Available online at www.ijpab.com

ISSN: 2320 – 7051
Int. J. Pure App. Biosci. **2 (1):** 169-188 (2014)

Review Article

International Journal of Pure & Applied Bioscience

Potential and application of molecular markers techniques for plant genome analysis

Mishra Kundan K[1]*, Fougat RS[1], Ballani A[2], Thakur Vinita[2], Jha Yachana[3] and Bora Madhumati[3]

[1]Centre of Excellence in Biotechnology, B. A. College of Agriculture, Anand Agricultural, University, Anand – 388 110, Gujarat, India

[2]Department of Biotechnology, N.V.Patel College of Pure and Applied Sciences V.V.Nagar-388 120, Gujarat and MGM'S Institutes of biosciences and technology, Aurangabad, Maharastra India.

[3]Department of Biotechnology, N.V.Patel College of Pure and Applied Sciences V.V.Nagar-388 120, Gujarat.

*Corresponding Author E-mail: kkmbt@yahoo.co.in, kkmishra71075@gmail.com

ABSTRACT

Molecular markers are based on naturally occurring polymorphisms in DNA sequences (i.e.: base pair deletions, substitutions, additions or patterns). Molecular markers represent one of the most powerful tools for the analysis of genomes and enable the association of heritable traits with underlying genomic variation. The current advancement in plant biology research encompassing generation of large number of molecular-genetic data, development of impressive methodological skills in molecular biology experimentation, and systems analyses, has set the stage to search for process to utilize the available resources to strengthen interdisciplinary efforts to find solutions to the challenging goals of plant breeding and agricultural biotechnology efforts ultimately leading to gainful applications in crop improvement. The presence of various types of molecular markers, and differences in their principles, methodologies, and applications require careful consideration. This review article provides detail for different important molecular marker techniques: restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter-simple sequence repeats (ISSRs), Simple Sequence Repeats (SSRs), single nucleotide polymorphisms (SNPs), expressed sequence tags (ESTs), and Micro arrays technology. All molecular markers technique can be used for several different applications including: germplasm characterization, genetic diagnostics, characterization of transformants, study of genome organization, marker assisted selection (MAS) and phylogenic analysis.

Keywords: *Molecular markers; plant biotechnology; genetic diversity; polymorphism; PCR; AFLP; DNA markers; Hybridization; ISSR; RAPD; RFLP; SSRs; SNPs; ESTs; genome sequencing.*

INTRODUCTION

A molecular markers a DNA sequence that is readily detected and whose inheritance can be easily be monitored. The uses of molecular markers are based on the naturally occurring DNA polymorphism, which forms basis for designing strategies to exploit for applied purposes. There are different types of markers viz. morphological, biochemical and DNA based molecular markers. These DNA based markers are differentiates in two types first hybridization based (RFLP) and second is PCR based markers (RAPD, AFLP, SSR, SNP, EST 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.

The majority of these molecular markers has been developed either from genomic DNA libraries (e.g. RFLPs and SSRs) or from random PCR amplification of genomic DNA (e.g. RAPDs) or both (e.g.

AFLPs). These DNA markers can be generated in large numbers and can prove to be very useful for a variety of purposes relevant to crop improvement. For instance, these markers have been utilized extensively for the preparation of saturated molecular maps (genetical and physical). Their association with genes/QTLs controlling the traits of economic importance has also been utilized in some cases for indirect marker-assisted selection (MAS) (e.g. Koebner 2004, Korzun 2002). Other uses of molecular markers include gene introgression through backcrossing, germplasm characterization, genetic diagnostics, characterization of transformants, study of genome organization and phylogenetic analysis (Jain et al. 2002). For plant breeding applications, SSR markers, among different classes of the existing markers, have been proven and recommended as markers of choice (Gupta and Varshney 2000). RFLP is not readily adapted to high sample throughput and RAPD assays are not sufficiently reproducible or transferable between laboratories. While both SSRs and AFLPs are efficient in identifying polymorphisms, SSRs are more readily automated (Shariflou et al. 2001). Although AFLPs can in principle be converted into simple PCR assays (e.g. STSs), this conversion can become cumbersome and complicated as individual bands are of multiple fragments (Shan et al. 1999), particularly in large genome templates. Despite this, recent and a new marker type, named SNP, for Single Nucleotide Polymorphism, is now on the scene and has gained high popularity, even though it is only a bi-allelic type of marker. Day by day development of such new and specific types of markers makes their importance in understanding the genomic variability and the diversity between the same as well as different species of the plants. Genetic mapping through molecular markers is necessary not only for the reliable detection, mapping and estimation of gene effects of important agronomic traits, but also for further research on the structure, organization, evolution and function of the plant genome.

