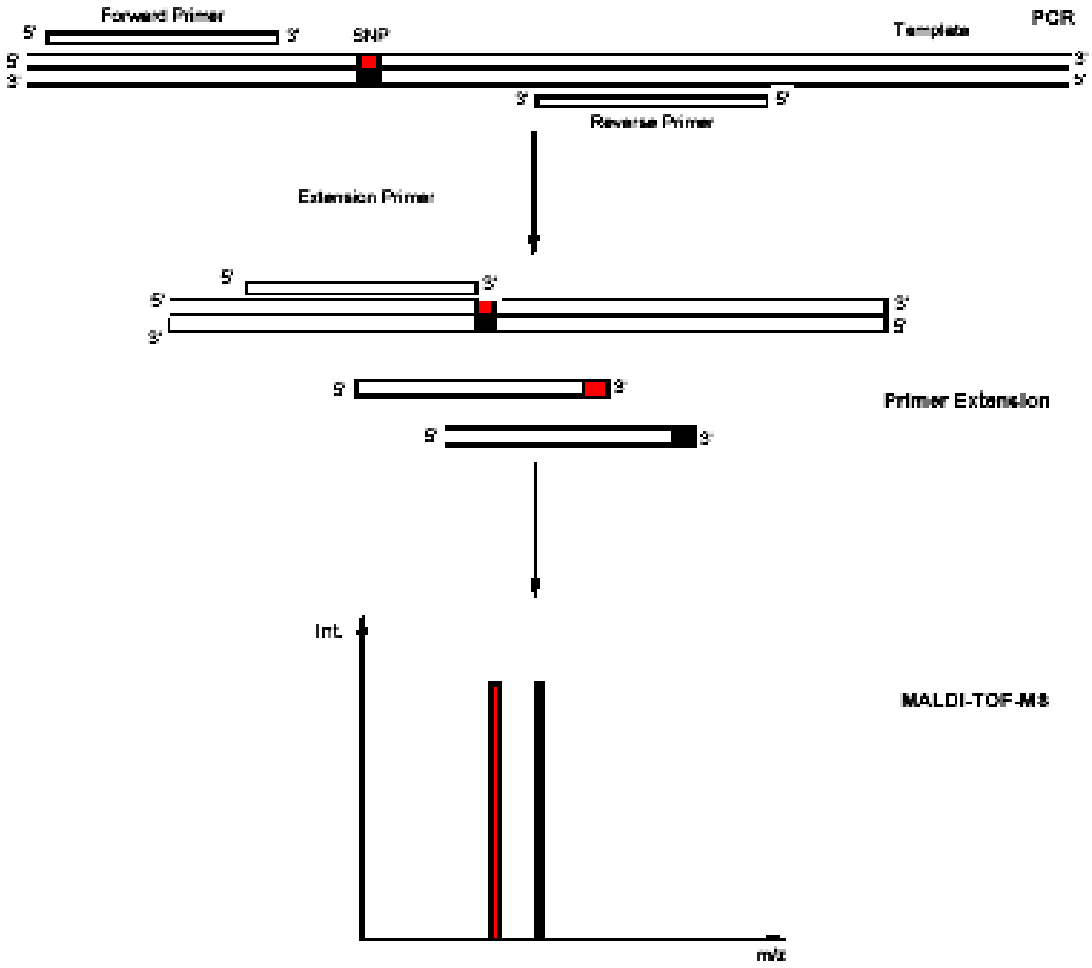

Figure 2. The principle of the primer extention reaction method is shown here. A DNA polymerase extends an oligonucleotide upstream of a SNP with a set of ddNTPs on a PCR product resulting in allelic-specific products for MALDI analysis. (Source: Sauer, 2006).

to the 3'UTRs of cDNA or to positions where introns are known to be located. In maize, SNP frequency has been found to be high and occur in every 28 - 124 bp (Tenaillon et al., 2001). SNP spoly-morphisms can also be identified from many individuals through direct sequencing of DNA segments (amplified by PCR) (Gaut and Clegg, 1993). Rafalski (2002) described this method of SNPs detection where PCR primers were designed to amplify 400-700 bp segments of DNA, which are derived from genes of interest or ESTs (Rafalski, 2002).

