

## **DNA markers in plant genome analysis**

With the advent of molecular markers, a new generation of markers has been introduced over the last two decades, which has revolutionized the entire scenario of biological sciences. DNA-based molecular markers have acted as versatile tools and have found their own position in various fields like taxonomy, physiology, embryology, genetic engineering, etc. They are no longer looked upon as simple DNA fingerprinting markers in variability studies or as mere forensic tools. Ever since their development, they are constantly being modified to enhance their utility and to bring about automation in the process of genome analysis. The discovery of PCR (polymerase chain reaction) was a landmark in this effort and proved to be an unique process that brought about a new class of DNA profiling markers. This facilitated the development of marker-based gene tags, map-based cloning of agronomically important genes, variability studies, phylogenetic analysis, synteny mapping, marker-assisted selection of desirable genotypes, etc. Thus giving new dimensions to concerted efforts of breeding and marker-aided selection that can reduce the time span of developing new and better varieties and will make the dream of super varieties come true. These DNA markers offer several advantages over traditional phenotypic markers, as they provide data that can be analysed objectively.

Plants have always been looked upon as a key source of energy for survival and evolution of the animal kingdom, thus forming a base for every ecological pyramid. Over the last few decades plant genomics has been studied extensively bringing about a revolution in this area. Molecular markers, useful for plant genome analysis, have now become an important tool in this revolution. In this article we attempt to review most of the available DNA markers that can be routinely employed in various aspects of plant genome analysis such as taxonomy, phylogeny, ecology, genetics and plant breeding.

During the early period of research, classical strategies including comparative anatomy, physiology and embryology were employed in genetic analysis to determine inter- and intra-species variability. In the past decade, however, molecular markers have very rapidly complemented the classical strategies. Molecular markers include biochemical constituents (e.g. secondary metabolites in plants) and macromolecules, viz. proteins and deoxyribonucleic acids (DNA). Analysis of secondary metabolites is, however, restricted to those plants that produce a suitable range of metabolites which can be easily analysed and which can distinguish between varieties. These metabolites which are being used as

markers should be ideally neutral to environmental effects or management practices. Hence, amongst the molecular markers used, DNA markers are more suitable and ubiquitous to most of the living organisms.

### **DNA-based molecular markers**

Genetic polymorphism is classically defined as the simultaneous occurrence of a trait in the same population of two or more discontinuous variants or genotypes. Although DNA sequencing is a straightforward approach for identifying variations at a locus, it is expensive and laborious. A wide variety of techniques have, therefore, been developed in the past few years for visualizing DNA sequence polymorphism.

The term DNA-fingerprinting was introduced for the first time by Alec Jeffrey in 1985 to describe bar-code-like DNA fragment patterns generated by multilocus probes after electrophoretic separation of genomic DNA fragments. The emerging patterns make up a unique feature of the analysed individual and are currently considered to be the ultimate tool for biological individualization. Recently, the term DNA fingerprinting/profiling is used to describe the combined use of several single locus detection systems and is being used as versatile tools for investigating various aspects of plant genomes. These include characterization of genetic variability, genome fingerprinting, genome mapping, gene localization, analysis of genome evolution, population genetics, taxonomy, plant breeding, and diagnostics.

### **Properties desirable for ideal DNA markers**

- Highly polymorphic nature
- Codominant inheritance (determination of homozygous and heterozygous states of diploid organisms)
- Frequent occurrence in genome
- Selective neutral behaviour (the DNA sequences of any organism are neutral to environmental conditions or management practices)
- Easy access (availability)
- Easy and fast assay
- High reproducibility
- Easy exchange of data between laboratories.

It is extremely difficult to find a molecular marker which would meet all the above criteria. Depending on the type of study to be undertaken, a marker system can be identified that would fulfill atleast a few of the above characteristics.

### **Types of molecular markers**

Various types of molecular markers are utilized to evaluate DNA polymorphism and are generally classified as hybridization-based markers and polymerase chain reaction (PCR)-based markers. In the former, DNA profiles are visualized by hybridizing the restriction enzyme-digested DNA, to a labelled probe, which is a DNA fragment of known origin or sequence. PCR-based markers involve *in vitro* amplification of particular DNA sequences or loci, with the help of specifically or arbitrarily chosen oligonucleotide sequences (primers) and a thermostable DNA polymerase enzyme. The amplified fragments are separated electrophoretically and banding patterns are detected by different methods such as staining and autoradiography. PCR is a versatile technique invented during the mid-1980s. Ever since thermostable DNA polymerase was introduced in 1988, the use of PCR in research and clinical laboratories has increased tremendously. The primer sequences are chosen to allow base-specific binding to the template in reverse orientation. 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.

For simplicity, we have divided the review in two parts. The first part is a general description of most of the available DNA marker types, while the second includes their application in plant genomics and breeding programmes.

### **Types and description of DNA markers**

#### ***Single or low copy probes***

*Restriction fragment length polymorphism* (RFLP). RFLPs are simply inherited naturally occurring Mendelian characters. They have their origin in the DNA rearrangements that occur due to evolutionary processes, point mutations within the restriction enzyme recognition site sequences, insertions or deletions within the fragments, and unequal crossing over.