Properties of ideal molecular markers

An ideal molecular marker must have some desirable properties which are enlisted as follows-

- Highly polymorphic nature: It must be polymorphic as it is polymorphism that is measured for genetic diversity studies.
- Codominant inheritance: determination of homozygous and heterozygous states of diploid organisms.
- Frequent occurrence in genome: A marker should be evenly and frequently distributed throughout the genome.
- Selective neutral behaviours: The DNA sequences of any organism are neutral to environmental conditions or management practices.
- Easy access (availability): It should be easy, fast and cheap to detect.
- Easy and fast assay
- High reproducibility
- Easy exchange of data between laboratories

DNA isolation for molecular marker analysis

DNA samples or Genomic DNA of plants can be isolated at the microlevel from young leaves using a CTAB-based extraction method of Altaf et al., (1997) with slight modifications. Approximately 0.5 g of fresh young leaf tissue can homogenized in 0.7 ml of extraction buffer [100 mM tris-HCl (pH 8.0), 20 mM EDTA (pH 8.0), 1 M NaCl, 2% CTAB, 2% PVP-40, 1 mM 1-10, phenanthroline] and 0.2% β-mercaptoethanol in a 1.5 ml centrifuge tube with the aid of a microtube pellet pestle and an electric hand drill. After incubation for 1 hr at 60ºC, the suspension can purified twice in chloroform:isoamyl alcohol (24:1) by centrifugation at 10,000 rpm on a desktop micro-centrifuge for 10 min at room temperature and precipitated with an equal volume of cold isopropanol. The recovered DNA can spooled out, or pelleted by centrifuging at 10,000 rpm for 5 min, washed twice with 80% Ethyl alcohal and 15 mM ammonium acetate and once with 95% Ethyl alcohal, air dried, and dissolved in 100 to 200 µl of 10mMTris buffer (pH 7.5). For Plant DNA sample 2 µl of Rnase A (10 mg/ml) per 100 µl of dissolved DNA can be added.

1. Restriction Fragment Length Polymorphism (RFLP)

This was the first molecular marked technique developed and used in MAS for plant breeding. Saturated molecular genetic maps based on RFLP markers have been developed for several crops. The technique

centers around the digestion of genomic DNA digested with restriction enzymes. These enzymes are isolated from bacteria and consistently cut DNA at specific base pair sequences which are called recognition sites. These recognition sites are not associated with any type of gene and are distributed randomly throughout the genome. When genomic DNA is digested with one of these restriction enzymes, (of which there are thousands, each cutting at a specific sequence), a series of fragment are produced of varying length. These fragments are separated using agarose or polyacrylamide gel electrophoresis (PAGE) and yield a characteristic pattern. DNA has a uniform charge per unit length when run under electrophoresis conditions which arises from the phosphates groups in its backbone. So when DNA fragments are separated via electrophoresis, the distance they travel is dependent only on their molecular weight. This allows their molecular weight to be determined with simple standard called DNA ladders which are run along side the DNA in the gel. When restriction fragments are separated on agarose gels a series of bands results. Each band corresponds to a restriction fragment of different length. The lighter they are the farther they have traveled. Variations in the characteristic pattern of a RFLP digest can be caused by base pair deletions, mutations, inversions, translocations and transpositions which result in the loss or gain of a recognition site resulting in a fragment of different length and polymorphism. Only a single base pair difference in the recognition site will cause the restriction enzyme not to cut. If the base pair mutation is present in one chromosome but not the other, both fragment bands will be present on the gel, and the sample is said to heterozygous for the marker. Only co-dominant markers exhibit this behavior which is highly desirable, dominant markers exhibit a present/absent behavior which can limit data available for analysis.

Procedure for RFLP

(i) DNA isolation – a significant amount of DNA must be isolated from the sample and purified to a fairly stringent degree as contaminants can often interfere with the restriction enzyme and inhibit its ability to digest the DNA.

(ii) Restriction Digest - Restriction enzyme is added to purified genomic DNA under buffered conditions. The enzyme cuts at recognition sites throughout the genome and leaves behind hundreds of thousands of fragments.

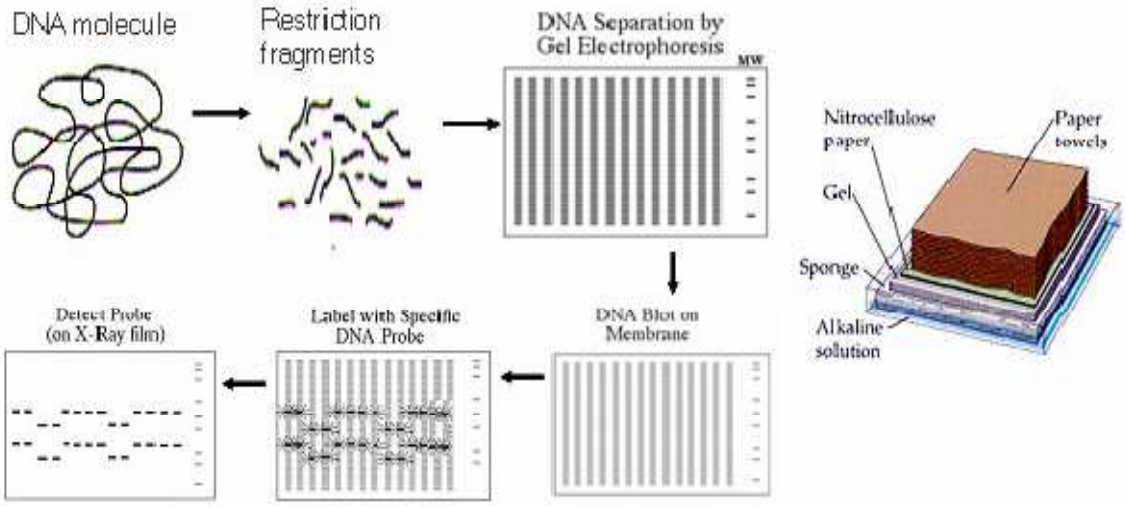
(iii) Gel electrophoresis – The digest is run on a gel and when visualized appears a smear because of the large number of fragments.

