EUROPEAN ACADEMIC RESEARCH
Vol. II, Issue 1/ April 2014

Impact Factor: 3.1 (UIF)
DRJI Value: 5.9 (B+)

ISSN 2286-4822
www.euacademic.org

Molecular Markers in Plants for Analysis of Genetic Diversity: A Review

MUHAMMAD IDREES[1]
Department of Biotechnology and Genetic Engineering
Kohat University of Science and Technology (KUST)
Kohat, Pakistan
MUHAMMAD IRSHAD
Department of Biotechnology
Bacha Khan University
Charsadda, Pakistan

Abstract:

The development and utilization of molecular markers for the exploitation and identification of plant genetic diversity is one of the most key developments in the field of molecular genetics studies. Various types of techniques are used to estimate genetic diversity such as dominant markers (Random Amplified Polymorphic DNA (RAPD), DNA amplification fingerprinting (DAF), Arbitrarily primed polymerase chain reaction (APPCR), Inter-simple sequence repeat (ISSR) and Amplified Fragment Length Polymorphism (AFLP), and co-dominant markers (Restriction Fragment Length Polymorphism (RFLP), Simple sequence repeats (SSRs), Sequence characterised amplified regions (SCARs), Cleaved amplified polymorphic sequence (CAPS), Expressed sequence tags (ESTs), Single Nucleotide Polymorphisms (SNPs) and sequence tagged sites (STSs). Today, new techniques are frequently being developed. No such techniques are ideal yet these fulfill all requirements needed by plant researchers. Each technique has its own advantages and limitations. The present review highlights a basic description of various molecular techniques that can be utilized for the analysis of plant genetic diversity and DNA fingerprinting.

[1] Corresponding author: idreesbiotech@yahoo.com

Key words: Genetic diversity, molecular markers, RAPD, SSR, ISSR, ESTs, STS

Molecular markers

Molecular marker is a sequence of DNA, which are located with a known position on the chromosome (Kumar 1999), or a gene whose phenotypic expression is frequently easily discerned and used to detect an individual, or as a probe to mark a chromosomes, nucleus, or locus (King and Stansfield 1990; Schulmann 2007). Markers show polymorphism, which may arise due to alteration of nucleotide or mutation in the genome loci (Hartl and Clark 1997) and make it possible to identify genetic differences between individual organisms or species (Collard et al. 2005). Molecular markers are used in many different areas such as genetic mapping, paternal tests, detect mutant genes which are connected to hereditary diseases, cultivars identification, marker assisted breeding of crops, population history, epidemiology and food safety, population studies (Hartl and Jones 2005). Collard et al. (2005) studies that genetic marker used to identify genetic variation among individual species or organisms. Schhulmann (2007) studies that genetic marker used to construct linkage maps and genetic diversity. Molecular marker also used in many different areas that are including detecting mutant genes, genetic mapping, epidemiology and for population studies (Hartl and Jones 2005).

The practice in agriculture has since the introduction of molecular makers in the 1980s broaden and simplified both for commercial and scientific uses to acknowledge new information such as large number of agronomic and disease resistance traits are available in major crop species (Phillips and Vasil 2001; Jain et al. 2002; Gupta and Varshney 2004). A conventional breeding programme thus involves crossing whole genomes, followed by selection of the superior recombinants from among the several segregation products. Indeed, such a

procedure is time consuming and laborious, involving several crosses, several generations, and careful phenotypic selection, and the linkage drag (tight linkage of the undesired loci with the desired loci) may make it further difficult to achieve the desired objective. Today many plant breeders utilize molecular markers to proof and identify desirable traits of important plants. Various plant researchers are used molecular markers to study genetic polymorphism and linkage maps construct (Schulmann 2007). There are many advantages of molecular markers compared with morphological and biochemical markers. For instance, when comparing morphological markers with molecular markers, expression of such markers is influenced both by dominant-recessive relationship and epistatic-pleiotropic interaction. The use of such markers has been criticized since the study of genetic variation among germplasm using morphological characteristics is labors and time consuming (Cooke 1984). When comparing with biochemical markers, their use is now limited due to large scale analysis and low level of genetic variation in a given species and subspecies (Rafinski and Babik 2000).

Ideal Desirable Properties of Molecular Markers

- Easily available
- Assay is rapid and easy
- Reproducible and highly polymorphic
- Co-dominant inheritance and recurrent occurrence in genome
  Selectively neutral to environmental conditions
- Data exchange between different laboratories should be easy.

It is really complicated to obtain molecular marker of above criteria. Depending on the type of study undertaken, a marker system can be recognized that would fulfill the above characteristics. Various types of molecular markers are used to estimate DNA polymorphism and are classified as

hybridization-based markers and polymerase chain reaction (PCR)-based markers. In hybridization based markers DNA profiles are visualized by hybridizing the restriction endonuclease digested DNA fragment, to a labelled probe, which is a DNA fragment of known sequence while PCR based markers involve in vitro amplification of particular DNA sequences with the help of specifically or arbitrarily chosen oligonucleotide sequences (primers) and a thermostable DNA polymerase enzyme. The amplified DNA fragments are separated by electrophoresis and banding patterns are detected by different methods such as staining (using ethidium bromide dye) and autoradiography. With the advent of thermostable DNA polymerase the use of PCR in research and clinical laboratories has increased tremendously. PCR is extremely sensitive and operates at a very high speed. Its application for diverse purposes has opened up a multitude of new possibilities in the field of molecular biology (Saiki et al. 1985; Saiki et al. 1988).

Types of molecular markers

Various types of molecular markers are available and new methods are frequently being developed. Today, no method is ideal for all applications so scientist teams must weigh both pros and cons of methods when starting a new project (Harlt and Jones 2005).