The single-strand conformation polymorphism analysis is one of the methods used for SNPs detection (Orita et al., 1989) and has been reviewed by Shi (2001). In single-strand conformation polymorphism, portions of the DNA with expected polymorphisms are first amplified by PCR. PCR products are denatured and this creates single-stranded DNAs, which are separated on a nonde-naturing polyacrylamide gel. Fragments are generated with a single base modification and forms a different conformer which migrates differently when compared with wild type DNA as explained in Figure 3. DNA micro-array technology can be described as a hybridisation-based method that can be used to detect SNPs. In microarrays, thousands of oligonucleotides at extremely high density are applied to a micro-scale format solid support in an ordered array. The DNA sample under investigations for SNPs detection is PCR amplified to incorporate fluorescent labelling and then hybridized to the solid support array. The perfect matches between the oligonucleotides immobilised on the support system and that of the amplified PCR products, may give stronger fluorescent signals over mismatched combina-tions, which give a weaker signal. There are a number of examples where DNA micro-array has been used to genotype SNPs (Wang et al., 1998; Bai et al., 2007).

In cases where the actual polymorphic sequence is not known, it is still possible to detect SNPs and insertions or deletions in the polymorphic sequences (Murphy et al.,

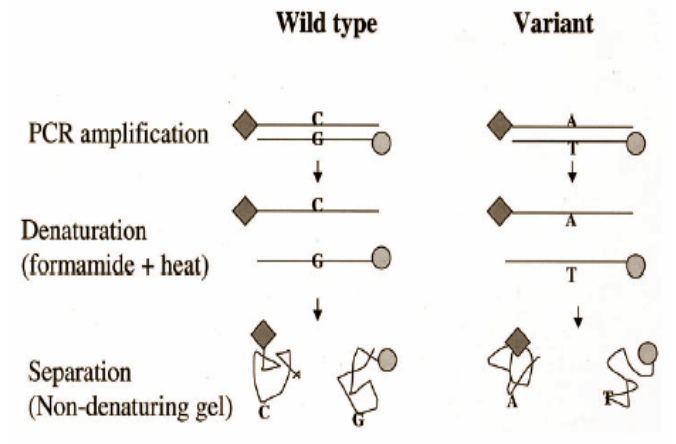

Figure 3. Single-strand conformation polymorphism analysis. Single-stranded DNAs were produced by denaturation of the PCR products and separated on a nondenaturing polyacrylamide gel. A fragment with a single-base modification generally forms a different conformer and migrates differently when compared with wild type. (Source: Shi, 2001).

2003). An-Ping Hsia et al. (2005) reported a novel technique known as the Temperature Gradient Capillary Electrophoresis (TGCE), which is a high-throughput method in discovering SNPs, insertions and deletion polymorphisms. TGCE principle is based on the formation of heteroduplex DNA. In heteroduplex DNA, two single- nucleotide chains from different origin form a double stranded (duplex) molecule. In a homoduplex however, single- stranded nucleic acid molecule containing inverted repeats may anneal to each other, so that double-stranded regions are generated which are linked by single stranded loops. Detection of heteroduplex is obtained through comparison of migration of homo /heteroduplex under different temperatures. At certain temperatures, homo/heteroduplex DNA molecules exhibit different electrophoretic mobilities. Heteroduplex DNA molecules have lower melting temperatures than homoduplex DNA molecules. This is due to the presence of mismatches in the former that results in the formation of secondary structures at lower temperatures. The writers showed that heteroduplex molecules exhibit slower electrophoretic mobilities than their homoduplex counterparts. They further demonstrated that they could detect a single SNP in amplicons of 800 bp using TGCE.