(iv) Southern blotting-transfer to nitrocellulose or nylon membrane filter

(v) Probe visualization – Because of the large number of fragments, probes must be constructed to visualize more specific bands in the digest. These probes consist of radio labeled oligonucleotide sequences which will anneal to the fragment sequences so that that they may be visualized on photographic paper using a technique called autoradiography.

(vi) Analysis-Number of RFLP loci can be analyzed after autoradiography.

Fig 1- Outline of the different steps of restriction fragment length polymorphism (RFLP) markers



Restriction fragment length polymorphisms (RFLPs) are very reliable markers in linkage analysis and crop breeding however, time consuming, expensive and require large amount of DNA for restriction and hybridization analysis (Paterson *et al*., 1993).

Examples of RFLPs

1. Methylene Tetra-Hydro-Folate Reductase (MTHFR) mutation detection (Creating Restriction Site):

Fig 1

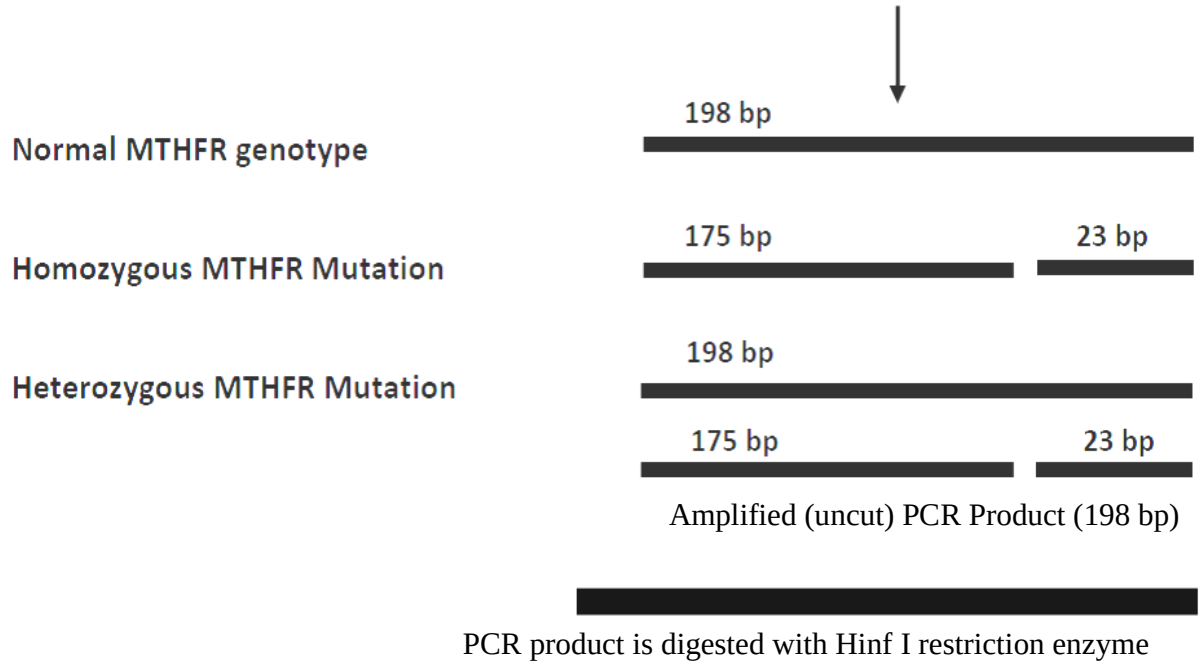
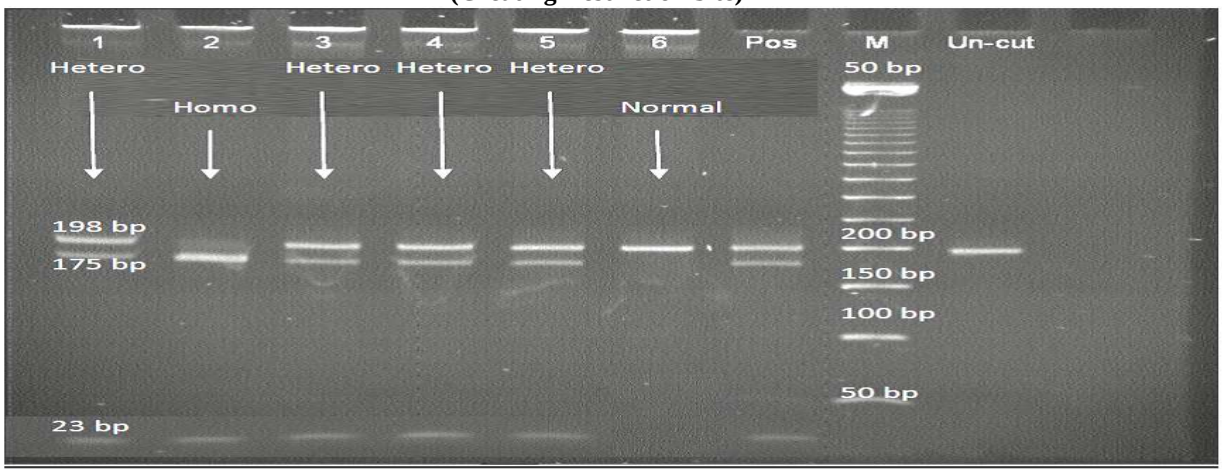