Molecular markers are grouped after their different abilities of showing homozygosity (dominant marker) or heterozygosity (co-dominant marker) (Hartl 1988). The most commonly used dominant DNA marker for genetic diversity in plants are: Random Amplified Polymorphic DNA (RAPD; Williams et al. 1990), DNA amplification fingerprinting (DAF; Caetano-Anolles et al. 1991), Arbitrarily primed polymerase chain reaction (AP-PCR; Welsh and McClelland 1990), Inter-simple sequence repeat (ISSR; Zietkiewicz et al. 1994) and

Amplified Fragment Length Polymorphisms (AFLP; Vos et al. 1995), whereas the most common used co-dominant markers are: Restriction Fragment Length Polymorphisms (RFLP; Botstein et al. 1980), Microsatellites (SSR; Akkaya et al. 1992); Sequence characterised amplified regions (SCAR; Paran and Michelmore 1993), Cleaved amplified polymorphic sequence (CAPS; Konieczny and Ausubel 1993), Expressed sequence tag (EST; Adams et al. 1991) and Single Nucleotide Polymorphism (SNP; Jordan and Humphries, 1994) and sequence tagged sites (STS; Olsen et al. 1989). Both dominant and co-dominant markers can be use to detect DNA polymorphism, which further used to assess the level of genetic variation in diverse populations and can indicate for instance population history, patterns of migration, and breeding structure (Hartl and Jones 2005).

Dominant DNA Markers

Arbitrarily Sequence Markers

Random Amplified Polymorphic DNA (RAPD), Arbitrarily Primed PCR (AP-PCR), and DNA Amplification Fingerprinting (DAF) have been collectively termed Multiple Arbitrary Amplicon Profiling (MAAP) (Caetano-Annolles 1994). These three techniques were first use to amplify any species DNA fragments without prior sequences information. The difference among MAAP techniques include modifications in amplification profiles by changing primer length, sequence and annealing temperature (Caetano-Anolles et al. 1992), the thermostable DNA polymerase (Bassam et al. 1992), the number of PCR cycles (Caetano-Anolles et al. 1991; Welsh and McClelland 1991; Micheli et al. 1993; Jain et al. 1994), enzymatic digestion of template DNA or amplification products (Caetano-Anolles et al. 1993) and alternative methods of fragment separation and staining. These three techniques produce markedly different amplification profiles, varying from quite simple (RAPD) to

highly complex (DAF) patterns. These marker techniques are quick, easily generated by PCR and require no prior sequence information.

The RAPD markers are small oligo nucleotide primers usually 10 bp in length of arbitrary sequence to generate bands profiles. These primers bind to the complementary sequences along the genome and PCR amplification occurs when the regions between the opposing primer sites are within amplifiable distances. The basic technique of RAPD involves

(i) isolation of highly pure DNA,
(ii) addition of single arbitrary primer,
(iii) polymerase chain reaction (PCR),
(iv) The resulting PCR products are generally resolved on 1.5- 2.0% agarose gels and stained with ethidium bromide (EtBr) or polyacrylamide gels in combination with either AgNO3 staining (Huff et al. 1993; Vejl, 1997; Hollingsworth et al. 1998), or fluorescently labeled primers or nucleotides (CorleySmith et al. 1997; Weller and Reddy 1997) or radioactivity (Pammi et al. 1994) are sometimes used.

Despite its low resolving power, the simplicity and low cost of agarose gel electrophoresis has made RAPD more rapid and popular than both AP-PCR and DAF.

Inter-Simple Sequence Repeat (ISSR)

Inter-Simple Sequence Repeat **(**ISSR) involves amplification of DNA segment present at an amplifiable distance in between two identical microsatellite repeat regions oriented in opposite direction. Inter-Simple Sequence Repeat usually 16-25 bp long as primers in a single primer PCR reaction targeting multiple genomic loci to amplify different sizes of inter-SSR sequences. The microsatellite repeats used as primer can be either di-nucleotides or tri-nucleotides. ISSR markers are highly polymorphic and are used on genetic diversity, gene tagging,

phylogeny, evolutionary biology and genome mapping studies (Reddy et al. 2002). ISSR PCR is a technique, which overcomes the problems like high cost of AFLP, low reproducibility of RAPD, and the flanking sequences to develop species specific primers for SSR polymorphism. ISSR segregate mostly as dominant markers following simple Mendelian inheritance. However, they have also been shown to segregate as codominant markers in some cases, thus enabling distinction between homozygote and heterozygote (Sankar and Moore 2001). ISSR have been used since 1994 for a wide range of organisms (Ziętkiewicz et al. 1994) in DNA fingerprinting, diversity analysis and genome mapping (Godwin et al. 1997; Bornet and Branchard 2001; de Vincente et al. 2004).

Advantages of ISSR

ISSR is quick, simple, highly reproducible and the use of radioactivity is not essential. ISSR markers usually show high polymorphism (Kojima et al. 1998), and with the most important advantage that no prior information about genomic sequence is required (Bornet and Branchard 2001).

Amplified Fragment Length Polymorphsim (AFLP)

AFLP technique combines the power of RFLP with the flexibility of PCR-based technology by ligating primer recognition sequences (adaptors) to the restricted DNA (Lynch and Walsh 1998). The key feature of AFLP is its capacity for "genome representation": the simultaneous screening of representative DNA regions distributed randomly throughout the genome. AFLP markers can be generated for DNA of any organism without initial investment in primer/probe development and sequence analysis. Both good quality and partially degraded DNA can be used for digestion but the DNA should be free of restriction enzyme and PCR inhibitors. Details of the AFLP methodology have been reviewed by many authors (Blears et al. 1998; Mueller and Wolfenbarger 1999; Ridout and

Donini 1999).