CONCLUSION AND FUTURE PERSPECTIVES

The extensive application of markers from the expressed region of the genome in plants, gives hope to plant biologists in that it is of extreme relevance in improving crops. Thus functional genomics is providing methods for profiling of transcripts, proteins and metabolites, and this may facilitate a much more characterization of plant status with changing environmental conditions (Wissel et al., 2003). Functional genomics driven by EST gene discovery project may significantly advance our understanding of the complexity of biological and celluar processes required for growth and development (Arpat et al., 2004). SNP technology may allow us to locate and detect candidate ESTs associated with agronomic traits and obtain a transcript map, which can be directly compared with an earlier detected quantitative trait loci. Single Nucleotide Polymorphism and indels are essential source of polymorphic markers for studying association of agronomic traits, which may be tightly linked to markers. SNPs based markers are thus becoming the marker system of choice because they are easily developed from sequence data, highly reproducible as reported by Rafalski (2002). It is envisaged that the knowledge gained from the understanding of plant functional genomics, ESTs and SNPs may find important application in breeding, agronomic practice and ecosystem research, especially in developing countries.

REFERENCES

Adams MD, Kelly JM, Gocayne JD, Dubnick M, Polymeropoulus MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991). Links complementary DNA sequencing: expressed sequence tags and human genome projects. Science: 252: 1651-1656.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408: 823-826.

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004). Functional genomics of cell elongatiobn in developing cotton fibers. Plant Mol. Biol. 54: 911-929.

Bai SL, Zhong XB, Ma LG, Zheng WJ, Fan LM, Wei N, Deng XW (2007). A simple and reliable assay for detecting specific nucleotide sequence in plants using optical thin-film biosensor chips. Plant J. 49 (2): 354-366.

Bennetzen JL (2000). Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell 12(7): 1021-1029.

Bertone P, Snyder M (2005). Prospects and challenges in proteomics. Plant Physiol. 138: 560-562.

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC, Tingey SV, Rafalski A (2002). Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol. Biol. 48: 539-547.

Borecki IB, Suarez BK (2001). Linkage and association: basic concepts. Adv. Genet. 42: 45-66.

Botstein B, white RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in map using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

Botstein D, Risch N (2003). Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nat. Genet 33: 228-237.

Bouche N, Bouchez D (2001) Arabidopsis gene knockout phenotypes wanted. Curr. Opin. Plant Biol. 4: 111-117.

Brenner ED, Stevenson DW, McCombie RW, Katari MS, Rudd SA, Mayer KFX, Palenchar PM, Runko SJ, Twigg RW, Dai GW, Martienssen RA, Benfey PN, Coruzzi GM (2003). Expressed sequence tag analysis in Cycas, the most primitive living seed plant. Genome Biol. 4: R78.

Brookes AJ (1999). The essence of SNPs. Gene 234:177-186.

Brown GR, Kadel EE III, Bassoni DL, Kiehne KL, Temesgen B, van Buijtenen JP, Sewell MED, Marshal KA, Neal DB (2001). Anchor reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genomics. 159: 799-809.

Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S (2000). Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in Arabidopsis. Plant Physiol. 129: 661-677.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet. 3: 19.

Chiang PW, Song WJ, Wu KY, Korenberg JR, Fogel EJ, Van Keuren ML, Lashkari D, Kurnit DM (1996). Use of a fluorescent-PCR reaction to detect genomic sequence copy number and transcriptional abundance. Genome 129: 863-872.

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubel E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat. Genet. 23: 203-207.

Choi DW, Jung JD, Ha YI, Park HW, In DS, Chung HJ, Liu JR (2005). Analysis of transcripts in methyl jasmonate-treated ginseng hairy roots to identify genes involved in the biosynthesis of ginsenosides and other secondary metabolites. Plant Physiol 129: 1095-1106.