Fig 2- Detection of RFLPs- Methylene Tetra-Hydro-Folate Reductase (MTHFR) mutation detection (Creating Restriction Site)



2. Factor V (FV) mutation detection (Deleting Restriction Site):

Amplified (uncut) PCR Product (143 bp)

PCR product is digested with Mnl II restriction enzyme

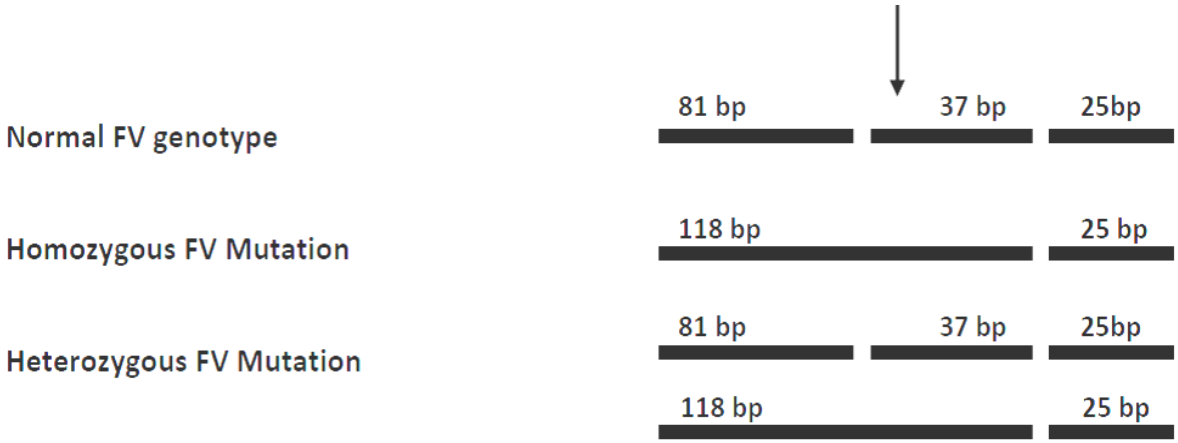
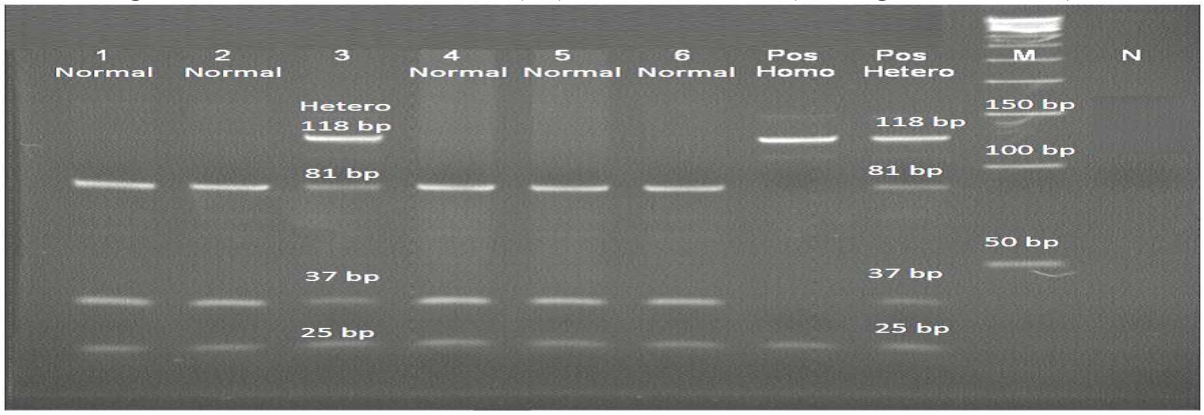


Fig 3- Detection of RFLPs- Factor V (FV) mutation detection (Deleting Restriction Site)



2. Random Amplified Polymorphic DNA (RAPD)

RAPD was the first PCR based molecular marker technique developed and it is by far the simplest. Short PCR primers (approximately 10 bases) are randomly and arbitrarily selected to amplify random DNA segments throughout the genome. The resulting amplification product is generated at the region flanking a part of the 10 bp priming sites in the appropriate orientation. RAPD often shows a dominant relationship due to primer being unable to bind (show 3:1 ration, unable to distinguish between homozygotes and heterozygotes) RAPD products are usually visualized on agarose gels stained with ethidium bromide.