AFLP analysis involves restriction digestion of genomic DNA with a combination of rare cutter (EcoRI or PstI) and frequent cutter (MseI or TaqI) restriction enzymes. Double-stranded oligonucleotide adaptors are then designed in such a way that the initial restriction site is not restored after ligation. Such adaptors are ligated to both ends of the fragments to provide known sequences for PCR amplification. PCR amplification will only occur where the primers are able to anneal to fragments which have the adaptor sequence plus the complementary base pairs to the additional nucleotides called selective nucleotides. An aliquot is then subjected to two subsequent PCR amplifications under highly stringent conditions with primers complementary to the adaptors and possessing 3` selective nucleotides of 1 – 3 bases. The first PCR (preamplification) is performed with primer combinations containing a single bp extension, while final (selective) amplification is performed using primer pairs with up to 3-bp extension. Because of the high selectivity, primers differing by only a single base in the AFLP extension amplify a different subset of fragments. A primer extension of one, two or three bases reduces the number of amplified fragments by factors of 4, 16 and 64, respectively. AFLP fragments are visualized either AgNO3 staining, or on denaturing polyacrylamide gels with autoradiography, or on agarose gel, or automatic DNA sequencers.

Advantages of AFLP

(i) It is highly reproducible and reliable (Jones et al. 1997).

(ii) It does not require any DNA sequence information from the organism under study.

(iii) It is information-rich due to its ability to analyze a large number of polymorphic loci simultaneously with a single primer combination on a single gel as compared to RFLPs and microsatellites (Russell et al. 1997).

(iv) Co-migrating AFLP amplification products are mostly homologous and locus specific with exceptions in polyploidy species.
(v) Both good quality and partially degraded DNA can be used for digestion but the DNA should be free of restricttion enzyme and PCR inhibitors.

Co-dominant markers

Restriction Fragment Length Polymorphism (RFLP)

RFLPs are inherited naturally occurring Mendelian characters. They have their DNA rearrangements due to evolutionary processes, unequal crossing over, mutations within the fragments, point mutations within the restriction enzyme recognition site (Schlotterer and Tautz 1992). In RFLP analysis, restriction enzyme digested genomic DNA and then resolved by gel electrophoresis and western blotting (Southern 1975). Specific banding patterns are then visualized by hybridization with labeled probe. These probes are mostly species-specific of about 0.5–3.0 kb in size, obtained from a cDNA library or a genomic library. The genomic libraries are easy to construct but a large number of scattered duplicates are found in inserts that makes complex patterns. This problem can be overcome by using methylation sensitive restriction enzyme PstI which facilitates DNA sequences of small sizes, preferred in RFLP analysis (Figdore et al. 1988). In contrast cDNA libraries are difficult to construct, however, they are more popular as actual genes are analyzed and they contain fewer repeat sequences (Miller and Tanksley 1990; Landry and Michelmore 1987)

.

Advantages of RFLP markers

(i) RFLP markers were used for constructing genetic maps.
(ii) RFLPs are codominant and reliable markers in linkage analysis, breeding.

(iii) can be easily determined in homozygous or heterozygous state of an individual.

Disadvantages of RFLP markers

(i) the large amount of DNA required for restriction digestion and Southern blotting.
(ii) Expensive, time-consuming and hazardous.
(iii) only one marker may be polymorphic, which is highly inconvenient especially for crosses between closely-related species and their inability to detect point mutations and polymorphism (Botstein et al. 1980; Winter and Kahl 1995)

Microsatellites (SSR)

Microsatellites, also known as simple sequence repeats (SSR), variable number tandem repeats (VNTR) and short tandem repeats (STR) are tandem repeats motifs of 1-6 nucleotides found at high frequency in the nuclear genomes of most taxa (Beckmann and Weber, 1992). For example, (A)11, (GT)12, (ATT)9, (ATCG)8, (TAATC)6 and (TGTGCA)5 represent mono-, di-, tri-, tetra-, penta- and hexa-nucleotide repeats, respectively. A microsatellite locus typically varies in length between 5 and 40 repeats, but longer strings of repeats are possible. Dinucleotide, trinucleotide and tetranucleotide repeats are the most common choices for molecular genetic studies. Dinucleotides are the dominant type of microsatellite repeats in most vertebrates characterized so far, although trinucleotide repeats are most abundant in plants (Beckmann and Weber, 1992; Kantety et al. 2002; Chen et al. 2006). Despite the fact that the mechanism of microsatellite evolution and function remains unclear, SSRs were being widely employed in many fields soon after their first description (Litt and Luty 1989; Tautz 1989; Weber and May 1989) because of the high variability which makes them very powerful genetic markers. Microsatellites have proven to be an extremely valuable tool for

genome mapping in many organisms (Knapiket al. 1998), but their applications span over different areas ranging from kinship analysis, to population genetics and conservation/management of biological resources (Jarne and Lagoda 1996). Microsatellites can be amplified for identification by the polymerase chain reaction (PCR), using two unique sequences which are complementary to the flanking regions as primers. This process results in production of enough DNA to be visible on agarose or polyacrylamide gels; only small amounts of DNA are needed for amplification as thermocycling in this manner creates an exponential increase in the replicated segment.

With the abundance of PCR technology, primers that flank microsatellite loci are simple and quick to use, but the development of correctly functioning primers is often a tedious and costly process. However, once they are developed and characterized in an organism, microsatellites are powerful for a variety of applications because of their reproducibility, multiallelic nature, codominant inheritance, relative abundance and good genome coverage (Liu and Cordes 2004).

Unlike conserved flanking regions, microsatellite repeat sequences mutate frequently by slippage and proofreading errors during DNA replication that primarily change the number of repeats and thus the length of the repeat string (Eisen 1999). Because alleles differ in length, they can be distinguished by high-resolution gel electrophoresis, which allows rapid genotyping of many individuals at many loci for a fraction of the price of sequencing DNA. Many microsatellites have high-mutation rates (between 10-2 and 10-6 mutations per locus per generation, and on average 5×10-4) that generate the high levels of allelic diversity necessary for genetic studies of processes acting on ecological time scales.