In RFLP analysis, restriction enzyme-digested genomic DNA is resolved by gel electrophoresis and then blotted on to a nitrocellulose membrane. Specific banding patterns are then visualized by hybridization with labelled probe. These probes are mostly species-

specific single locus probes 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 and almost all sequence types are included; however, a large number of interspersed repeats are found in inserts, that detect a large number of restriction fragments forming complex patterns. In plants, this problem is overcome to some extent by using methylation-sensitive restriction enzyme *Pst*I. This helps to obtain low copy DNA sequences of small fragment sizes, which are preferred in RFLP analysis. On the other hand cDNA libraries are difficult to construct, however, they are more popular as actual genes are analysed and they contain fewer repeat sequences. The selection of appropriate source for RFLP probe varies, with the requirement of particular application under consideration. Though genomic library probes may exhibit greater variability than gene probes from cDNA libraries, a few studies reveal the converse. This observation may be because cDNA probes not only detect variation in coding regions of the corresponding genes but also regions flanking genes and introns of the gene.

RFLP markers were used for the first time in the construction of genetic maps by Botstein *et al.* RFLPs, being codominant markers, can detect coupling phase of DNA molecules, as DNA fragments from all homologous chromosomes are detected. They are very reliable markers in linkage analysis and breeding and can easily determine if a linked trait is present in a homozygous or heterozygous state in an individual, an information highly desirable for recessive traits<sup>12</sup>. However, their utility has been hampered due to the large amount of DNA required for restriction digestion and Southern blotting. The requirement of radioactive isotope makes the analysis relatively expensive and hazardous. The assay is time-consuming and labour-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.

### ***Restriction landmark genomic scanning (RLGS)***

This method, introduced for the first time by Hatada *et al.*, for genomic DNA analysis of higher organisms, is based on the principle that restriction enzyme sites can be used as landmarks. It employs direct labelling of genomic DNA at the restriction site and two-dimensional (2D) electrophoresis to resolve and identify these landmarks. The technique has proven its utility in genome analysis of closely-related cultivars and for obtaining polymorphic markers that can be cloned by spot target method. It has been used as a new fingerprinting technique for rice cultivars.

## **RFLP markers converted in to PCR based-markers**

### ***Sequence-tagged sites (STS)***

RFLP probes specifically linked to a desired trait can be converted into PCR-based STS markers based on nucleotide sequence of the probe giving polymorphic band pattern, to obtain specific amplicon. Using this technique, tedious hybridization procedures involved in RFLP analysis can be overcome. This approach is extremely useful for studying the relationship between various species. When these markers are linked to some specific traits, for example powdery mildew resistance gene or stem rust resistance gene in barley, they can be easily integrated into plant breeding programmes for marker-assisted selection of the trait of interest.

### ***Allele-specific associated primers (ASAPs)***

To obtain an allele-specific marker, specific allele (either in homozygous or heterozygous state) is sequenced and specific primers are designed for amplification of DNA template to generate a single fragment at stringent annealing temperatures. These markers tag specific alleles in the genome and are more or less similar to SCARs.

### ***Expressed sequence tag markers (EST)***

This term was introduced by Adams *et al.* Such markers are obtained by partial sequencing of random cDNA clones. Once generated, they are useful in cloning specific genes of interest and synteny mapping of functional genes in various related organisms. ESTs are popularly used in full genome sequencing and mapping programmes underway for a number of organisms and for identifying active genes thus helping in identification of diagnostic markers. Moreover, an EST that appears to be unique helps to isolate new genes. EST markers are identified to a large extent for rice, *Arabidopsis*, etc. wherein thousands of functional cDNA clones are being converted in to EST markers.

### ***Single strand conformation polymorphism (SSCP)***

This is a powerful and rapid technique for gene analysis particularly for detection of point mutations and typing of DNA polymorphism. SSCP can identify heterozygosity of DNA fragments of the same molecular weight and can even detect changes of a few nucleotide bases as the mobility of the single-stranded DNA changes with change in its GC content due to its conformational change. To overcome problems of reannealing and complex banding patterns, an improved technique called asymmetric-PCR SSCP was developed, wherein the denaturation step was eliminated and a large-sized sample could be loaded for

gel electrophoresis, making it a potential tool for high throughput DNA polymorphism. It was found useful in the detection of heritable human diseases. In plants, however, it is not well developed although its application in discriminating progenies can be exploited, once suitable primers are designed for agronomically important traits.

### **Multi locus probes**

#### **Repetitive DNA**

A major step forward in genetic identification is the discovery that about 30–90% of the genome of virtually all the species is constituted by regions of repetitive DNA, which are highly polymorphic in nature. These regions contain genetic loci comprising several hundred alleles, differing from each other with respect to length, sequence or both and they are interspersed in tandem arrays ubiquitously. The repetitive DNA regions play an important role in absorbing mutations in the genome. Of the mutations that occur in the genome, only inherited mutations play a vital role in evolution or polymorphism. Thus repetitive DNA and mutational forces functional in nature together form the basis of a number of marker systems that are useful for various applications in plant genome analysis. The markers belonging to this class are both hybridization-based and PCR-based.

