MOLECULAR MARKERS: MOMENTOUS TOOLS FOR EXPLORING PLANT BIOTECHNOLOGY

Ritesh Kaur, Amandeep Hora and C.P. Malik*
School of Life Sciences, Jaipur National University, Jaipur 302017, Rajasthan, India
*Author for Correspondence

ABSTRACT
All through the preceding years, the use of molecular markers, for genotyping and revealing polymorphism at the DNA level, is one of the most significant developments in the field of plant biotechnology and their molecular genetic studies. There are different types of markers such as morphological, biochemical and DNA based molecular markers. These DNA based markers are differentiated in two types first non PCR based (RFLP) and second is PCR based markers (RAPD, AFLP, SSR, SNP etc.) amongst others, the microsatellite DNA marker has been the most widely used, due to its easy use by simple PCR, followed by a denaturing gel electrophoresis for allele size determination and to the high degree of information provided by its large number of alleles per locus. Despite this, SNPs, CCMPs and DArT are now on the scene and has gained high popularity. Now days, a variety of different genetic markers has been proposed to assess genetic variability as a complementary strategy to more traditional approaches in genetic resources management, in understanding the genomic variability and the diversity between the same as well as different species of the plants. In this review, we will discuss about various types of molecular markers.

Key Words: Genomic Variability, Molecular Markers, Plant Biotechnology, Polymorphism

INTRODUCTION
During the past few decades’s classical strategies of evaluating genetic variability such as comparative anatomy, morphology, embryology and physiology have been increasingly complemented by various molecular techniques which comprises of the analysis of chemical constituents (e.g. plant secondary compounds) and most importantly the characterization of macromolecules. On the basis of polymorphism found in proteins or DNA, the development of molecular markers has greatly facilitated research in a variety of disciplines such as taxonomy, ecology, phylogeny, genetics and plant breeding. Molecular markers have been verified to be a powerful tool to assess the genetic diversity of plants and to investigate genetic factors controlling quantitatively inherited traits. In establishing evolutionary relationship between related species, phylogenetic studies, studying biochemical pathways at the cellular level and in population and conservation genetics, these tools have played a very significant role and have given new insights in the world of plant genomes.

Some inherent properties of the molecular markers make them more suitable and advantageous over morphological markers (Paramita et al., 2012). The efficacy of the molecular marker is derived from these inherent properties which are as follows-

- Genotypes of molecular loci can be determined at a whole plant, tissue and cellular levels whereas phenotypes of most morphological markers can only be distinguished at a whole plant level.
- At molecular loci, relatively large number of naturally occurring alleles can be found. At morphological loci distinguishable alleles occur less frequently and often must be induced through the application of exogenous mutagens.
- Phenotypic neutrality: usually no deleterious effects are associated with alternate alleles of molecular markers. This is not the case with morphological markers which are often accompanied by undesirable phenotypic effects.
Review Article

- Co dominant nature: alleles of most molecular markers are co dominant allowing all possible genotypes to be distinguished in any segregating generation. Alleles at morphological marker loci usually interact in a dominant recessive manner, limiting the use in many crosses.

- Epistatic and pleiotropic effect: Strong epistatic effects limit number of segregating markers that can be unequivocally scored in the segregating generation with morphological loci. Fewer epistatic or pleiotropic effects are observed with molecular markers, thus a virtually limitless number of segregating markers can be monitored in a single population.

Moreover, molecular markers have advantages over other kinds, where they show genetic differences on a more detailed level without interferences from environmental factors, and where they involve techniques that provide fast results detailing genetic diversity (Binneck et al., 2002; Garcia et al., 2004; Saker et al., 2005; Goncalves et al., 2008; Souza et al., 2008). However, the discovery of high throughput platforms increases number of data per run and reducing the cost of the data and increasing map resolution.

The eukaryotic genome comprises of variety of repetitive as well as non repetitive sequences. Molecular methods which reveal these specific DNA sequence distribution patterns are generally described as DNA profiling/fingerprinting methods. DNA fingerprinting has extensive applicability and can resolve individual specific polymorphic patterns. With the help of this, it is promising to trace gene flow amongst the individuals in populations and can also establish genetic relatedness. From multi locus markers, DNA fingerprints are produced and an individual genotype can be identified at the molecular level on the basis of polymorphism in the sequence of its DNA. The various molecular markers can be classified into different groups based on:

1) Mode of transmission (biparental nuclear inheritance, maternal nuclear inheritance, maternal organelle inheritance or paternal organelle inheritance)
2) Mode of gene action (dominant or co dominant markers) and
3) Methods of analysis (hybridization-based or PCR-based markers).

Large number of molecular markers available these days. Some molecular markers have been discussed that are used to address the questions of diversity and systematic of plants.