Advantages of SSRs markers

(i) SSR are used for plant breeding, conservation biology

and population genetics as forensics, paternity analysis and gene mapping (Coates and Byrne 2005).

(ii) require little amount of DNA, which does not have to be of high quality.

(iii) the simple interpretation of results (de Vicente and Fulton 2003).

Disadvantage of SSRs markers

(i) the requirement of a known sequence to be amplified (Weising et al. 2005).

(ii) Developing new microsatellites are expensive and time consuming (Coates and Byrne 2005).

(iii) the phenomena null-alleles, which are nonamplifying alleles and appears frequently. Null-alleles leave no bands when having homozygousity, but when having heterozygousity it leaves one band visual. This will interfere and complicate reading of data, since it will be registered as a homozygote individual when actually being a heterozygote. To reduce error due to null alleles, population studies should contain many diverse SSR primers so that different multiple microsatellite loci are investigated (Weising et al. 2005).

Sequence characterized amplified regions for amplification of specific band (SCAR)

Michelmore (1991) and Martin (1991) was the first to introduce this technique, in which RAPD marker termini are sequenced and longer primers of 22–24 nucleotide bases long are designed for specific amplification of a particular locus. It shows similarity with STS markers in construction and application. The presence or absence of the band represent variation in sequence.

Advantages of SCARs markers

a) SCARs are advantageous over RAPD markers as

(i) they detect only a single locus.
(ii) their amplification is less sensitive to reaction conditions.
(iii) they can potentially be converted into codominant markers (Paran and Michelmore 1993).

b) SCARs compared to arbitrary primers
   (i) SCARs exhibit several advantages in mapping studies (codominant SCARs are informative for genetic mapping than dominant RAPDs)
   (ii) map based cloning as they can be used to screen pooled genomic libraries by PCR.
   (iii) locus specificity and physical mapping.

SCARs allow comparative mapping or homology studies among related species, thus making it an extremely adaptable concept in the near future (Tanaka et al. 2006; Michhelmore et al. 1991).

Cleaved Amplified Polymorphic Sequence (CAPS)

CAPS are a combination of the RFLP and PCR and it was originally named PCRRFLP (Maeda et al., 1990). The technique involves amplification of a target DNA through PCR, followed by digesting with restriction enzymes (Michaels and Amasino 1998). Hence, CAPS markers rely on differences in restriction enzyme digestion patterns of PCR fragments caused by nucleotide polymorphism between samples. Critical steps in the CAPS marker approach include DNA extraction, and the number or distribution of polymorphic sites, and PCR conditions.

Advantage of CAPS markers

(i) since analysis of restriction fragment length polymorphisms is based on PCR amplification, it is much easier and less time-consuming than analyzing alternative types of markers that require southern hybridizations.

(ii) CAPS primers developed from ESTs are more useful as genetic markers for comparative mapping study than those markers derived from non-functional sequences such as genomic microsatellite markers.

CAPS markers are inherited mainly in a co-dominant manner (Matsumoto and Tsumura 2004).

Expressed Sequence Tags (EST)

Each gene must be converted or transcribed into messenger RNA (mRNA) that serves as a template for protein synthesis. The resulting mRNA then guides the synthesis of a protein through a process called translation. The problem is that mRNA is very unstable outside of a cell; therefore, scientists use an enzyme called reverse transcriptase to convert mRNA to complementary DNA (cDNA). cDNA production is the reverse of the usual process of transcription in cells because the procedure uses mRNA as a template rather than DNA. cDNA is a much more stable compound and it represents only expressed DNA sequence because it is generated from mRNA that represents exons by excising (splicing) introns. Once cDNA representing an expressed gene has been isolated, scientists can then sequence a few hundred nucleotides from either the 5' or 3' end to create 5' expressed sequence tags (5' ESTs) and 3' ESTs, respectively. A 5' EST is obtained from the portion of a transcript (exons) that usually codes for a protein. These regions tend to be conserved across species and do not change much within a gene family. The 3' ESTs are likely to fall within non-coding (introns) or untranslated regions (UTRs), and therefore tend to exhibit less cross-species conservation than do coding sequences. The challenge associated with identifying genes from genomic sequences varies among organisms and is dependent upon genome size as well as the presence or absence of introns, which are the intervening DNA sequences interrupting the protein coding sequence of a gene. The production of ESTs starts with the construction of cDNA

libraries. The identification of ESTs has proceeded rapidly, with over 6 million ESTs now available in computerized databases. ESTs were originally intended as a way to identify gene transcripts, but have since been instrumental in gene discovery, for obtaining data on gene expression and regulation, sequence determination, and for developing highly valuable molecular markers, such as EST-based RFLPs, SSRs, SNPs, and CAPS. ESTs have been used for designing probes for DNA microarrays that is used to determine gene expression. ESTs also allow the efficient development of single or low-copy RFLP markers. RFLP markers developed from ESTs (EST-RFLP) have been extensively used for the construction of high-density genetic linkage maps (e. and physical maps (Kurata et al. 1997). Often EST-based RFLP 2556 markers allow comparative mapping across different species, because sequence conservation is high in the coding regions. Hence, marker development and mapbased cloning in one species will profit directly from data, which are available in any other species. ESTs also allow a computational approach to the development of SSR and SNP markers (Eujayl et al. 2001) for which previous development strategies have been expensive. Pattern-finding programs can be employed to identify SSRs in ESTs. The available sequence information allows the design of primer pairs, which can be used to screen cultivars of interest for length polymorphisms. A modest 1% to 5% of the ESTs in various plant species have been found to have SSRs of suitable length (20 bp or more) for marker development (Kantety et al. 2002). It should be possible to find a large number of these SSRs in an organism for which a great number of ESTs have been generated. EST-SSRs also have a higher probability of being functionally associated with differences in gene expression than the genomic SSRs (Gao et al. 2004). Most of the large scale, multispecies in silico mining efforts for developing EST-SSRs seem to have focused primarily on monocotyledonous crops (Kantety et al. 2002; Thiel et al. 2003), although ESTs of a few dicot species have been explored

for SSR mining (Qureshi et al. 2004; Varshney et al. 2005).