#### ***Microsatellites and minisatellites***

The term microsatellite was coined by Litt and Lutty, while the term minisatellite was introduced by Jeffrey. Both are multilocus probes creating complex banding patterns and are usually non-species specific occurring ubiquitously. They essentially belong to the repetitive DNA family. Fingerprints generated by these probes are also known as oligonucleotide fingerprints. The methodology has been derived from RFLP and specific fragments are visualized by hybridization with a labelled micro- or minisatellite probe.

Minisatellites are tandem repeats with a monomer repeat length of about 11–60 bp, while microsatellites or short tandem repeats/simple sequence repeats (STRs/SSRs) consist of 1 to 6 bp long monomer sequence that is repeated several times. These loci contain tandem repeats that vary in the number of repeat units between genotypes and are referred to as variable number of tandem repeats (VNTRs) (i.e. a single locus that contains variable number of tandem repeats between individuals) or hypervariable regions (HVRs) (i.e. numerous loci containing tandem repeats within a genome generating high levels of polymorphism between individuals). Microsatellites and minisatellites thus form an ideal marker system creating complex banding patterns by simultaneously detecting

multiple DNA loci. Some of the prominent features of these markers are that they are dominant fingerprinting markers and codominant STMS (sequence tagged microsatellites) markers. Many alleles exist in a population, the level of heterozygosity is high and they follow Mendelian inheritance.

### **Minisatellite and microsatellite sequences converted into PCR-based markers**

#### ***Sequence-tagged microsatellite site markers (STMS)***

This method includes DNA polymorphism using specific primers designed from the sequence data of a specific locus. Primers complementary to the flanking regions of the simple sequence repeat loci yield highly polymorphic amplification products. Polymorphisms appear because of variation in the number of tandem repeats (VNTR loci) in a given repeat motif. Tri- and tetranucleotide microsatellites are more popular for STMS analysis because they present a clear banding pattern after PCR and gel electrophoresis. However, dinucleotides are generally abundant in genomes and have been used as markers e.g.  $(CA)_n(AG)_n$  and  $(AT)_n$ . The di- and tetranucleotide repeats are present mostly in the non-coding regions of the genome, while 57% of trinucleotide repeats are shown to reside in or around the genes. A very good relationship between the number of alleles detected and the total number of simple repeats within the targeted microsatellite DNA has been observed. Thus larger the repeat number in the microsatellite DNA, greater is the number of alleles detected in a large population.

#### ***Direct amplification of minisatellite DNA markers (DAMD-PCR)***

This technique, introduced by Heath *et al.*, has been explored as a means of generating DNA probes useful for detecting polymorphism. DAMD-PCR clones can yield individual-specific DNA fingerprinting pattern and thus have the potential as markers for species differentiation and cultivar identification.

#### ***Inter simple sequence repeat markers (ISSR)***

In this technique, reported by Zietkiewicz *et al.*, primers based on microsatellites are utilized to amplify inter-SSR DNA sequences. Here, various microsatellites anchored at the 3' end are used for amplifying genomic DNA which increases their specificity. These are mostly dominant markers, though occasionally a few of them exhibit codominance. An unlimited number of primers can be synthesized for various combinations of di-, tri-, tetra- and pentanucleotides [ $(4)^3 = 64$ ,  $(4)^4 = 256$ ] etc. with an anchor made up of a few bases and can be exploited for a broad range of applications in plant species.

## **Other repetitive DNA-type markers**

### ***Transposable elements***

A large number of transposable repeat elements have been studied in plants; however, only a few have been exploited as molecular markers. In evolutionary terms, they have contributed to genetic differences between species and individuals by playing a role in retrotransposition events promoting unequal crossing over. Retrotransposon-mediated fingerprinting has been shown to be an efficient fingerprinting method for detection of genetic differences between different species.

### ***Alu-repeats***

This strategy was developed to fingerprint genotypes using semi specific primers, complementary to repetitive DNA elements called 'Alu-repeats', in human genome analysis. *Alu* repeats are a class of randomly repeated interspersed DNA, preferentially used for *Alu* PCR as they reveal considerable levels of polymorphism. These representatives of short and long interspersed nuclear elements are known as *SINES*. *Alu* elements are approximately 300 bp in size and have been suggested to be originated from special RNA species that have been reintegrated at a rate of approximately one integration event per 10000 years. These repeats have been studied largely in humans, while their function in plants remains largely unexplored.

### ***Repeat complementary primers***

As an alternative to the interspersed repeats, primers complementary to other repetitive sequence elements were also successfully used for generation of polymorphisms, e.g. introns/exons splice junctions, tRNA genes, 5sRNA genes and Zn-finger protein genes. Primers complementary to specific exons, resulting in the amplification of the intervening introns have been studied by Lessa *et al.* One of the strengths of these new strategies is that they are more amenable to automation than the conventional hybridization-based techniques.