**Biochemical or Protein Based Molecular Markers**

The primary molecular tools for measuring genetic diversity from the late 1960s until the 1980s are protein markers. The composition of proteins is - amino acids linked by covalent peptide bonds which results to form polypeptides. These protein (enzyme-and non-enzyme) based molecular markers provide indirect information about plant genome structure. Molecular markers based on protein polymorphism depend on the migration properties of proteins which allow separation by electrophoresis and revealed by histochemical stains specific to the enzyme to be assayed. In the presence of electric field the changes in the mobility of enzymes reflect changes in the encoding DNA sequences (Sharma et al., 2013).

Two general forms of protein data can be assembled simultaneously using electrophoretic methods. One is derived from isozymes, which are all functionally similar forms of enzymes including all polymers of subunits produced by different gene loci or by different alleles at the same locus. Second data set comprises of allozymes which are a subset of isozyme and are variants of polypeptides representing different allelic alternatives of the same gene locus. Both forms of molecular data are important in molecular systematics and involve proteins that can be separated on the basis of net charge and size.

Protein polymorphisms were the first markers used for genetic studies in livestock. However, the number of polymorphic loci that can be assayed and the level of polymorphisms observed at the loci are often low, which greatly limits their application in genetic diversity studies. With the development of new technologies, DNA polymorphisms have become the markers of choice for molecular-based surveys of genetic variation.

The inheritance in a codominant fashion in allozymes are the main advantages of protein marker, because of this, it is possible to distinguish between plants that are homozygous and heterozygous at a gene locus. Gene frequencies can be calculated by identifying the heterozygotes and homozygotes, and then these frequencies can be used for a comparison of populations for phylogenetic purposes. The analysis are
simple, cheap and relatively easy to perform, which makes it possible to study an excessive amount of individuals. Therefore, isozyme (allozyme) has been widely applied for genetic diversity and population structure analysis in a large range of plant species.

The main demerits of protein based molecular markers are - used only to detect variation in protein coding loci and therefore provide fewer markers compared to DNA-based methods. Furthermore, a protein marker often show low variability and in some cases no variability at all due to a low rate of mutational events.

**DNA Based Molecular Markers**

The traditional morphological and biochemical markers are very efficiently overcome by DNA based markers because of their advancements. Although, these markers can be used to assess genetic diversity and show little variation at the intra-specific level and they are strongly influenced by environmental factors (Sangwan et al., 2001).

Molecular markers based on DNA sequences detect more polymorphism than morphological and protein based marker and constitute a generation of genetic markers. DNA markers with higher resolving power have been used to discern exact genetic relationships, information on domestication, dispersion and evolution of crop sciences (Gepts, 1993). Therefore, DNA based molecular markers are very often used to obtain more consistent estimates of genetic diversity amongst them. Based on the loci detected and mode of inheritance the molecular markers can be single locus and multilocus. Polymorphism at a single locus is characterized, usually through use of a specific probe or specific PCR primers. Because the single locus detected by this method is characterized, one obtains a DNA genotype from single locus methods. Protein (allozymes and isozymes) and DNA based markers such as RFLP, SSR and SNPs are single locus codominant markers which are capable of distinguishing allelic variations. Polymorphism identified at multiple loci is multi locus fingerprinting. This can be performed by application of a mixture of single locus probes that identifies multiple similar sequence polymorphisms. RAPD, DAMD, ISSR and AFLP constitutes the dominant (allelic variations cannot be distinguished) multi locus markers.

DNA-based molecular techniques that are utilized to evaluate DNA polymorphism are as follows:

**Hybridization Based DNA Methods**

In the hybridization based markers, DNA profiles are visualized by hybridizing the restriction enzyme digested DNA to a labeled probe which is a DNA fragment of known /unknown sequences. Polymorphisms are detected by presence or absence of bands upon hybridization.

**Restriction Fragment Length Polymorphism (RFLP)**

RFLP is the most broadly used hybridization-based molecular marker. These markers were first used to identify DNA sequence polymorphisms for genetic mapping of a temperature-sensitive mutation of adeno-virus serotypes. It was then used for human genome mapping and after that adopted for plant genomes. RFLPs are inherited as naturally occurring Mendelian characters and have their DNA rearrangements due to evolutionary processes, point mutations within the restriction enzyme recognition site, mutations within the fragments, and unequal crossing over. The advantage of RFLPs is that they are co dominant markers and are very reliable in linkage analysis and breeding.

The limitations of RFLP marker are that large amount of DNA is required for restriction digestion and Southern blotting. The RFLP is relatively expensive and hazardous due to the requirement of radioactive isotope. The assay is time-consuming and labor-intensive and only one out of several markers may be polymorphic, which is highly inconvenient especially for crosses between closely-related species. Their inability to detect single base changes restricts their use in detecting point mutations occurring within the regions at which they are detecting polymorphism.