Chory J, Ecker JR, Briggs S, Caboche M, Coruzzi GM, Cook D, Dangl J, Grant S, Guerinot ML, Henikoff S, Martienssen R, Okada K, Raikhel NV, Somerville, CR, Weigel D (2000). National Science Foundation-Sponsored Workshop report: "The 2010 Project" – Functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. Plant Physiol. 123(2): 423-425.

Coles ND, Coleman CE, Christensen SA, Jellen EN, Stevens MR, Bonifacio A, Rojas-Beltran JA, Fairbanks DJ, Maughan PJ (2005). Development and use of an expressed sequence tag library in quinoa (*Chenopodium quinoa* Willd.) for the discovery of single nucleotide polymorphisms. Plant Sci. 168: 439-447.

Demartinville B, Wyman AR, White R, Francke U (1982). Assignment of the 1st Random Restriction Fragment Length Polymorphism (RFLP) locus (D14S1) to a region of human chromosome-14. Am. J. Hum. Genet. 34(2): 216-226.

Devos KM, Gale MD (2000). Genome Relationships: the grass model in current research. Plant Cell. 12: 637-646.

Devos KM, Atkinson MD, Chinoy CN, Francis HA, Harcourt RL, Koebner RMD, Liu CJ, Masojc P, Xie DX, Gale MD, (1993). Chromosomal rearrangements in the rye genome relative to that of wheat. Theor. Appl. Genet. 5: 673-680.

Dubcovsky J, Ramakrrishna W, SanMiguel PJ, Busso CS, Yan YL, Shiloff BA, Bennetzen JL (2001). Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol. 125: 1342-13-1353.

Ehlting J, Matheus N, Aeschliman SD, Li E, Hamberger B, Cullis FI, Zhuang J, Kaneda M, Mansfield DS, Samuels L, Ritland K, Ellis EB, Bohlmann J Douglas JC (2005). Global transcript of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of Fibre differentiation. Plant J. 42: 618-640.

Fjellheim S, Rognli OA (2005). Genetic diversity within and among Nordic meadow fescue (*Festuca pratensis* Huds.) cultivars determined on the basis of AFLP markers. Crop Sci. 45 (5): 2081-2086.

Gannibal PB, Klemsdal SS, Levitin MM (2007). AFLP analysis of Russian Alternaria tenuissima populations from wheat kernels and other hosts. Eur. J. Plant Pathol. 119: 175-182.

Gaut BS, Clegg MT (1993). Nucleotide polymorphism in the Adh1 locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics 135: 1091-1097.

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1995). Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. Science 269: 843-846.

Griffin TJ, Smith LM (2000). Single-nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. Trends Biotechnol. 18: 77-84.

Goulard F, Diouris M, Deslandes E, Floc'h JY (2001). An HPLC method for the assay of UDP-glucose pyrophosphorylase and UDP-glucose-4-epimerase in Solieria chlordalis (Rhodophyceae). Phytochem. Anal. 12: 363-365.

Gunther K (2001). The Dictionary of Gene Technology. Wiley-VCH Verlag GmbH, D- 69469, Weinheim, pp: 941.

Gupta PK Rustgi S (2004). Molecular markers from the transcribed /expressed region of the genome in higher plants. Funct. Integr. Genomics 4, 139-162.

Gupta PK, Varshney RK, Prasad M (2002). Molecular markers: principles and methodology. In: Jain SM, Brar, DS, Ahloowalia, BS (eds). Molecular techniques in crop improvement. Kluwer, Dordrect.

Gutterson N, Zhang ZZ (2004). Genomics applications to biotech traits: a revolution in progress? Curr. Opin. Plant Biol. 7: 226-230.

Hacia JG, Collins FS (1999). Mutational analysis using oligonucleotide microarrays. J. Med. Genet. 36: 730-736.