### ***Arbitrary sequence markers***

#### ***Randomly-amplified polymorphic DNA markers (RAPD)***

In 1991 Welsh and McClelland developed a new PCR-based genetic assay namely randomly amplified polymorphic DNA (RAPD). This procedure detects nucleotide sequence polymorphisms in DNA by using a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer anneals to the genomic DNA at two different sites on

complementary strands of DNA template. If these priming sites are within an amplifiable range of each other, a discrete DNA product is formed through thermocyclic amplification. 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. However, due to the stoichiometric nature of DNA amplification with random sequence primers, it is important to optimize and maintain consistent reaction conditions for reproducible DNA amplification. They are dominant markers and hence have limitations in their use as markers for mapping, which can be overcome to some extent by selecting those markers that are linked in coupling<sup>43</sup>. RAPD assay has been used by several groups as efficient tools for identification of markers linked to agronomically important traits, which are introgressed during the development of near isogenic lines. The application of RAPDs and their related modified markers in variability analysis and individual-specific genotyping has largely been carried out, but is less popular due to problems such as poor reproducibility faint or fuzzy products, and difficulty in scoring bands, which lead to inappropriate inferences.

### **Some variations in the RAPD technique include**

#### **DNA amplification fingerprinting (DAF)**

Caetano-Anolles *et al.* employed single arbitrary primers as short as 5 bases to amplify DNA using polymerase chain reaction. In a spectrum of products obtained, simple patterns are useful as genetic markers for mapping, while more complex patterns are useful for DNA fingerprinting. Band patterns are reproducible and can be analysed using polyacrylamide gel electrophoresis and silver staining. DAF requires careful optimization of parameters; however, it is extremely amenable to automation and fluorescent tagging of primers for early and easy determination of amplified products. DAF profiles can be tailored by employing various modifications such as predigesting of template. This technique has been useful in genetic typing and mapping.

#### **Arbitrary primed polymerase chain reaction (AP-PCR)**

This is a special case of RAPD, wherein discrete amplification patterns are generated by employing single primers of 10–50 bases in length in PCR of genomic DNA. In the first two cycles, annealing is under non-stringent conditions. The final products are structurally similar to RAPD products. Compared to DAF, this variant of RAPD is not very popular as it involves autoradiography. Recently, however, it has been simplified by separating the fragments on agarose gels and using ethidium bromide staining for visualization.

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

Michelmore *et al.* and Martin *et al.* introduced this technique wherein the RAPD marker termini are sequenced and longer primers are designed (22–24 nucleotide bases long) for specific amplification of a particular locus. These are similar to STS markers in construction and application. The presence or absence of the band indicates variation in sequence. These are better reproducible than RAPDs. SCARs are usually dominant markers, however, some of them can be converted into codominant markers by digesting them with tetra cutting restriction enzymes and polymorphism can be deduced by either denaturing gel electrophoresis or SSCP. Compared to arbitrary primers, SCARs exhibit several advantages in mapping studies (codominant SCARs are informative for genetic mapping than dominant RAPDs), map-based cloning as they can be used to screen pooled genomic libraries by PCR, physical mapping, locus specificity, etc. SCARs also allow comparative mapping or homology studies among related species, thus making it an extremely adaptable concept in the near future.

### **Cleaved amplified polymorphic sequences (CAPs)**

These polymorphic patterns are generated by restriction enzyme digestion of PCR products. Such digests are compared for their differential migration during electrophoresis. PCR primer for this process can be synthesized based on the sequence information available in databank of genomic or cDNA sequences or cloned RAPD bands. These markers are codominant in nature.

### **Randomly amplified microsatellite polymorphisms (RAMPO)**

In this PCR-based strategy, genomic DNA is first amplified using arbitrary (RAPD) primers. The amplified products are then electrophoretically separated and the dried gel is hybridized with microsatellite oligonucleotide probes. Several advantages of oligonucleotide fingerprinting, RAPD and microsatellite-primed PCR are thus combined, these being the speed of the assay, the high sensitivity, the high level of variability detected and the nonrequirement of prior DNA sequence information. This technique has been successfully employed in the genetic fingerprinting of tomato, kiwi fruit and closely-related genotypes of *D. bulbifera*.

### **Amplified fragment length polymorphism (AFLP)**

A recent approach by Zabeau *et al.*, known as AFLP, is a technique based on the detection of genomic restriction fragments by PCR amplification and can be used for DNAs of any

origin or complexity. The fingerprints are produced, without any prior knowledge of sequence, using a limited set of generic primers. The number of fragments detected in a single reaction can be ‘tuned’ by selection of specific primer sets. AFLP technique is reliable since stringent reaction conditions are used for primer annealing. This technique thus shows an ingenious combination of RFLP and PCR techniques and is extremely useful in detection of polymorphism between closely related genotypes.

### **AFLP procedure mainly involves 3 steps**

- (a) Restriction of DNA using a rare cutting and a commonly cutting restriction enzyme simultaneously (such as *Mse*I and *Eco*RI) followed by ligation of oligonucleotide adapters, of defined sequences including the respective restriction enzyme sites.
- (b) Selective amplifications of sets of restriction fragments, using specifically designed primers. To achieve this, the 5' region of the primer is made such that it would contain both the restriction enzyme sites on either sides of the fragment complementary to the respective adapters, while the 3' ends extend for a few arbitrarily chosen nucleotides into the restriction fragments.
- (c) Gel analysis of the amplified fragments. AFLP analysis depicts unique fingerprints regardless of the origin and complexity of the genome. Most AFLP fragments correspond to unique positions on the genome and hence can be exploited as landmarks in genetic and physical mapping. AFLPs are extremely useful as tools for DNA fingerprinting and also for cloning and mapping of variety-specific genomic DNA sequences. Similar to RAPDs, the bands of interest obtained by AFLP can be converted into SCARs. Thus AFLP provides a newly developed, important tool for a variety of applications.