Two strategies have been employed for SNP development based on ESTs.

One strategy uses ESTs from the 3'-end of cDNA clones, which consists mainly of 3'-UTRs, to maximize the chance of finding sequence variations. Primer pairs can be derived from the EST sequences and the amplification of corresponding regions from several genotypes followed by sequence comparison may reveal SNPs.

Alternatively, one can use clusters of ESTs which contain sequences from different cultivars and identify potential SNPs computationally.

Single nucleotide polymorphism (SNP)

Single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide (A, T, G or C) differs among members of a species. SNP is the most abundant marker system both in animal and plant genomes and has recently emerged as the new generation molecular markers for various applications. Being binary or co-dominant status, they are able to efficiently discriminate between homozygous and heterozygous alleles. Moreover, unlike microsatellites their power comes not from the number of alleles but from the large number of loci that can be assessed (Foster et al. 2010). Once the rare SNPs are discovered in a low diversity species, the genetic population discrimination power can be equivalent to the same number of loci in a genetically diverse species. The more evolutionary conserved nature of SNPs makes them less subject to the problem of homoplasy (Brumfield et al. 2003). Most importantly, SNPs are amenable to high throughput automation, allowing rapid and efficient genotyping of large numbers of samples (Tsuchihashi and Dracopoli 2002). In plants, SNP can be designed from ESTs (Coles et al. 2005), and single-stranded pyrosequencing (Miller et al. 2003). A high-throughput genome analysis method called diversity array

technology (DArT), based on microarray platform, has been developed for the analysis of plant DNA polymorphism (Jaccoud et al. 2001). Eijk et al. (2004), described a novel SNP genotyping technique, SNPWave. Chip-based SNP arrays use thousands of oligonucleotide probes attached to a solid surface (e.g., glass, silicon wafer) allowing for a large number of SNPs to be interrogated simultaneously (Rapley and Harbron 2001). The ABI PRISM SNaPshot Multiplex Kit is designed to interrogate up to 10 single nucleotide polymorphisms (SNPs) at known locations on one to 10 DNA templates in a single tube. In brief, the protocol includes preparation of sample reactions using template and primer, performing SNaPshot reactions by thermal cycling and conduction of post-extension treatment of the products. Then automated electrophoresis of the samples and finally, analyzing the data. SNP is able to determine genetic diversity in plants, particularly in species with limited genetic diversity. Determination of population genetic structure of Castor bean (Ricinus communis) using SNPs from genome-wide comparisons showed low levels of genetic diversity and mixing of genotypes, leading to minimal geographic structuring of castor bean populations worldwide (Foster et al. 2010).

Clustering of Castor bean indicated five main groups worldwide and a repeated pattern of mixed genotypes in most countries; most molecular variance occurred within-populations (74%) followed by 22% among-populations and 4% among-continents (Foster et al. 2010).. Recently, a single nucleotide primer extension (SNuPE) assay targeting gyrB gene has been developed to identify bacteria belonging to the Burkholderia cepacia complex, which are very difficult to identify using commonly used phenotypic and molecular techniques (Coenye et al. 2001). Novel SNP based technique allowed the successful detection and distinction of specific genetic variations and is effectively applied in routine medical diagnosis since it permits to analyze routinely many samples in a short time (Ferri et al. 2010). Similar approaches need to be utilized in plants that are

difficult to discriminate for low level of genetic diversity.

Sequence Tagged Site (STS)

STS was first developed by Olsen et al. (1989) as DNA landmarks in the physical mapping of the human genome, and later adopted in plants. STS is a short, unique sequence whose exact sequence is found nowhere else in the genome. Two or more clones containing the same STS must overlap and the overlap must include STS. Any clone that can be sequenced may be used as STS provided it contains a unique sequence. In plants, STS is characterized by a pair of PCR primers that are designed by sequencing either an RFLP probe representing a mapped low copy number sequence (Blake et al. 1996) or AFLP fragments. STS markers are codominant, highly reproducible, suitable for high throughput and automation, and technically simple for use (Reamon-Buttner and Jung 2000).

Competing interest: The authors declare that they have no competing interests.

Acknowledgments: The authors are thankful to the Department of Biotechnology and Genetic Engineering, KUST, Kohat, Pakistan.

BIBLIGRAPHY

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubrick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. 1991. "Complementary DNA sequencing: expressed sequence tags and human genome project." *Science* 252: 1651–1656.

Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B. 1992. "Length polymorphisms of simple sequence repeat DNA in soybean." *Genetics* 132: 1131-1139.

Bassam, B.J., Caetano-Anolles, G. and Gresshoff, P.M. 1992.

"DNA amplification fingerprinting of bacteria." *Appl. Microbiol. Biotechnol* 38:70-76.

Beckmann, J.S. and Weber, J.L. 1992. "Survey of human and rat microsatellites." *Genomics* 12: 627–631.

Blake, T.K., Kadyrzhanova, D., Shepherd, K.W., Islam, A.K.M.R., Langridge, P.L., McDonald, C.L., Erpelding, J., Larson, S., Blake, N.K. and Talbert, L.E. 1996. "STS-PCR markers appropriate for wheat-barley introgression." *Theor. Appl. Genet* 82:715-721.