**PCR-based Methods**

Polymerase Chain Reaction (PCR) is the technique of *in vitro* amplification of specific DNA sequences and discovered by Mullis and Faloona (1987) which added a new dimension to the genetic analysis. PCR which is an extremely sensitive and a versatile technique that uses thermostable DNA polymerases which has changed the total scenario of molecular biology and has brought about a crowd of new possibilities in
molecular marker research. The technique involves in vitro amplification of particular DNA sequence or loci, with the help of specific or arbitrary oligonucleotide primers and the thermostable DNA polymerase enzyme thereby evading the molecular cloning. In this technique the double stranded template DNA is denatured at high temperature to form two single strands of DNA. The short oligonucleotide primers bind at low annealing temperature to these single strands of DNA at complementary motifs in reverse orientation. With the help of thermostable Taq DNA polymerase synthesis takes place by primer extension of the target sequence at high temperatures. These newly synthesized target DNA sequences are again denatured at high temperature and the cycle is repeated. The targeted DNA can be amplified exponentially if there is sufficient amount of polymerase, primers and nucleotides in the PCR reaction mix.

Of the various molecular approaches the PCR based technology offers a maximum potential not only for genetic analysis but also for phylogenetic and systematic studies. In case of PCR-based markers the primers of unknown/known sequences and length are used to amplify random or specific genomic DNA segments. These amplified products are then visualized by gel electrophoresis.

Cleaved Amplified Polymorphic Sequences (CAPS)
The principle behind this technique is that PCR fragments are digested with a restriction enzyme that is sized by gel electrophoresis. Several abbreviations have been created for these techniques which are PCR-RFLP and CAPS (Konieczny and Ausubel, 1993) are more frequently used. CAPS markers are generated in two steps: defined DNA sequence is first amplified by a sequence-specific primer pair and subsequently digested with a restriction enzyme and the resulting fragments are separated by gel electrophoresis. Mutations in the restriction sites can prevent the restriction and thus fragments of differing size are obtained. CAPS approach does not require radioactivity or blotting steps as RFLP but instead exhibits all the attributes of PCR-based techniques. CAPS markers are codominantly inherited thus the possibility of distinguishing homo and heterozygous states makes this procedure particularly attractive for mapping purposes (Drenkard et al., 1998).

Random Amplification of Polymorphic DNA (RAPD)
William and his coworkers in 1990 developed a PCR-based genetic assay named as Randomly Amplified Polymorphic DNA (RAPD), which employs single primer usually with 10 nucleotide bases as oligonucleotide primers and a GC content of at least 50% to amplify discrete fragments of DNA in low stringency of polymerase chain reaction. In this method essentially a single primer of arbitrary nucleotide sequence anneals to the genomic DNA, if these primer annealing sites on opposite strands are and within an amplifiable distance (generally less than 3000 bp) of each other then discrete products are formed through thermocyclic amplification. The PCR products are separated on agarose gel by electrophoretic procedure and bands are detected by ethidium bromide staining and visualized in UV transilluminator. The number of amplified products depends on the genome size and the homology between primer and the template DNA sequence. On an average, each primer directs amplification of several discrete loci in the genome, making the assay useful for efficient screening of nucleotide sequence polymorphism between individuals. The presence or absence of the PCR product is assumed to represent mutations in the primer-binding sites of the genomic DNA.

The variants of RAPD procedure are the DAF (DNA Amplification Fingerprinting) that uses very short primers usually 5-8 nucleotides long with either low or high stringency annealing steps and two temperatures instead of three temperatures cycling programme. The resulting fragments are separated on polyacrylamide gels and visualized by silver staining another variant of RAPD procedure is Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) that uses oligonucleotides of 20 or more nucleotides. Here the radiolabelled PCR products are separated by polyacrylamide gel electrophoresis and made visible by autoradiography (Welsh and McClelland, 1990).

Among the PCR based DNA marker RAPDs are cost effective, most versatile and relatively easy to perform. The technique requires no prior knowledge of DNA sequence and utilizes minor quantities of
DNA material therefore can be applied to even rare plant species. The main limitations encountered with the use of RAPD markers are repeatability of banding patterns and dominant inheritance. Despite the limitations associated with RAPD markers they have been successfully used in a wide variety of species because of its practical advantages. They have been employed widely for the determination of genetic diversity, phenetic relationships, and the identification of cultivars in number of plant species (Fernandez et al., 2002), estimation of breeding systems (Fritsch and Rosenberg, 1992), patterns of intrapopulation (Roelofs and Bachmann, 1995), conservation genetics (Fritsch and Rieseberg, 1996) etc.