Handberg K, Stougaard J (1992). Lotus japonicus, an autogamous, diploid legume species for classical and molecular genetics. Plant J. 2: 487-496.

Hearne CM, Ghoh S, Todd JA (1992). Microsatellites for linkage analysis of genetic traits. Trends Genet 8: 288-294.

Hertzberg M, Sievertzon M, Aspeborg H, Nilsson P, Sandberg G, Lundeberg J (2001). cDNA microarray analysis of small plant tissue samples using a cDNA tag traget amplication protocol. Plant J. 25(5): 585-591.

Hsia An-Ping, Wen TJ, Chen HD, Liu Z, Marna D, Yandeau N, Wei Y, Guo L, Schnable PS (2005). Temperature gradient capillary electrophoresis (TGCE)-a tool for the high-throughput discovery and mapping of SNPs and IDPs. Theor. Appl. Genet, 111: 218-225.

Huang A, Madan A (1999). CAP3: A DNA sequence assembly program. Genome Res. 9: 868-877.

Hudson RR (2001). Linkage disequilibrium and recombination. In Handbook of Statistical Genetics. Edited by Balding DJ, Bishop M, cannings Cchichester: John Wiley and sons, Ltd., pp. 309-324.

Ingham DJ, Beer S, Money S, Hansen G (2001). Quantitative real-time PCR assay for determining transgene copy number in transformed maize. Biotechniques 31: 132-140.

Jacobson A, Hedrén M (2007). Phylogenetic relationships in Alisma (Alismataceae) based on RAPDs, and sequence data from ITS and trnL. Pl. Syst. Evol. 265: 27-44.

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002). *Arabidopsis* map-based cloning in the plant-genome era. Plant Physiol. 129: 440-450.

Jarvis P, Lister C, Szabo V, Dean C (1994). Integration of CAPS markers into the RFLP map generated using recombinant inbred lines of *Arabidopsis thaliana*. Plant Mol. Biol. 24: 685-687.

Ji W, Li Y, Li J, Dai CH, Wang X, Bai X, Cai H, Yang L, Zhu Y (2006). Generation and analysis of expressed sequence tags from NaCl-treated *Glycine soya*. BMC Plant Biology, 6:4.

Johnston JW, Harding K, Bremmer DH, Souch G, Green J, Lynch PT, Grout B, Benson EE (2005). HPLC analysis of plant DNA methylation: a study of critical methodological factors. Plant Physiol. Biochem. 43: 844-853.

Kim MK, Lee B-S, In J-G Sun H, Yoon JH, Yang D-C (2006). Comparative analysis of expressed tags (ESTs) of ginseng leaf. Plant Cell Rep., 25: 599-606.

Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R (2003). Apparent homology of expressed genes from wood –forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. Proc Nactl Acad Sci USA, 100: 7383-7388.

Kodym A, Afza R (2003). Physical and Chemical mutagenesis. Methods Mol Biol 236: 189-204.

Konieczny A, Ausubel FM (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4(2), 403-410.

Leutenegger CM, Mislin CN, Sigrist BE, Markus U, Hofmann-Lehmann R, Lutz H (1999). Quantitative real-time PCR for the measurement of