Blears, M.J., De Grandis, S.A., Lee, H. and Trevors, J.T. 1998. "Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications." *J. Industrial Microbiol. Biotechnol* 21: 99–114.

Bornet, B. and Branchard, M. 2001. "Nonanchored inter simple sequence repeat (ISSR) markers: Reproducible and specific tools for genome fingerprinting." *Plant Molecular Biology Reporter* 19(3): 209-215.

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. 1980. "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." *Am. J. Hum. Genet* 32: 314-331.

Brumfield, R.T., Beerli, P., Nickerson, D.A., and Edwards, S.V. 2003. "The utility of single nucleotide polymorphisms in inferences of population history." *Trends Ecol. Evol* 18: 249–256.

Caetano-Anolles, G., Bassam, B.J. and Gresshoff, P.M. 1991. "DNA amplification fingerprinting using very short arbitrary oligonucleotide primers." *Biotechnol* 9: 553-557.

Caetano-Anolles, G., Bassam, B.J. and Gresshoff, P.M. 1992. "Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides." *Mol. General Genet* 235: 157-165.

Caetano-Anolles, G., Bassam, B.J. and Gresshoff, P.M. 1993.

"Enhanced detection of polymorphic DNA by multiple arbitrary amplicons profiling of endonuclease digested DNA: identification of markers linked to the supernodulation locus in soybean." *Mol. General Genet* 241: 57-64.

Caetano-Anolles, G. 1994. "MAAP- A versatile and universal tool for genome analysis." *Plant Mol. Biol* 25: 1011-1026.

Chen, C.X., Zhou, P., Choi, Y.A., Huang, S. and Gmitter, F.G. 2006. "Mining and characterizing microsatellites from citrus ESTs." *Theor. Appl. Genet* 112: 1248-1257.

Coenye, T., Vandamme, P., Govan, J.R.W. and LiPuma, J.L. 2001. "Taxonomy and Identification of the Burkholderia cepacia Complex." *J. Clin. Microbiol* 39: 3427–3436.

Coetes, D.J. and Byrne, M. 2005. "Genetic variation in plant populations." In *Plant diversity and Evolutions-Genotypic and phenotypic variation in higher plants*, edited by R.J. Henry, ch. 9: 139-164. Wallingford, Oxfordshire, UK: CABI Publishing.

Coles, N.D., Coleman, C.E., Christensen, S.A., Jellen, E.N., Stevens, M.R., Bonifacio, R.A., Rojas-Beltran, J.A., Fairbanks, D.J. and Maughan, P.J. 2005. "Development and use of an expressed sequenced tag library in quinoa (Chenopodium quinoa Willd.) for the discovery of single nucleotide polymorphisms." *Plant Soc* 168: 439–447.

Collard, B.C.Y., Jahufer, M.Z.Z. and Pang, E.C.K. 2005. "An introduction to markers, quantitative trait loci (QTL) mapping and marker-assissted selection for crop improvement: The basic concepts." *Euphytica* 142(1-2): 169-196.

Cooke, R.J. 1984. "The characterization and identification of crop cultivars by electrophoresis." *Elect* 5: 59-72.

Corley Smith, G.E., Lim, C.J., Kalmar, G.B. and Brandhorst, B.P. 1997. "Efficient detection of DNA polymorphisms by fluorescent RAPD analysis." *BioTechniques* 22: 690-692.

Dalmastri, C., Chiarini, L., Bevivino, A., Manno, G., Mentasti,

M. and Fani, R. 2010. "Application of multiplex single nucleotide primer extension (mSNuPE) to the identification of bacteria: The Burkholderia cepacia complex case." *J. Microbiol. Methods* 80: 251–256.

de Vicente, M.C. and Fulton T. 2003. "Using molecular marker technology in studies on plant genetic diversity." [Online] IPGRI, Rome, Italy and Institute for Genetic Diversity, Ithaca, New York, USA.

Eijk, M.J.T., Broekhof, J.L.N., van-der-Poel, H.J.A., Hogers, R.C.J., Schneiders, H., Verstege, K.J.E., van-Aart, J.W., Geerlings, H., Buntjer, J.B., van-Oeveren, A.J. and Vos, P. 2004. "SNPWave: A flexible multiplexed SNP genotyping technology." *Nucl. Acids Res* 32: e47.

Eisen, J.A. 1999. "Mechanistic basis for microsatellite instability." In *Microsatellites: Evolution and applications*, edited by Goldstein, D.B. & Schlötterer, C., 34-48. Oxford, UK: Oxford University Press.

Eujayl, I., Sorrells, M., Baum, M., Wolters, P. and Powell, W. 2001. "Assessment of genotypic variation among cultivated durum wheat based on ESTSSRs and genomic SSRs." *Euphytica* 119: 39–43.

Ferri, L., Perrin, E., Campana, S., Tabacchioni, S., Taccetti, G., Cocchi, P., Ravenni, N., Dalmastri, C., Chiarini, L., Bevivino, A., Manno, G., Mentasti, M. and Fani, R. 2010. "Application of multiplex single nucleotide primer extension (mSNuPE) to the identification of bacteria: The Burkholderia cepacia complex case." *J. Microbiol. Methods* 80: 251–256.

Figdore, S.S., Kennard, W.C., Song, K.M., Slocum, M.K. and Osborn, T.C. 1988. "Assessment of the degree of restriction fragment length polymorphism in Brassica." *Theor. Appl. Genet* 75: 833–40.

Foster, J.T., Allan, G.J., Chan, A.P., Rabinowicz, P.D., Ravel, J., Jackson, P.J. and Keim, P. 2010. "Single nucleotide polymorphisms for assessing genetic diversity in castor

bean (Ricinus communis)." *BMC Plant Biol* 10: 13–23.