Inter Simple Sequence Repeats (ISSR)
Inter-SSR (ISSR) genetic markers were developed from the common SSR motifs present in eukaryotic organisms (Gupta et al., 1994; Zietkiewcz et al., 1994) and are tandem repeats of 1-10 bp DNA sequence motif as the “simple sequence repeats”, interspersed evenly throughout the genome (Epplen et al., 1991), with regions either unanchored (Wu et al., 1994) or anchored at the 5’ or 3’ end by two or four arbitrary, often degenerate nucleotides (Fang et al., 1997). The addition of a different base at the 5’or 3’end renders their binding sites more specific and reproducible (Barth et al., 2002). The sequence between the two binding sites in opposite orientation within suitable distance is amplified, and indels within this region and loss or gain of binding sites are detected as band polymorphism. Since these are in abundant throughout the genome and it reveals a much larger number of fragments per primer than RAPD (Wolfe and Liston, 1998). ISSR markers are inherited as dominant genetic markers (Gupta et al., 1994) and are multilocus markers so, no prior DNA sequence information is needed, development costs are low, and laboratory procedures can easily be transferred to any plant species (Barth et al., 2002). ISSR markers are highly polymorphic and provide a novel quick, reliable and highly informative system for DNA fingerprinting approach and have been extensively used in genetic diversity investigations (Zietkiewicz et al., 1994) and for taxonomy and phylogeny to evaluate genetic relationships (Fernandez et al., 2002). These markers have been used as a mapping tool in a wide range of organisms (Zietkiewicz et al., 1994). Although ISSR markers are highly polymorphic and robust and tend to be evenly distributed throughout the plant genomes, they are dominant (Casasoli et al., 2001) in nature which limit their use in various studies.

Simple Sequence Repeats (SSR)
Microsatellites are also known as simple sequence repeats SSRs; short tandem repeats (STRs) or simple sequence length polymorphisms SSLPs are the smallest class of simple repetitive DNA sequences. They are known to be universal in prokaryote and eukaryote genomes and are present both in coding and non-coding regions. Microsatellites may be 2–8 bp repeats (Armour et al., 1999), 1–6 bp repeats (Goldstein and Pollock, 1997) or even 1–5 bp repeats (Schlotterer, 1998), that occur at multiple sites (upto 10^5) in eukaryotic genomes (Wang et al., 1994). The repeated sequence is often simple, consisting of two, three or four nucleotides (di-, tri-, and tetranucleotide repeats, respectively). One common example of a microsatellite is a dinucleotide repeat (CA)n, where n refers to the total number of repeats. A key feature of this class of repetitive DNA is an extraordinary high level of variation among taxa, mainly expressed as variation in the copy number of tandem repeats at a particular locus (Geiatlunger et al., 1997), which makes them a very powerful genetic marker in various studies. Microsatellites are abundant and occur frequently and randomly in all eukaryotic nuclear DNAs (Gupta et al., 1996). Microsatellite diversity is detected by amplifying DNA using PCR. Unique conserved sequences or primers flanking microsatellites are used to define the DNA segment that is to be amplified. The resulting DNA fragments are separated according to the size using electrophoresis. The main advantages of microsatellites include their codominant nature, high abundance, enormous extent of allelic diversity and the ease of assessing size variation by PCR with pairs of flanking primers. The main disadvantage is that sequence information is required for primer design.

Directed Amplification of Minisatellite DNAs (DAMD)
Minisatellites are tandemly repeated DNA regions of eukaryotic genomes, many of which showed high levels of allelic length variation due to differences in the number of repeated units (Jeffreys et al., 1985). They are dispersed throughout the genome and mutate very frequently. The sequences consist of...
repetitive, generally GC-rich, variant repeats that range in length from 10 to over 100 bp which are highly informative as genetic markers that have been used extensively in many areas of genetics. The high abundance of minisatellites in eukaryotic genomes allows the use of minisatellite complementary oligonucleotides as PCR primers to generate numerous polymorphic amplification products. Minisatellites are abundant in humans, animals and plant genomes (Zhou and Gustafson, 1995).

Heath et al., (1993) reported a technique, called directed amplification of minisatellite region DNA (DAMD) to direct the PCR mediated amplification of minisatellite DNA region. The technique involves the primers derived from these minisatellite regions of the genome containing hypervariable regions or Variable Number of Tandem Repeats (VNTRs). It has been speculated that when a portion of a minisatellite DNA array is involved in an inversion, a single primer makes PCR possible for the amplification of minisatellite core region (Heath et al., 1993). The DAMD-PCR technique offers several advantages to other methods (Bebeli et al., 1997). Since minisatellite core sequences which are used as primers are longer than RAPD-PCR primers, DAMD-PCR can be effectively carried out at relatively high stringency reactions.

Amplified Fragment Length Polymorphism (AFLP)

Vos et al., (1995) described AFLP technology. The key feature of AFLP–PCR is its capacity for the simultaneous screening of many different DNA regions distributed randomly throughout the genome.