Table 3. Comparison of the five most widely used DNA markers in plants.

	RFLP	Microsatellite	RAPD	AFLP	ISSR
Genomic abundance	high	medium	very high	very high	medium
Part of genome surveyed	low copy coding regions	whole genome	whole genome	whole genome	whole genome
Amount of DNA required	high	low	low	medium	low
Type of polymorphism	single base changes, insertion, deletion	changes in length of repeats	single base changes, insertion, deletion	single base changes, insertion, deletion	single base changes, insertion, deletion
Level of polymorphism <sup>a</sup>	medium	high	high	very high	high
Effective multiplex ratio <sup>b</sup>	low	medium	medium	high	medium
Marker index <sup>c</sup>	low	medium	medium	high	medium
Inheritance	codominant	codominant	dominant	dominant	dominant
Detection of alleles	yes	yes	no	no	no
Ease of use	labour intensive	easy	easy	difficult initially	easy
Automation	low	high	medium	medium	medium
Reproducibility (reliability)	high	high	intermediate	high	medium to high
Type of probes/primers	low copy genomic DNA or cDNA clones	specific repeat DNA sequence	usually 10 bp random nucleotides	specific sequence	specific repeat DNA sequence
Cloning and/or sequencing	yes	yes	no	no	no
Radioactive detection	usually yes	no	no	yes/no	no
Development/start-up costs	high	high	low	medium	medium
Utility for genetic mapping	species specific	species specific	cross specific	cross specific	cross specific
Proprietary rights status	No	No (some are licensed)	licensed	licensed	no

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

<sup>b</sup> Effective multiplex ratio is the number of polymorphic loci analysed per experiment in the germplasm tested.

<sup>c</sup> Marker index is the product of the average expected heterozygosity and the effective multiplex ratio.

Table 1. Advantages and disadvantages of most commonly-used DNA markers for QTL analysis

Molecular marker	Codominant (C) or Dominant (D)	Advantages	Disadvantages	References
Restriction fragment length polymorphism (RFLP)	C	<ul style="list-style-type: none"> <li>• Robust</li> <li>• Reliable</li> <li>• Transferable across populations</li> </ul>	<ul style="list-style-type: none"> <li>• Time-consuming, laborious and expensive</li> <li>• Large amounts of DNA required</li> <li>• Limited polymorphism (especially in related lines)</li> </ul>	Beckmann & Soller (1986), Kochert (1994), Tanksley et al. (1989)
Random amplified polymorphic DNA (RAPD)	D	<ul style="list-style-type: none"> <li>• Quick and simple</li> <li>• Inexpensive</li> <li>• Multiple loci from a single primer possible</li> <li>• Small amounts of DNA required</li> </ul>	<ul style="list-style-type: none"> <li>• Problems with reproducibility</li> <li>• Generally not transferable</li> </ul>	Penner (1996), Welsh & McClelland (1990), Williams et al. (1990)
Simple sequence repeats (SSRs)* or 'microsatellites'	C	<ul style="list-style-type: none"> <li>• Technically simple</li> <li>• Robust and reliable</li> <li>• Transferable between populations</li> </ul>	<ul style="list-style-type: none"> <li>• Large amounts of time and labour required for production of primers</li> <li>• Usually require polyacrylamide electrophoresis</li> </ul>	McCouch et al. (1997), Powell et al. (1996), Taramino & Tingey (1996)
Amplified fragment Length Polymorphism (AFLP)	D	<ul style="list-style-type: none"> <li>• Multiple loci</li> <li>• High levels of polymorphism generated</li> </ul>	<ul style="list-style-type: none"> <li>• Large amounts of DNA required</li> <li>• Complicated methodology</li> </ul>	Vos et al. (1995)

\*SSRs are also known as sequence tagged microsatellite site (STMS) markers (Davierwala et al., 2000; Huettel et al., 1999; Mohapatra et al., 2003; Winter et al., 1999).

## **Applications of molecular markers in plant science**

Molecular markers have been looked upon as tools for a large number of applications ranging from localization of a gene to improvement of plant varieties by marker-assisted selection. They have also become extremely popular markers for phylogenetic analysis adding new dimensions to the evolutionary theories. If we look at the history of the development of these markers, it is evident that they have been improved over the last two decades to provide easy, fast and automated assistance to scientists and breeders. Genome analysis based on molecular markers has generated a vast amount of information and a number of databases are being generated to preserve and popularize it.