Gao, L.F., Jing, L.R., Huo, N.X., Li, Y., Li, X.P., Zhou, R.H., Chang, X.P., Tang, J.F., Ma, Z.Y. and Jia, J.Z. 2004. "One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat." *Theor. Appl. Genet* 108: 1392–1400.

Godwin, I.D., Aitken, E.A.B. and Smith, L.W. 1997. "Application of inter simple sequence repeat (ISSR) markers to plant genetics." *Electrophoresis* 18(9): 1524-1528.

Gupta, P.K. and Varshney, R.K. 2004. "Cereal genomics: an overview." In *Cereal genomics*, edited by Gupta P.K. and Varshney, R.K., 639. Dordrecht, The Netherlands: Kluwer Academic Press.

Hartl, D.L. and Clark, A.G. 1997. *Principles of population genetics*. 3rd ed. Sunderland, Massachsetts: Sinauer Assoc., 163.

Hartl, D.L. 1988. "Genetic variation." In *A primer of population genetics*. 2nd ed., Ch. 1:1-67. Massachusetts: Sinauer Associates.

Hartl, D.L. and Jones, E.W. 2005. "DNA Structure and DNA manipulation." In *Genetics: analysis of genes and genomes*. 5th ed., Ch. 2: 36-85. Sudbury: Jones and Bartlett Pub.

Hollingsworth, W.O., Christie, C.B., Nichols, M.A. and Neilson, H.F. 1998. "Detect ion of variation among and within asparagus hybrids using random amplified DNA (RAPD) markers." *New Zealand J. Crop Hortic. Sci* 26: 1-9.

Huff, D.R., Peakall, R. and Smouse, P.E. 1993. "RAPD variation within and among natural populations of outcrossing buffalograss [(Buchloe dactyloides (Nutt.) Engelm]." *Theor. Appl. Genet* 86: 927-934.

Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. 2001. "Diversity arrays: A solid state technology for sequence information independent genotyping." *Nucl. Acids Res*

29: e25.

Jain, A., Bhatia, S., Banga, S. and Prakash, S. 1994. "Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (Brassica juncea) and its relationship to heterosis." *Theor. Appl. Genet* 88: 116-122.

Jain, S.M., Brar, D.S. and Ahloowalia, B.S. 2002. Molecular techniques in crop improvement. The Netherlands: Kluwer Academic Publishers.

Jarne, P. and Lagoda, P.J.L. 1996. "Microsatellites, from molecules to populations and back." *Trends in Ecol. Evol* 11: 424-429.

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F. and van de Wiel, C. 1997. "Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories." *Mol. Breed* 3: 381-390.

Jordan, S.A. and Humphries, P. 1994. "Single nucleotide polymorphism in exon 2 of the BCP gene on 7q31-q35." *Hum. Mol. Genet* 3: 1915.

Kantety, R.V., Rota, M. L., Matthews, D.E. and Sorrells, M.E. 2002. "Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat." *Plant Mol. Biol* 48: 501-510.

Knapik, E.W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M.C. and Jacob, H.J. 1998. "A microsatellite genetic linkage map for zebrafish (Danio rerio)." *Nature Genet* 18: 338-343.

Kojima, T., Nagaoka, T., Noda, N. and Ogihara, Y. 1998. "Genetic linkage map of ISSR and RAPD markers in Einkorn wheat in relation to that of RFLP markers." *Theor. Appl. Genet* 96: 37–45.

Konieczny, A. and Ausubel, F.M. 1993. "A procedure for mapping Arabidopsis mutations using co-dominant

ecotype-specific PCR-based markers." *Plant J* 4: 403–410.

Kurata, N., Umehara, Y., Tanoue, H. and Sasaki, T. 1997. "Physical mapping of the rice genome with YAC clones." *Plant Mol. Biol* 35: 101–113.

Landry, B.S. and Michelmore, R.W. 1987. *An Agricultural Perspective.* Edited by G. Bruening, J. Harada and A. Hollaender. New York: Plenum, 134–57.

Litt, M. and Luty, J.A. 1989. "A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene." *American J. Human Genet* 44: 397-401.

Liu, Z.J. and Cordes, J.F. 2004. "DNA marker technologies and their applications in aquaculture genetics." *Aquaculture* 238: 1-37.

Lynch, M. and Walsh, B. 1998. *Genetics and analysis of quantitative traits.* Sunderland, MA: Sinauer Associates.

Matsumoto, A. and Tsumura, Y. 2004. "Evaluation of cleaved amplified polymorphic sequence markers." *Theor. Appl. Genet* 110: 80–91.

Micheli, M.R., Bova, R., Calissano, P. and Dambrosio, E. 1993. "Randomly amplified polymorphic DNA fingerprinting using combinations of oligonucleotide primers." *Biotechniques* 15: 388-390.

Michelmore, R.W., Paran, I.and Kesseli, R. V. 1991. "Identification of markers linked to disease resistance gene by bulk segregant analysis: A rapid methods to detect markers in specific genome region using segregating populations." *Proc. Natl. Acad. Sci* 88: 9828–32.

Michaels, S.D. and Amasino, R.M. 1998. "A robust method for detecting single-nucleotide changes as polymorphic markers by PCR." *Plant J.* 14: 381–385.

Miller, J.C. and Tanksley, S.D. 1990. "RFLP analysis of phylogenetic relationship and variation in genus

lycopersicon." *Theor. Appl. Genet* 80: 437–48 9.

Miller, P.T. and Henry, R. 2003. "Single-nucleotide polymorphism detection in plants using a single stranded pyrosequencing protocol with a universal biotinylated primer." *Anal. Biochem* 317: 166–170.

Mueller, U.G. and Wolfenbarger, L.L. 1999. "AFLP genotyping and fingerprinting." *Trends Ecol. Evol* 14: 389-394.