Table 1: Comparison of commonly used genetic markers (modified from Semagn et al., 2006)

<table>
<thead>
<tr>
<th>Features</th>
<th>RFLP</th>
<th>Microsatellites</th>
<th>RAPD</th>
<th>AFLP</th>
<th>ISSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic abundance</td>
<td>High</td>
<td>Medium</td>
<td>very high</td>
<td>very high</td>
<td>Medium</td>
</tr>
<tr>
<td>Part of genome surveyed</td>
<td>low copy</td>
<td>whole genome</td>
<td>whole genome</td>
<td>whole genome</td>
<td>whole genome</td>
</tr>
<tr>
<td>Amount of DNA required</td>
<td>Low</td>
<td>Low single base</td>
<td>Low single base</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Type of polymorphism</td>
<td>High</td>
<td>insertion, deletion</td>
<td>insertion, deletion</td>
<td>deletion</td>
<td>insertion, deletion</td>
</tr>
<tr>
<td>Level of polymorphism</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>very high</td>
<td>High</td>
</tr>
<tr>
<td>Effective multiplex ratio</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>high</td>
<td>Medium</td>
</tr>
<tr>
<td>Marker index</td>
<td>Low</td>
<td>Codominant</td>
<td>Medium</td>
<td>Dominant</td>
<td>Medium</td>
</tr>
<tr>
<td>Inheritance</td>
<td>Codominant</td>
<td>Codominant</td>
<td>Medium</td>
<td>Dominant</td>
<td>Dominant</td>
</tr>
<tr>
<td>Detection of alleles</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Ease of use</td>
<td>labor intensive</td>
<td>Easy</td>
<td>Easy</td>
<td>difficult initially</td>
<td>Easy</td>
</tr>
<tr>
<td>Automation</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Reproducibility</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>medium to high</td>
</tr>
<tr>
<td>Type of probes/primers</td>
<td>low copy genomic DNA or cDNA clones</td>
<td>specific repeat DNA sequence</td>
<td>usually 10 bp random nucleotides</td>
<td>specific sequence</td>
<td>specific repeat DNA sequence</td>
</tr>
<tr>
<td>Cloning and/ or sequencing</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Radioactive detection</td>
<td>usually yes</td>
<td>No</td>
<td>no</td>
<td>yes/no</td>
<td>No</td>
</tr>
<tr>
<td>Development/start-up costs</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

\(a\) Level of polymorphism (average heterozygosity) is an average of the probability that two alleles taken at random can be distinguished

\(b\) Effective multiplex ratio is the number of polymorphic loci analyzed per experiment in the germplasm tested.

\(c\) Marker index is the product of the average expected heterozygosity and the effective multiplex ratio.
To achieve high reliability of the screen, genomic DNA is prepared in an ingenious, but technically straightforward way that combines the strengths of two methods, the replicability of restriction fragment analysis and the power of the PCR. This technique comprises of two principle steps. The first step is the restriction digestion of the genomic DNA and ligation of the adapters and the second step involves the selective amplification of the sets of restriction fragment. In the first step the DNA is digested with two restriction enzymes the hexa and tetra cutter producing sticky ends and double stranded synthetic adapters (oligonucleotide) of a defined sequence are then ligated to both cut ends of all restriction fragments. Adapters and restriction site sequences then provide universal primer binding sites for subsequent PCR. The selective amplification of the subset of the restriction fragment is carried out using primers that extend into the restriction fragments amplifying those fragments in which the primer extensions match the nucleotides flanking the restriction sites. The amplified restriction fragments are separated on denaturing polyacrylamide gel then visualized in autoradiograph.

AFLP markers have been used to infer phylogenetic relationships based on measures of genetic distance (Semblat et al., 1998). AFLPs are widely used to study inter and intra population genetic variation in many plant taxa (Ribeiro et al., 2002). Advantages of the AFLP are that no prior sequence information of the genome is required. A large number of polymorphic bands are produced and the technique is highly reproducible and standardized kits are available. The weaknesses are that there are number of steps involved in this technique. Additional expenditure is required for purchasing the enzymes and licensed kits. It is necessary to use polyacrylamide gel electrophoresis in a sequencing gel apparatus to separate radio-labelled fragments and for nonradioactive labeled fragments access to an automated sequencer is required.

**Gene Sequencing**

DNA sequencing is the most fundamental measure of diversity because it detects polymorphisms within the DNA's building blocks themselves. DNA sequence-based studies differ from DNA fingerprinting in data generation. Sequencing is essentially totally reproducible and generates large quantities of data very quickly. The choice of regions to sequence is increasing rapidly as our knowledge of the plant genome improves. DNA sequences often give higher resolution than other molecular markers (Kass and Wink, 1997). Recent development and use of molecular sequence data have increased significantly the understanding of plant systematics at various taxonomic levels (Soltis and Soltis, 1995).

Several genes from the organelle genomes e.g. mitochondria and chloroplast have been utilized for various molecular systematic studies, in particular the chloroplast genome, that has been extensively surveyed to reconstruct plant phylogeny (Olmstead and Palmer, 1994). The studies based on rbcL gene sequences located in the large single copy region, encoding the large subunit of ribulose 1, 5-biphosphate carboxylase/oxygenase, atpB located in the large single copy region encoding ß subunit of ATP synthetase, matK protein encoding region located in the large single copy region of the chloroplast genome, etc. have been successfully carried out resolving phylogeny of various taxa. Identification of easily amplifiable and relatively rapidly evolving unambiguously alignable, DNA regions that provide sufficient suitable variation within a short sequence segment from nuclear genome for carrying out lower level phylogenetic studies is greatly needed to supplement the wealth of chloroplast restriction site information that has accumulated over the period of time (Baldwin et al., 1995).