### **Mapping and tagging of genes: Generating tools for marker-assisted selection in plant breeding**

Plant improvement, either by natural selection or through the efforts of breeders, has always relied upon creating, evaluating and selecting the right combination of alleles. The manipulation of a large number of genes is often required for improvement of even the simplest of characteristics. With the use of molecular markers it is now a routine to trace valuable alleles in a segregating population and mapping them. These markers once mapped enable dissection of the complex traits into component genetic units more precisely, thus providing breeders with new tools to manage these complex units more efficiently in a breeding programme.

The very first genome map in plants was reported in maize, followed by rice, *Arabidopsis* etc. using RFLP markers. Maps have since then been constructed for several other crops like potato, barley, banana, members of Brassicaceae, etc. Once the framework maps are generated, a large number of markers derived from various techniques are used to saturate the maps as much as is possible. Microsatellite markers, especially STMS markers, have been found to be extremely useful in this regard. Owing to their quality of following clear Mendelian inheritance, they can be easily used in the construction of index maps, which can provide an anchor or reference point for specific regions of the genome. About 30 microsatellites have already been assigned to five linkage groups in *Arabidopsis*, while their integration into the genetic linkage maps is still in progress in rice, soybean, maize, etc. The very first attempt to map microsatellites in plants was made by Zhao and Kochert in rice using (GGC)*n*, followed by mapping of (GA)*n* and (GT)*n* by Tanksley *et al.* and (GA/AG)*n*, (ATC) 10 and (ATT) 14, by Panaud *et al.* in rice. The most recent microsatellite map has

been generated by Milbourne *et al.* for potato. Similar to microsatellites, looking at the pattern of variation, generated by retrotransposons, it is now proposed that apart from genetic variability, these markers are ideal for integrating genetic maps.

Once mapped, these markers are efficiently employed in tagging several individual traits that are extremely important for a breeding programme like yield, disease resistance, stress tolerance, seed quality, etc. A large number of monogenic and polygenic loci for various traits have been identified in a number of plants, which are currently being exploited by breeders and molecular biologists together, so as to make the dream of marker-assisted selection come true. Tagging of useful genes like the ones responsible for conferring resistance to plant pathogen, synthesis of plant hormones, drought tolerance and a variety of other important developmental pathway genes, is a major target. Such tagged genes can also be used for detecting the presence of useful genes in the new genotypes generated in a hybrid programme or by other methods like transformation, etc. RFLP markers have proved their importance as markers for gene tagging and are very useful in locating and manipulating quantitative trait loci (QTL) in a number of crops. The very first reports on gene tagging were from tomato, availing the means for identification of markers linked to genes involved in several traits like water use efficiency, resistance to *Fusarium oxysporum* (the 12 gene), leaf rust resistance genes LR 9 and 24 (refs 78, 79), and root knot nematodes (*Meloidogyne* sp.) (the mi gene). Recently, Xiao *et al.* have shown the utility of RFLP markers in identifying the trait improving QTL alleles from wild rice relative *O. rufipogon*.

Allele-specific associated primers have also exhibited their utility in genotyping of allelic variants of loci that result from both size differences and point mutations. Some of the genuine examples of this are the waxy gene locus in maize, the Glu D1 complex locus associated with bread making quality in wheat, the *Lr1* leaf rust resistance locus in wheat, the Gro1 and H1 alleles conferring resistance to the root cyst nematode *Globodera rostochiensis* in potato, and allele-specific amplification of polymorphic sites for detection of powdery mildew resistance loci in cereals. A number of other traits have been tagged using ASAPs in tomato, lettuce, etc. Besides ASAPs, AFLP and SSR markers have been identified to be associated with quantitative resistance to *Globodera pallida* (stone) in tetraploid potato, which can be very well employed in marker-assisted selection.

STMS markers have displayed a potential use as diagnostic markers for important traits in plant breeding programmes, e.g. (AT) 15 repeat has been located within a soybean heat shock protein gene which is about 0.5 cM from (Rsv) a gene conferring resistance to soybean mosaic virus. Several resistance genes including peanut mottle virus (Rpv), phytophera (Rps3) and Javanese rootknot nematode are clustered in this region of the soybean genome.

Similar to specific markers like RFLPs, STMS and ASAPs, arbitrary markers like RAPDs have also played important role in saturation of the genetic linkage maps and gene tagging. Their use in mapping has been especially important in systems, where RFLPs have failed to reveal much polymorphism that is so very important for mapping. One of the first uses of RAPD markers in saturation of genetic maps was reported by Williams *et al.* They have proven utility in construction of linkage maps among species where there has been inherent difficulty in producing F<sub>2</sub> segregating populations and have large genome size, e.g. conifers. RAPD markers in near isogenic lines can be converted into SCARs and used as diagnostic markers. SCAR/STS marker linked to the translocated segment on 4 AL of bread wheat carrying the *Lr28* gene has been tagged by Naik *et al.* Recently, ISSRs, which too belong to the arbitrary marker category, but are found to be devoid of many of the drawbacks shared by RAPD class of markers, have been employed as a reliable tool for gene tagging. An ISSR marker (AG) 8YC has been found to be linked closely ( $3.7 \pm 1.1$  cM) to the rice nuclear restorer gene, RF1 for fertility. RF1 is essential for hybrid rice production and this marker would be useful not only for breeding both restorer and maintainer lines, but also for the purity management of hybrid rice seeds. Similarly ISSR marker (AC)<sub>8</sub> YT has been found to be linked to the gene for resistance to *Fusarium* wilt race 4 in repulsion at a distance of 5.2 cM in chickpea.