RAPD markers offers many advantages such as higher frequency of polymorphism, rapidity technical simplicity, use of fluorescence, requirement of only a few nanogram of DNA, no requirement of prior information of the DNA sequence and feasibility of automation (Subudhi and Huang, 1999). The use of such techniques for germplasm characterization facilitates the conservation and utilization of plant genetic resources, permitting the identification of unique accessions or sources of genetically diverse germplasm. (Kapteyn and Simon, 2002). The technique is widely used to analyze the genetic relatedness in several crop species (Chalmers et al., 1992; Koller et al., 1993; Zhu et al., 1997).

Several factors have been reported to influence the reproducibility of RAPD reactions: quality and quantity of template DNA, PCR buffer, concentration of magnesium chloride, primer to template ratio, annealing temperature, Taq DNA polymerase brand or source, and thermal cycler brand (Wolff et al., 1993). The concern about reproducibility of RAPD markers, however, could be overcome through choice of an appropriate DNA extraction protocol to remove any contaminants (Micheli et al., 1994), by optimizing the parameters used (Ellsworth et al., 1993; Skroch and Nienhuis, 1995), by testing several oligonucleotide primers and scoring only the reproducible DNA fragments (Kresovich et al., 1992; Yang and Quiros, 1993), and by using appropriate DNA polymerase brand. The presence of artifactual bands

(false positives) corresponding to rearranged fragments produced by nested primer binding sites (Schierwater et al., 1996; Rabouam et ,al. 1999) and intrastrand annealing and interactions during PCR (Hunt and Page, 1992; Caetano- Anolles et al., 1992) have also been reported to influence the reliability of RAPD data. The presence of both false negatives and false positives may, if frequent, seriously restrict the reliability of RAPDs for various purposes, including genetic diversity and mapping studies. All pair wise comparison of RAPD fragments along samples begins with the assumption that co-migrating bands (i.e.,bands that migrate equal distance) represent homologous loci. However, as in any study based on electrophoretic resolution, the assumption that equal length equals homology may not be necessarily true, especially in polyploid species. For example, some RAPD bands scored as identical (equal length) have been found not to be homologous (e.g., Thormann et al., 1994; Pillay and Kenny, 1995); more accurate resolution of fragment size using polyacrylamide gels and AgNO3 staining have been reported to reduce such errors (e.g., Huff et al., 1993). The other limitation of RAPD markers is that the majority of the alleles segregate as dominant markers, and hence the technique does not allow identifying dominant homozygotes from heterozygotes. The RAPD assays produce fragments from homozygous dominant or heterozygous alleles. No fragment is produced from homozygous recessive alleles because amplification is disrupted in both alleles.

Fig- 4 RAPD profile of different cotton genotypes

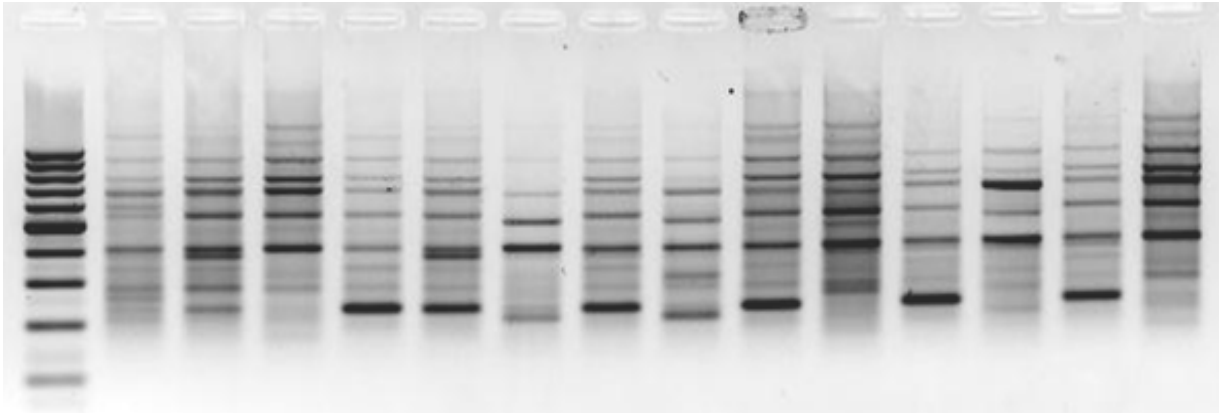
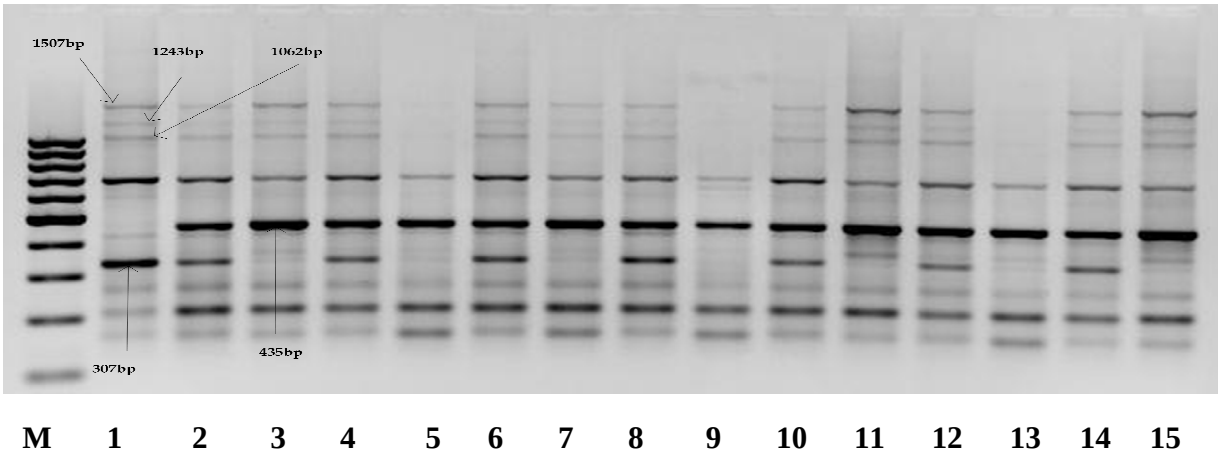