**Internal Transcribed Spacers (ITS) of nuclear ribosomal DNA**

Ribosomal DNA is the most conserved region in the genome, with capabilities of phylogenetic divergence. The whole rRNA gene contains a small subunit (SSU) 18S rRNA, 5.8S rRNA, and a large subunit (LSU) 28S rRNA. Internal transcribed spacers (ITS) are a part of the transcriptional unit of the nuclear ribosomal DNA (nrDNA). These internal transcribed spacers (ITS1 and ITS2), are one of the most extensively sequenced region of nrDNA. These are non coding regions of the eukaryotic cistron of ribosomal DNA that exist in several hundred copies and are located in one or several loci and distributed in one or several chromosomes. The ITS1 is located between the genes coding for 18S and 5.8S rRNA (an
Review Article

evolutionary highly conserved sequence), whereas ITS2 is located between the genes coding for 5.8S and 28S rRNA genes.

The internal transcribed spacers (ITS) have been the most popular target region in the nuclear genome for evolutionary studies of diverse plant groups (Hughes et al., 2006). The ITS region along with the other components of the nrDNA multigene family is highly repetitive in the plant nuclear genome. The entire nrDNA repeat unit is present in up to many thousands of copies arranged in tandem repeats at a chromosomal locus or at multiple loci (Du et al., 2011). The high copy number promotes detection, amplification, cloning and sequencing of nrDNA.

This gene family undergoes concerted evolution through unequal crossing over and gene conversion, which is most important from the standpoint of phylogeny reconstruction. This property promotes intragenomic uniformity of repeat units, in some cases between nrDNA loci and non homologous chromosomes (Wendel et al., 1995), and in general promotes accurate reconstruction of species relationships from these sequences. Also the small size of the ITS region (less than 700 bp in angiosperms) and the presence of highly conserved sequences flanking each of the two spacers make this region easy to amplify even from herbarium material (White et al., 1990).

ITS1 and ITS2 including 5.8S can be easily amplified and sequenced by PCR using universal primers ITS4 and ITS5 (White et al., 1990), and are routinely used to distinguish related species and to infer phylogenetic relationships from populations to families and even higher taxonomic levels (Coleman and Vacquier, 2002).

Single Nucleotide Polymorphism (SNPs)

SNPs are more advanced molecular markers based on DNA sequencing occurring when a single nucleotide- A-T-C or G- in the genome differs between members of a species or individuals. SNPs may occur in the coding, non-coding and intergenic regions of the genome. Since majority of the SNPs are located in noncoding DNA, they are infelicitously called noncoding SNPs. These are the most frequent type of variations found in DNA 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. SNPs are consequences of transition or transversion event and are highly abundant, their density differing substantially in different regions of the genome (Wiesing et al., 2005). SNP locus can have two, three or four alleles in a population, but biallelic SNPs massively prevail.

Several methods are available for SNP detection/sequencing, like pyrosequencing (Fakhrai-Rad et al., 2002), polymorphism ratio sequencing (Blazej et al., 2003), degenerate oligonucleotide primer PCR (Jordan et al., 2002), ecotilling (Comai et al., 2004), and SNP Hunter (Wang et al., 2005). SNPs are excellent markers for association mapping of genes controlling complex traits and provide the highest map resolution (Bhattramakki et al., 2002). SNPs has been successively used in plants species such as Barley (Kanazin et al., 2002; Rice (Bormans et al., 2002; Larkin et al., 2003), Maize (Batley et al., 2003), Wheat (Somers, 2003) and Sugar beet (Schneider et al., 2001).

Consensus Chloroplast Microsatellite Primers (CCMPs)

Information on the genetics of species would be valuable for scheming suitable plant breeding program, conservation of genetic resources, gene sequencing, gene mining or tagging etc. Genetic preservation strategies for their improved developmental progress of inherited information, inhabitants organization knowledge of plants, acquainted with eminence variability of chloroplast DNA, mitochondria and nuclear genome of plants due to its uniparental inheritance, the absence of recombination and premeditated mutation rates chloroplast genome is used (Provan et al., 2001). Since, the chloroplast genomes in higher plants are more conservative than mitochondrial and nucleic genomes (Wolfe et al., 1987), chloroplast microsatellites have the potentiality in the phylogenetic studies among plants with great taxonomic distances than nuclear or mitochondrial microsatellites (Ishii and McCouch, 2000). Therefore, the chloroplast genetic investigation was considered as a trustworthy technique to mark out the origin, evolution and phylogeny of many plant species. Usually, RFLPs and other genetic markers are the most
conventional method for the study of cytoplasmic inheritance in somatic hybrids of higher plants (Cheng et al., 2003). During last decade, cleaves amplified polymorphic sequence (CAPS) using mitochondrial or chloroplast specific primers are universal primers and are simple, inexpensive, valuable and extra competent in application in comparison to RFLPs (Bastia et al., 2001; Guo et al., 2002; Cheng et al., 2003). Chloroplast microsatellites or chloroplast simple sequence repeats (cpSSRs) are characteristically mononucleotide tandem repeats which has various realistic implications for genetic investigation (Weising and Gardner, 1999). Its amplification protocol has been described by Andrianoeina et al., (2006) who defined a combination of the different alleles as a chlorotype which is established at each locus for the reason of the non recombining quality of the chloroplast genome (chlorotypes were then treated as alleles at a distinct locus). For chloroplast microsatellites, the genetic diversity within D. monticola is one of the uppermost among tropical tree species assessed with chloroplast microsatellites (Muller et al., 2009) prevent the species extinction.