Apart from mapping and tagging of genes, an important utility of RFLP markers has been observed in detecting gene introgression in a backcross breeding programme, and synteny mapping among closely related species. Similar utility of STMS markers has been observed for reliable pre selection in a marker assisted selection backcross scheme. Apart from specific markers, DAMD-based DNA fingerprinting in wheat has also been useful for monitoring backcross-mediated genome introgression in hexaploid wheat.

### **Phylogeny and evolution**

Most of the early theories of evolution were based on morphological and geographical variations between organisms. However, it is becoming more and more evident that the techniques from molecular biology hold a promise of providing detailed information about the genetic structure of natural population, than what we have been able to achieve in the past. RFLP, DNA sequencing, and a number of PCR-based markers are being used extensively for reconstructing phylogenies of various species. The techniques are speculated to provide path-breaking information regarding the fine time scale on which closely related species have diverged and what sort of genetic variations are associated with species formation. Furthermore, these studies hold a great promise for revealing more about the pattern of genetic variation within species.

Efforts are now being made for studying the genetic variation in plants, so as to understand their evolution from wild progenitors and to classify them into appropriate groups. The taxonomic classification is an essential first step to determine whether any germplasm is a part of the primary, secondary or tertiary gene pool of the system concerned. This is especially important in cases where morphological markers can prove to be inaccurate and misleading. A genuine example of this is the lines Azucena and PR 304 which have been classified as indicas using morphological characters, whereas they behave like japonicas in crossing studies. These samples are however, clearly revealed to be japonicas upon being analysed by RAPD markers.

RFLPs have been used in evolutionary studies for deducing the relationship between the hexaploid genome of bread wheat and its ancestors. Similarly a number of transposon elements like *tos1-1*, *tos2-1* and *tos3-1* retrotransposons have been used to detect the genetic differences between different species of rice and even between ecotypes of cultivated rice, wherein they were found to distinguish between the cultivars of Asian and African rice, *O. sativa* and *O. glaberrima*. Retroelement *Wis-2* has been found to detect genomic variation within individual plants of wheat variety and also within and between varieties of wheat. This element has also been found to occur in the genomes of other grasses like barley, rye, oats and *Aegilops* species, indicating common ancestral elements in grasses. Though RFLPs, microsatellites, minisatellites and transposons are useful for carrying out genetic variability analysis, the trend is now shifting towards the use of PCR-based markers. Specific markers are preferred over arbitrary primers. However, arbitrary primers are found to be the markers of choice in the analysis of complex genomes like

wheat, where genetic variation is extremely difficult to dissect. Sen *et al.* have used DAF as a new source of molecular markers in fingerprinting of bread wheat. Recent studies in our laboratory have revealed the utility of RAPD and ISSR markers in evolutionary studies of wheat and rice, respectively.

Specific markers like STMS (sequence-tagged microsatellite markers) ALPs (Amplicon length polymorphisms) or STS markers have proved to be extremely valuable in the analysis of gene pool variation of crops during the process of cultivar development, and classification of germplasm. These markers are extremely sensitive and can detect allelic variability during cultivar development. STS markers specific to chloroplast or mitochondrial DNA have been useful in providing seed and pollen specific markers which can be utilized for the detection of length variation at multiple physically linked sites and may be used to provide haplotype data and thus genotypically unique individual plants. Also a comparison of patterns of variability detected with biparentally (nuclear) and uniparentally (organellar) transmitted markers can provide complementary information to population and evolutionary biologists. Excellent examples of this are the Poly A mononucleotide repeats in maize, Poly (TA/AT) dinucleotide repeats found in liverworts, maize, pea and non photosynthetic green plant *Epifagus virginiana*, and a total of 500 chloroplast SSRs identified with repeat motifs greater than 10 repeat units in rice, tobacco, black pine, liverwort and maize. Though all these marker types provide valuable information regarding the evolution and phylogeny of various species being studied in any given set of samples, the trend is now shifting towards the use of ESTs (expressed sequence tags) for such analysis. This may be so, because in such studies, one actually looks at the evolution of functional genes.

### **Diversity analysis of exotic germplasm**

Following domestication, genetic variation in crop plants has continued to narrow due to continuous selection pressure for specific traits i.e. yield, thus rendering them more vulnerable to disease and insect epidemics and jeopardizing the potential for sustained genetic improvement over a long term. This risk was brought sharply into focus in 1970 with the outbreak of southern corn leaf blight which drastically reduced corn yield in USA, and was attributed to extensive use of a single genetic male sterility factor which was unfortunately linked to the disease susceptibility. Thus it is extremely important to study the genetic composition of the germplasm of existing modern-day cultivars in comparison with their ancestors and related species. This will not only provide information on their phylogenetic relationship but will also indicate a chance of finding new and useful genes, as

the accessions with most distinct DNA profiles are likely to contain a greater number of novel alleles. DNA profiling to make such sampling decisions is now underway in most crops. The exotic germplasm for breeding is selected on the basis of certain characteristic features such as (a) the exotic germplasm must possess a significant number of unique DNA polymorphisms (throughout the genome) relative to the modern-day cultivars and (b) each exotic germplasm has to be genetically dissimilar (on the basis of DNA profiling). This is necessary to understand the genetic variations between the existing cultivars, the cultivars in comparison with their wild progenitors and a number of wilds that still exist in nature.