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

Pammi, S., Schertz, K., Xu, G., Hart, G. and Mullet, J.E. 1994. "Randomamplified- polymorphic DNA markers in sorghum." *Theor. Appl. Genet* 89: 80-88.

Paran, I. and Michelmore, R.W. 1993. "Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce." *Theor. Appl. Genet* 85: 985-993.

Phillips, R.L. and Vasil, I.K. 2001. "DNA-based markers in plants." In *DNA-based markers in plants.* Edited by Phillips RL and Vasil IK, 497. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Qureshi, S.N., Saha, S., Kantety, R.V. and Jenkins, J.N. 2004. "EST-SSR: a new class of genetic markers in cotton." *J. Cotton Sci* 8: 112–123.

Rafinski, J. and Babik, W. 2000. "Genetic differentiation among northern and southern populations of the moor frog Rana arvalis Nilsson in central Europe." *Hered* 84(5): 610-618.

Rapley, R. and Harbron, S. 2001. *Molecular Analysis and Genome Discovery*. Chichester, UK: John Wiley & Sons Ltd.

Reamon-Buttner, S.M. and Jung, C. 2000. "AFLP derived STS markers for the identification of sex in Asparagus officinalis L." *Theor. Appl. Genet* 100: 432–438.

Reddy, M.P., Sarla, N. and Siddiq, E.A. 2002. "Inter simple sequence repeat (ISSR) polymorphism and its

application in plant breeding." *Euphytica* 128: 9–17.

Ridout, C. and Donini, P. 1999. "Use of AFLP in cereals research." *Trends Plant Sci* 4: 76-79.

Rojas-Beltran, J.A., Fairbanks, D.J. and Maughan, P.J. 2005. "Development and use of an expressed sequenced tag library in quinoa (Chenopodium quinoa Willd.) for the discovery of single nucleotide polymorphisms." *Plant Soc* 168: 439–447.

Russell, J.R., Fuller, J.D., Macaulay, M., Hatz, B.G., Jahoor, A., Powell, W. and Waugh, R. 1997. "Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs." *Theor. Appl. Genet* 95: 714-722.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. 1985. "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." *Science* 230: 1350–54.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. 1988. "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase." *Nature* 239: 487–97.

Sankar, A.A. and Moore, G.A. 2001. "Evaluation of inter-simple sequence repeat analysis for mapping in Citrus and extension of genetic linkage map." *Theor. Appl. Genet* 102: 206–214.

Schlotterer, C. and Tautz, D. 1992. "Slippage synthesis of simple sequence DNA." *Nucleic Acids Res* 20: 211–15.

Schulmann, A.H. 2007. "Molecular markers to assess genetic diversity." *Euphytica* 158(3): 313-321.

Southern, E.M. 1975. "Detection of specific sequences among DNA fragments separated by gel electrophoresis." *J. Mol. Biol* 98: 503–17.

Tautz, D. 1989. "Hypervariability of simple sequences as a general source for polymorphic DNA markers." *Nucleic*

Acids Res 17: 6463-6471.

Tanaka, H., Fukuda, N. and Shoyama, Y. 2006. "Identification and differentiation of Panax species are using ELISA, RAPD and Eastern blotting." *Phytochem. Anal* 17: 46-55.

Thiel, T., Michalek, R.K., Varshney, R.K. and Graner, A. 2003. "Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (Hordeum vulgare L.)." *Theor. Appl. Genet* 106: 411–422.

Tsuchihashi, Z. and Dracopoli, N.C. 2002. "Progress in high-throughput SNP genotyping methods." *Pharmacogenomics* J 2: 103–110.

Varshney, R.K., Graner, A. and Sorrells, M.E. 2005. "Genic microsatellite markers in plants: features and applications." *Trends Biotechnol* 23: 48–55.

Vejl, P. 1997. "Identification of genotypes in hop (Humulus lupulus L.) by RAPD analsysis using program Gel Manager for Windows." *Rostlinna Vyroba* 43: 325-331.

Verstege, K.J.E., van-Aart, J.W., Geerlings, H., Buntjer, J.B., van-Oeveren, A.J. and Vos, P. 2004. "SNPWave: A flexible multiplexed SNP genotyping technology." *Nucl. Acids Res* 32: e47.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van De Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. 1995. "AFLP: a new technique for DNA fingerprinting." *Nucleic Acids Res* 23: 4407- 4414.

Weber, J.L. and May P.E. 1989. "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction." *American J. Human Genet* 44: 388-396.

Weising, K., Nybom, H., Wolff, K. and Kahl, G. 2005. "Detecting DNA variation by molecular markers." In *DNA fingerprinting in plants: principles, methods, and applications*, Ch. 2: 21-74. 2nd ed. Boca Raton, FL: Taylor and Francis.

Welsh, J. and McClelland, M. 1990. "Fingerprinting genomes using PCR with arbitrary primers." *Nucleic Acids Res* 18: 7213-7218.

Welsh, J. and McClelland, M. 1991. "Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers." *Nucleic Acids Res* 19: 5275-5279.

Weller, J.W. and Reddy, A. 1997. "Fluorescent detection and analysis of RAPD amplicons using the ABI PRISM DNA sequencers." In *Fingerprinting methods based on arbitrarily primed PCR,* edited by Micheli MR and Bova R., 81-92. Springer Lab Manual.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. 1990. "DNA polymorphisms amplified by arbitrary primers are useful as genetic markers." *Nucl. Acids Res* 18: 6531–6535.

Winter, P. and Kahl, G. 1995. "Molecular marker technologies for plant improvement." *World J. Microbiol. Biotechnol* 11: 438–48.

Zietkiewicz, E., Rafalski, A. and Labuda, D. 1994. "Genome fingerprinting by simple sequence re-peats (SSR)-anchored polymerase chain reaction amplification." *Genomics* 20(2): 176-83.