Different studies revealed that chloroplast SSR have been employed for information of execution of genetic conservation approach and restoration of populations in Madagascar (Crandall et al., 2000; Lhuillier et al., 2006; Muller et al., 2009).

Chloroplast SSR (cpSSR) markers have demonstrated their utility in studying genetic relationships, to assess the maternal and paternal plastid inheritance (Cato and Richardson, 1996), evaluation of interspecific polymorphism, detection of hybridization, introgression, phylogeny of plant population and are also applicable tool in the research of plant population genetics, understanding crop evolution, domestication, and phylogenetics (Provan et al., 2001). Comparing with RFLPs and CAPS, cpSSR are proved to be more convenient, proficient, simpler and less expensive for organelle analysis of Citrus somatic hybrids at a very early regeneration stage. It was useful to verify the chloroplast genomic origin of citrus somatic hybrids which is the first information on cytoplasmic inheritance analysis hybrids in higher plants by cpSSR. DNA sequence information of the chloroplast genome is necessary for the development of cpSSR primer pairs. These primers were designed according to the conserved nature of intron regions in chloroplast of higher plants. The relatively low numbers of repeats are typical of cpSSRs in other plant species, where long stretches of mono-nucleotide repeats are very exceptional (Powell et al., 1996; Provan et al., 1999a). In spite of these comparatively short repeat lengths, it has been revealed that cpSSR primers designed in one species will generate polymorphic products in other species and even in assorted genera (e.g. primers derived from Nicotiana tabacum revealing cpSSR polymorphism in Solanum spp.; Provan et al., 1999a).

They unanimously disclose intra-specific dissimilarity in repeat number when sited in the non-coding regions of the chloroplast genome (cpDNA). The results revealed that the chloroplast genomes in the somatic hybrids were indiscriminately inherited from either parent, which were documented in the previous reports on Citrus somatic hybrids being based on RFLPs analysis (Grosser et al., 1996, 2000; Moreria et al., 2000; Guo and Deng, 2001). In the past several years, cpSSR has turn out to be useful method for evaluating genetic diversity of the cpDNA genome in plant species such as pine (Pinus contorta) (Powell et al., 1995), potato (Solanum tuberosum) (Bryan et al., 1999), barley (Hordeum vulgare) (Provan et al., 1999), rice (Oryza sativa) (Ishii and McCouch, 2000), soybean (Glycine max) (Powell et al., 1996; Xu et al., 2002) and Kiwifruit (Actinidia delicosa) (Weising and Gardner, 1999). cpSSRs are used for identification (for example Dalbergia monticola) that may previously not have been recognized using any other molecular markers such as RFLPs, RAPDs, AFLPs, etc. whether there is any genetic discrepancy present in the chloroplast genome of numerous plants) and to study genetic variation within and among populations and geographical structure in established populations. Chloroplast SSRs have been used in population and systematic studies in a variety of species (Powell et al., 1996; Provan et al., 1999a). This facilitates the perceptive of both chronological and current actions. It provides a balancing observation of gene flow blueprint because in angiosperms chloroplasts are transmitted by seeds (Petit et al., 2005). In Pinaceae, to measure discontinuity within an efficiently haploid genome uniparental inheritance of mitochondrial and chloroplast genomes provides opportunities. However, in
plants due to a high rate of sequence reformation mitochondrial genomes have not typically been helpful for phylogenetic analyses (Sederoff et al., 1981; Wu et al., 1998). Chloroplast microsatellites are predominantly valuable markers for paternal inheritance of chloroplast genomes in most conifers for studying mating systems, uniparental lineages, and gene flow via both pollen and seeds in most conifers (Neale et al., 1986). Predominant paternal inheritance of chloroplast DNA has been established in European Abies with previously categorized (Vendramin et al., 1996) Pinus thunbergii primers at two extremely inconsistent microsatellite loci (Vendramin and Ziegenhagen, 1997; Ziegenhagen et al., 1998; Vendramin et al., 1999). Intra-specific diversity and population structure in heterologous amplification with these primers have also been demonstrated (Cato and Richardson, 1996; Morgante et al., 1998). In wild plant species and their varieties the potential of cpSSRs into biological and evolutionary processes has yet to be entirely accepted because of the intensifying outline of studies employing cpSSRs, while studies of economically imperative plants and their varieties remains obscured. Chloroplast markers are further sensitive to drift, one of the major effects of disintegration. The only limitation is because of the same number of repeats may evolve in two different microsatellite lineages through independent mutational events i.e. its homoplasy effect, (Navascue´s and Emerson, 2005) instead of that recent simulations have verified that chloroplast microsatellites are capable of studying genetic organization and gene flow (Hansen et al., 2005). Consequently to determine levels of inconsistency amongst populations, have great influence in estimating population structure information of mutation rates at SSR loci is significant. The low chloroplast DNA haplotypes is closely related to slow mutation rates (Provan et al., 2001). Till date there have been no statistics published for mutation rates at simple repeat loci in the chloroplast genome. Discrepancy can be detected among plant species using cpDNA-specific universal primers can be utilized for many applications in plant science. Among these, systematics and evolutionary relationships studies at the different taxonomic levels (e.g. interfamilial, inter-generic, inter-specific, intra-specific and inter-population level) have been by far the most widespread.