feline cytokine mRNA. Vet. Immunol. Immunopathol. 71: 291-305.
Lopez C, Piegu B, Cooke R, Delseny M, Tohme J, Verdier V (2005). Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 110: 425-431.
Lyamichev V, Brow MAD, Dahlberg JE (1993). Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science 260: 778-783.
MacIntosh GC, Wilkerson C, Green PJ (2001). Identification and analysis of analysis of Arabidopsis expressed sequence tags characteristic of non-coding RNAs. Plant Physiol. 127(3): 765-776.
Maeda M, Uryu N, Murayaman N, Ishii H, Ota M, Tsuji K, Inoko H (1990). A simple and rapid method for HLA-DP genotyping of digestion of PCR amplified DNA with allele specific restrictions endonycleases. Hum. Immunol. 27: 111-211.
Martienssen RA, Rabinowicz PD**,** Shaughnessy A, McCombie WR (2004). Sequencing the maize genome. Curr. Opin. Plant. Biol. 7, 102-107.
McCallum CM, Comai L, Green EA, Henikoff S (2000). Targeting induced local lesions in genomes (TILLING) for plant functional genomics. Plant Physiol. 123: 439-442.
Michaelis SD, Amasino RM (1998). A robust method for detecting single nucleotide polymorphism functional analysis. Brief Bioinform. 6: 44-56.
Mindrinos M, Katagiri F, Yu GL, Ausubel FM (1994). The *Arabidopsis thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. Cell 78: 1089-1099.
Mohan M, Nair S, Bentur JS, Prasada Rao U, Bennet J (1994). RFLP and RAPD mapping of the rice *Gm2* gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theor. Appl. Genet. 87: 782-788.
Mohan M, Nair S, Bhagwat A, Krishna TG. Yano M, Bhatia CR, Sasaki T (1997). Genome mapping molecular markers and marker assisted selection in crop plants. Mol Breed 3:87-103.
Mohan M, Sathyanarayanan PV, Kumar A, Srivastava MN, Nair S (1997). Molecular mapping of resistance-specific PCR-based marker linked to a gall midge resistance gene (*Gm4t*) in rice. Theor. Appl. Genet. 95: 777-782.
Morales M, Roig E, Monforte AJ, Arus P, Garcia-Mas J (2004). Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). Genome 47: 352-360.
Mstumura H, Nirasawa S, Terauchi R (1999). Transcript profiling in rice (*Oryza sativa* L) seedlings using serial analysis of gene expression (SAGE)- Plant J. 20(6): 719-726.
Murphy KM, Hafez MJ, Philips J, Yarnell K, Gutshall KR, Berg KD (2003). Evaluation of temperature gradient capillary electrophoresis for detection of the Factor V Leiden mutation. Mol. Diagn. 7: 35-40.
Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M (2003). Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant 'evolution. Proc. Natl. Acad. Sci. USA, 100(13): 8007-8012.
Orita M, Suzuki Y, Sekiya T, Hayashi K (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics, 5: 874-879.
Quesada V, Garcia Martinez S, Piqueras P, Ponce MR, Micol JL (2002). Genetic architecture of NaCl tolerance in Arabidopsis. Plant Physiol, 130: 951-963.
Rafalski A (2002). Applications of single nucleotide polymorphism in crop genetics. Curr. Opin. Plant Biol. 5: 94-100.
Rudd S (2003). Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. 8: 321-329.
Salvi S, Tuberosa R (2005). To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 10: 6.
Sauer S (2006). Typing of single nucleotide polymorphisms by MALDI mass spectrometry: principles and diagnostic applications. Clin. Chim. Acta 363: 95-105.
Shi MM (2001). Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. Clin. Chem. 47(2): 164-172.
Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat. Biotechnol. 23: 75-81.
Slade AJ, Knauf VC (2005). Tilling moves beyond functional genomics into crop improvement. Transgenic Res. 14: 109-115.
Song P, Cai CQ, Skokut M, Kosegi BD, Petolino JF (2002). Quantitative real-time PCR as a screening tool for estimating transgene copy number in WHISKERS^{TM}-derived transgene maize. Plant Cell. Rep. 20: 948-954.
Tang W, Lou X, Nelson A, Collver H, Kinken K (2003). Functional genomics of wood quality and properties. Genome Prot. Bioinform. 1: 263-278.
Tautz D, Rentz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nature 322: 652-656.
Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L). PNAS 98: 9161-9166.
Till BJ, Reynolds SH, Green EA, Codomo CA, Enns LC, Johnson JE, Burtrner C, Odden AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. Genet. Res. 13: 524-530.
Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996). A pre Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc. Natl. Acad. Sci. USA. 93: 196-200.
Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995). Serial analysis of gene expression sequence. Science, 270: 484-487.
Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23: 4407-4414.
Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R et al (1998). Large- scale identification, mapping, mapping and genotyping of single nucleotide polymorphisms in the human genome. Science, 280: 1077-82.
Weber D, Helentjaris T (1989). Mapping RFLP loci in maize using B-A translocations. Genetics 121: 583-590.
Welsh J, McClelland M (1990). Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18: 7213-7218.
Whitefield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson M, Soares MB, Robinson GE (2002). Annotated expressed sequence tags and cDNA microarrays for studies of brain and behaviour in the honey bee. Genome Res. 12: 555-566.
Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531-6535.
William MNV, Pande N, Nair S, Mohan M, Bennett J (1991). Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa L.*) genomic DNA. Theor. Appl. Genet. 82: 489-498.
Wing R, Fang Z, Coe EH (2006). Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map. Crop Sci. 46: 12-21.
Wissel K, Petterson F, Berglund A, Janson S (2003). What affects mRNA levels in leaves of field-grown aspen? A study of developmental and environmental influences. Plant Physiol. 133: 1190-2011.
Wullschleger SD, Jansson S, Taylor G (2002). Genomics and Forest biology: Populus emerges as the perennial favorite. Plant Cell 14(11): 2651-1655.
Yang W, De Olivear AC, Schertz GK, Bennetzen JL (1996). Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. Crop Sci. 36: 1669-1676.
Yizhong C, Mei S, Harold C (2005). HPLC characterization of Betalins from plants in the Amaranthaceae. J. Chromatogr. Sci. 43(9): 454-460.
Young ND (1990). Potential applications of map-based cloning to plant pathology. Physiol. Mol. Plant Pathol. 37: 81-94.
Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ, Cao ML, Liu J, Sun JD, Tang JB, Chen YJ, Huang XB, Lin W, Ye C, Tong W, Cong LJ, Geng JN, Han YJ, Li L, Li W, Hu GQ,