Many DNA markers both specific as well as arbitrary have been used so far, for DNA fingerprinting of various classes of germplasm. Further studies with STMS markers may also throw light on the domestication process involved in crop plants and provide useful criteria for enriching the gene pool of crop plants and determine how efficient plant breeders have been in accessing preexisting forms of variation. AFLP, a new class of molecular markers, has gained popularity as marker for the study of genetic polymorphism especially in species where polymorphism is extremely rare using other types of markers. Pakniyat *et al.* used AFLP for studying variation in wild barley with reference to salt tolerance and associated eco geography, and a number of reports are coming up each day for different systems. Similarly the potential of ISSR markers has been exploited for diversity analysis of pine, rice and also in wheat. These studies have helped in the classification of existing biodiversity among plants, which can be further exploited in wild gene introgression programs.

### **Genotyping of cultivars**

The repetitive and arbitrary DNA markers are markers of choice in genotyping of cultivars. Microsatellites like  $(CT)_{10}$ ,  $(GAA)_5$ ,  $(AAGG)_4$ ,  $(AAT)_6$  (ref. 123),  $(GATA)_4$ ,  $(CAC)_5$  and minisatellites have been employed in DNA fingerprinting for the detection of genetic variation, cultivar identification and genotyping. This information is useful for quantification of genetic diversity, characterization of accessions in plant germplasm collections and taxonomic studies. Microsatellites have been useful for generation of STMS markers, revealing polymorphisms within closely related cultivars. The first application of microsatellites in plants has been in cultivar identification, wherein microsatellites have been used to genotype unequivocally diverse materials like rice, wheat, grapevine, soybean, etc. This is important especially for protection of proprietary germplasm. Similarly

microsatellite markers have also been advantageous in pedigree analysis as they represent single locus. The multi allelism of these markers facilitates comparative allelic variability detection reliably across a wide range of germplasm and allows individuals to be ubiquitously genotyped, so that gene flow and paternity can be established.

One of the most recent applications of these markers has been shown in sex identification of dioecious plants, wherein microsatellite probe (GATA)<sub>4</sub> is found to reveal sex-specific differences in Southern analysis and can be used as a diagnostic marker in this system where male and female plants do not show any sex-specific morphological difference until flowering. Similarly, Di Stilio *et al.* have identified a randomly-amplified (RAPD) DNA marker for pseudo-autosomal plant sex chromosome in *Silene dioica* (L.).

### **Indian scenario for development of molecular markers in crop improvement programmes**

Agriculture is one of the most important occupations in India with almost 70% of the population being dependent on it. A noteworthy research in conventional breeding for several years has made this country self-sufficient in many respects. However, the ever-increasing population has alarmed food security in India and attempts have been initiated to integrate modern biotechnology tools in conventional breeding to improve the most important crops such as rice, wheat and legumes.

Extensive research using DNA markers is in progress in many institutions all over India. Markers tagged and mapped with specific genes have been identified and some such examples include resistance genes for blast and gall midge using RFLP- and PCR-based approaches in rice. Similarly, in wheat, leaf rust resistance gene *LR* 28, and pre-harvest sprout tolerance gene have been tagged. QTLs such as protein content in wheat and heterosis in rice have also been identified. While efforts for tagging genes providing resistance to BPH, WBPH, sheath rot and drought are going on, many attempts are also being made towards pyramiding different resistance genes for a specific disease or pest attack like blast, bacterial blight, gall midge, BPH, WBPH, etc. in rice in order to increase the field life of the crop.

Germplasm analysis to study genetic diversity is another important area in which a lot of efforts have been put in. Fingerprinting of crops like rice, wheat, chickpea, pigeonpea, pearl

millet etc. is being carried out extensively. This information has potential in strategic planning of future breeding towards crop sustainability in India. Apart from use of molecular markers in crop plants, efforts are also underway in other horticultural plants. Early identification of sex in dioecious papaya using molecular marker is one such example.

Thus in the last few years there are many reports of amalgamation of classical breeding and modern biotechnological approaches which have unlimited scope in Indian agriculture.

## Questions

9. EST markers were used for the first time by .....

- a). Alec Jeffrey
- b). Botstein
- c). Hatada
- d). Adams**

10. The multi locus probes include .....

- a). Microsatellites and minisatellites
- b). STMS
- c). ISSR
- d). All the above**

11. The term microsatellite was coined by .....

- a). Litt
- b). Lutty
- c). Both a and b**
- d). None of the above

12. The term minisatellite was coined by .....

- a). Litt
- b). Lutty
- c). Jeffrey**
- d). None of the above

13. RAPD was developed by .....

- a). Welsh and McClelland**
- b). Litt and Lutty
- c). Jeffrey
- d). None of the above