Fig- 5 Genetic purity testing of hybrids through molecular markers RAPD



Limitation of RAPD

(1) It is not always reproducible.

(2) It shows dominant inheritance.

(3) Sometime it reveals homology.

3. Amplified Fragment Length Polymorphism (AFLP)

AFLP is the latest form of marker assisted selection and is a highly sensitive method based on the combined concepts of RFLP and RAPD. This technique is applicable to all species giving very reproducible results. The basis of AFLP is the PCR amplification of restriction enzyme fragments of genomic DNA. The key feature of AFLP is its capacity for "genome representation": the simultaneous screening of representative DNA regions distributed randomly throughout the genome. AFLP markers can be generated for DNA of any organism without initial investment in primer/probe development and sequence analysis. Both good quality and partially degraded DNA can be used for digestion but the DNA should be free of restriction enzyme and PCR inhibitors. Details of the AFLP methodology have been reviewed by various authors (e.g., Blears et al., 1998; Mueller and Wolfenbarger, 1999; Ridout and Donini, 1999).Overall steps of AFLPs can be represented by following steps-

(i) DNA is cut with two specific restriction enzymes, one frequent cutter (3 bp recognition site) and one rare cutter (6 bp recognition site).

(ii) Oligonucleotide "adapters" are ligated to the ends of each fragment. One end with a complimentary sequence for the rare cutter and the other with the complimentary sequence for the frequent cutter. This way only fragments which have been cut by the frequent cutter and rare cutter will be amplified.

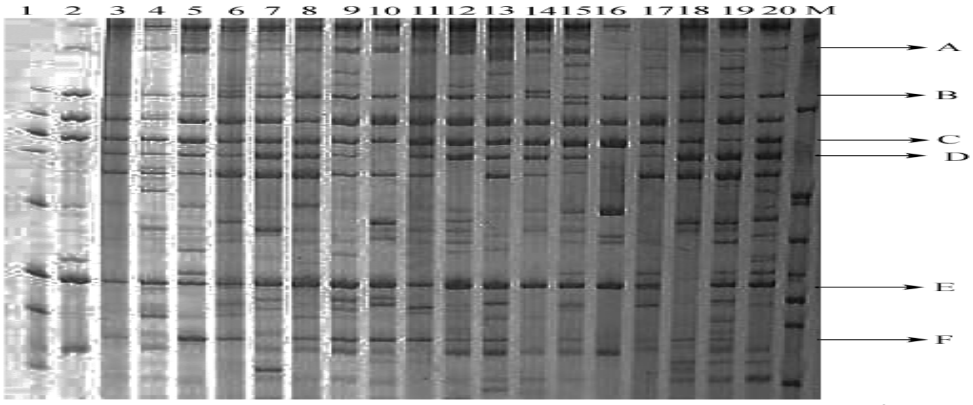
(iii) Primers are designed from the known sequence of the adapter, plus 1-3 selective nucleotides which extend into the fragment sequence. Sequences not matching these selective nucleotides in the primer will not be amplified.

(iv) PCR performed

(v) Visualized on agarose gels with ethidium bromide

Typical results give 50-100 bands despite selective nucleotides and rare/frequent selection. This high number of bands eases analysis by providing more chance of polymorphism. AFLP technique shares some characteristics with both RFLP and RAPD analysis. It combines the specificity of restriction analyses with PCR amplification. The sequence variation detected is the same as that with RFLP analyses but the number of polymorphism detected per analysis is higher. Both RFLP and AFLP require southern blotting, radioactive labeling and auto-radiography and as such are expensive techniques that many breeding programmes in the developing countries cannot afford particularly due to non availability of radioactive material. AFLP can be used to distinguish closely related individuals at the sub-species level and can also map genes. Like that for RFLP, high quality DNA is also required. AFLP is extremely sensitive technique and the added use of fluorescent primers for automated fragment analysis systems, and sophisticated software packages to analyze the biallelic data, makes the AFLP well suited for high throughput analysis. Mode of transmission of AFLP is reported to be of bi-parental nuclear through many loci with unknown number of alleles per locus. Their mode of action is dominant at some loci and codominant at others. Thus, they act both as gene and genetic markers. Level of variability is abundant as each and every pattern is unique.