Huang XG, Li WJ, Li J, Liu ZW, LiL., Liu JP, Qi QH, Liu JS, Li L, Li T, Wang XG, Lu H, Wu TT, Zhu M, Ni PX, Han H, Dong W, Ren XY, Feng XL, Cui P, Li XR, Wang H, Xu X, Zhai WX, Xu Z, Zhang JS, He SJ, Zhang JG, Xu JC, Zhang KL, Zheng XW, Dong JH, Zeng WY, Tao L, Ye J, Tan J, Ren XD, Chen XW, He J, Liu DF, Tian W, Tian CG, Xia HG, Bao QY, Li G, Gao H, Cao T, Wang J, Zhao WM, Li P, Chen W, Wang XD, Zhang Y, Hu JF, Wang J, Liu S, Yang J, Zhang GY, Xiong YQ, Li ZJ, Mao L, Zhou CS, Zhu Z, Chen RS, Hao BL, Zheng WM, Chen SY, Guo W, Li GJ, Liu SQ, Tao M, Wang J, Zhu LH, Yuan LP, Yang HM (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). Science 296(5565): 79-92.

Yuanqing J, Deyholos MK (2006). Comprehensive transcriptional profiling of NaCl-stressed Arabidopsis roots reveals novel classes of responsive genes. BMC Plant Biol. 6: 25.

Zhang ZJ (2003). Overexpression analysis of plant transcription factors. Curr. Opin. Plant Biol. 6: 430-440.

Zhu YL, Song QJ, Hyten DL, Van Tasell, CP, Matukumali, LK, Grim DR, Hyatt, SM, Fickus EW, Young ND, Cregan PB (2003). Single-nucleotide polymorphisms in soybean. Genetics, 163: 1123-1134.

Zietkiewicz E, Rafalski A, Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored Polymerase Chain Reaction Amplification. Genomics 20: 176-183.