**Diversity Array Technology (DArt)**

Because of the limitations of existing marker technologies, Diversity Arrays Technology (DArT), a novel method to discover and score genetic polymorphic markers has been discovered (Jaccoud et al., 2001). DArT, developed by CAMBIA involve a new use of microarrays, detects single base change as well as insertions and deletions and also detects differences in DNA methylation depending on the enzyme used to generate the fragments. It does not require sequence knowledge, and thus may become very useful to crop researchers. This technique offers a low cost high throughput, robust system with minimal DNA sample requirement capable of providing comprehensive genome coverage even in organisms without any DNA sequence information (Jaccound et al., 2001). Data aquisition and analysis is fast. DArT genome profiles can be used for the recognition and management of biodiversity, for example in germplasm collections. Identification of duplicate accessions and a better understanding of the genetic relationships between the accessions could help to control the costs of maintaining these collections.

The main disadvantage of this technique is that these are dominant markers and are technically demanding.

As evident from the above paragraphs, there are large numbers of molecular markers that can be used to carry out the genetic diversity studies. Ideally the marker should have the following desirable properties:

- Marker should be highly polymorphic.
- Mode of inheritance should be codominant (which allows the discrimination of homo-and heterozygous states in diploid organisms).
- Frequent occurrence in the genome.
- Unambiguous assignment of the allele.
- Even distribution throughout the genome.
- Selectively neutral behaviour (i.e. no pleiotropic effects).
- Easy access (i.e. by purchasing or fast procedures).
- Easy and fast assay (e.g. by automated procedures).
Review Article

➢ High reproducibility.
➢ Easy exchange of data between laboratories.
➢ Low cost for both marker development and assay.

There is not a single marker which fulfills all of these criteria. The choice from the several molecular markers may be made, each of which combines at least some if not all of the above properties on the basis of the kind of study to be undertaken.

CONCLUSION

Molecular characterization can play a role in uncovering the history, and estimating the diversity, distinctiveness and population structure. Different marker types have different usefulness in studying genetic diversity. Fast changing markers can be used in studying closely related species. RFLP was the first molecular marker to be used in genetic diversity. Although reproducible, it is time consuming and in policy species it is of low frequency. RAPD marker is easy to perform, however; it has inherited problem of reproducibility. ISSR is more reproducible and polymorphic than RAPD. The SSR have been the marker of choice for the last two decades especially before the discovery of SNP’s. SNP are the most widespread sequence variation in the genome. They are numerous, more stable and easier to score than SSR. AFLP, although having high discriminatory power, it has medium reproducibility and alleles are not easily recognized. On the other hand the development of microarray based technologies such as DArT which have the merits of SNP’s without going through sequencing. It is medium- to ultra-high-throughput genotyping at a low cost. They have been shown to be particularly useful for genomes, where the level of polymorphism is low. They are expected to play an important role in crop improvement and will be used for a variety of studies including the development of high-density molecular maps, which may then be used for QTL interval mapping and for functional and evolutionary studies. One important characteristics of ITS region is that it is highly conserved intraspecifically, but variable between different species. It provides novel insights into plant evolution and hybridization and considered one of the most successfully used of nuclear genome in studying phylogenetic and genomic relationships of plants. The revolution in plant genomics has opened up new perspectives and opportunities for the plant breeders, who can now apply molecular markers to assess and enhance diversity in their germplasm collections, to introgress valuable traits from new sources and to identify genes that control key traits. Awareness of the level of genetic diversity and the proper management of genetic resources are important issues in modern scenario.

REFERENCES


Review Article


Review Article


Review Article


Review Article


Review Article


Wu JKV and Strauss SH (1999). Nuclear DNA diversity, population differentiation, and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. Genome 42 893-908.