Fig 6-AFLP profile of six genotypes of cotton



Lane 1, Female parent; lane 2, male parent; lanes 3—20, the progeny; M, marker. A, Presence of polymorphism in the male parent; B, polymorphism of intensity difference between parents; C, monomorphic in parents and progeny; D, presence of polymorphism in the female parent; E, polymorphic in progeny, but monomorphic in parents; F, absent in parents, but polymorphic in progeny.

The advantages of AFLP include

(1) It is highly reliable and reproducible

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

(3) It is information-rich due to its ability to analyze a large number of polymorphic loci simultaneously (effective multiplex ratio) with a single primer combination on a single gel as compared to RAPDs, RFLPs and microsate-llites

(4) Co-migrating AFLP amplification products are mostly homologous and locus specific with exceptions in polyploidy species.

limitations of AFLP include:

(1) It requires more number of steps to produce the result.

(2) It requires template DNA free of inhibitor compounds that interferes with the restriction enzyme.

(3) The technique requires the use of polyacrylamide gel in combination with AgNO3 staining, radioactivity, or fluore-scent methods of detection, which will be more expensive and laborious than agarose gels.

(4) It involves additional cost to purchase both restriction and ligation enzymes as well as adapters.

(5) Most AFLP loci are dominant, which does not differentiate dominant homozygotes from heterozygotes. This reduces the accuracy of AFLP markers in population genetic analysis, genetic mapping, and marker assisted selection (MAS).

4. Inter Simple Sequence Repeat (ISSR)

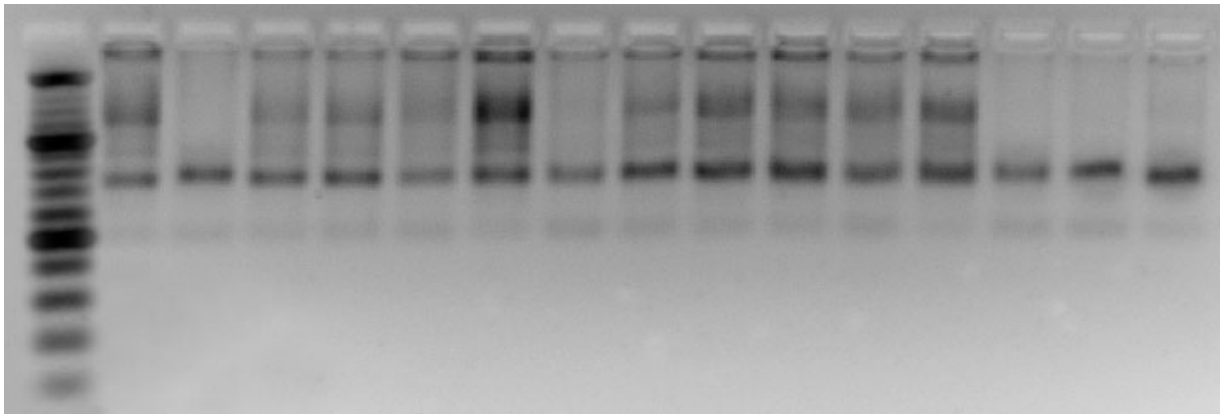
ISSR involves amplification of DNA segments present at an amplifiable distance in between two identical microsatellite repeat regions oriented in opposite direction. The technique uses microsatellites as primers in a single primer PCR reaction targeting multiple genomic loci to amplify mainly inter simple sequence repeats of different sizes. The microsatellite repeats used as primers for ISSRs can be di-nucleotide, tri-nucleotide, tetranucleotide or penta-nucleotide. The primers used can be either unanchored (Meyer et al., 1993; Gupta et al., 1994; Wu et al., 1994) or more usually anchored at 3` or 5` end with 1 to 4 degenerate bases extended into the flanking sequences. ISSRs use longer primers (15–30 mers) as compared to RAPD primers (10 mers), which permit the subsequent use of high annealing temperature leading to higher stringency. The annealing temperature depends on the GC content of the primer used and ranges from 45 to 65ºC. The amplified products are usually 200–2000 bp long and amenable to detection by both agarose and polyacrylamide gel electrophoresis.

In contrast to the SSR marker technique that amplifies with primers located on the flanking single copy DNA, microsatellites anchored primers that anneal to an SSR region can amplify regions between adjacent SSRs. The ISSR technique uses primers that are complimentary to a single SSR and anchored at either the 5' or 3' end with a one- to three-base extension (Zietkiewicz et al., 1994). The amplicons generated consist of regions between neighbouring and inverted SSRs. As a result, the high complex banding pattern obtained will often differ greatly between genotypes of the same species. Liu and Wendel (2001) reported ISSR as an easy and informative genetic marker system in cotton for revealing both inter and intraspecific variations.

ISSRs exhibit the specificity of microsatellite markers, but need no sequence information for primer synthesis enjoying the advantage of random markers (Joshi et al., 2000). The primers are not proprietary and can be synthesized by anyone. The technique is simple, quick, and the use of radioactivity is not essential. ISSR markers usually show high polymorphism (Kojima et al., 1998) although the level of polymorphism has been shown to vary with the detection method used. Polyacrylamide gel electrophoresis (PAGE) in combination with radioactivity was shown to be most sensitive, followed by PAGE with AgNO3 staining and then agarose gel with EtBr system of detection. Like RAPDs, reproducibility, dominant inheritance and homology of co migrating amplification products are the main limitations of ISSRs. Fang and Roose (1997) reported a reproducibility level of more than 99% after performing repeatability tests for ISSR markers by using DNA samples of the same cultivar grown in different locations, DNA extracted from different aged leaves of the same individual, and by performing

separate PCR runs. In other cases, the reproducibility of ISSRs amplification products ranged from 86 to 94%, with the maximum being when polyacrylamide gel electrophoresis and AgNO3 staining were used and weak bands excluded from scoring (Moreno et al., 1998). ISSRs segregate mostly as dominant markers (Gupta et al., 1994; Tsumura et al., 1996; Ratnaparkhe et al., 1998; Wang et al., 1998), although co-dominant segregation has been reported in some cases (Wu et al., 1994; Akagi et al., 1996; Wang et al., 1998; Sankar and Moore, 2001). There is also a possibility as in RAPD that fragments with the same mobility originate from non-homologous regions (Sanchez et al., 1996).

Fig 7- ISSR profile of different cotton genotypes



M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

5. Simple Sequence Repeat (SSR)

Simple sequence repeats are present in the genomes of all eukaryotes and consists of several to over a hundred repeats of a 1-4 nucleotide motif. Some common motifs are: Mono: A, T Di: AT, GA Tri: AGG Tetra: AAAC. These repeated motifs are denoted (AAAC)n, where n is the number of tandem repeats. The sequences flanking these microsatellites are often conserved and can be used to design primers. These primers can be designed by constructing a novel genomic library and sequencing segments of the subject genome. Already discovered sequence (i.e.: GENEBANK online database) can also be searched for SSRs and primers designed from that. Polymorphism is based on the number of tandem repeats and therefore the length of the PCR products. SSR is a co dominant marker such as RFLP and is usually visualized on metaphor agarose or polyacrylamide gels. Presence of short tandem repeats of varying length is characteristic of microsatellite loci (Akkaya et al., 1992). The simple sequence repeats (SSR) or microsatellites (sometimes referred to as a variable number of tandem repeats or VNTRs) that are similar in nature have been shown to be abundant and highly polymorphic in eukaryotic genomes. They maybe dinucleotide repeats (AC)n, (AG)n and (AT)n or trinucleotide or tetranucleotide repeats. SSRs are analyzed by PCR-amplification of a short genomic region containing the repeated sequence and size estimation of the repeat length by gel separation. Very little DNA is required for SSR analysis , if the primer sequences are known a laboratory set up to perform RAPD analysis could be used for the purpose. Microsatellites are highly useful markers for cultivar identification as they have been shown to be highly polymorphic and genotype specific. The technique is more robust and reproducible. Hence this is fast replacing RAPDs as a tool for cultivar identification. Powell et al. (1996) reported that among different classes of molecular markers, SSR markers are useful for a variety of applications in plant molecular biology, genetics and breeding because of their reproducibility, multiallelic nature, codominant inheritance, relative\ abundance and good genome coverage. Gupta and Varshney (2000) also reported that SSR markers have been useful for integrating the genetic, physical and sequence-based physical maps in plant species, and simultaneously have provided breeders and geneticists with an efficient tool to link phenotypic and genotypic variation.

The genomes of higher organisms contain tree types of multiple copies of simple repetitive DNA sequences (satellite DNAs, minisatellites, and microsatellites) arranged in arrays of vastly differing size (Armour et al., 1999; Hancock, 1999). Microsatellites (Litt and Luty, 1989), also known as simple sequence repeats (SSRs; Tautz et al., 1986), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs; McDo-nald and Potts, 1997), are the smallest class of simple repetitive DNA sequences. Some authors (e.g. Armour et al., 1999) define microsatellites as 2–8 bp repeats, others (e.g., Goldstein and Pollock, 1997) as 1–6 or even 1–5 bp repeats (Schlotterer, 1998). Chambers and MacAvoy (2000) suggested following a strict definition of 2–6 bp repeats, in line with the descriptions of the original authors. Microsatellites are born from regions in which variants of simple repetitive DNA sequence motifs are already over represented (Tautz et al., 1986). It is now well established that the predominant mutation mechanism in microsatellite tracts is 'slipped-strand mispairing' (Levinson and Gutman, 1987). This process has been well described by Eisen (1999). When slipped-strand mispairing occurs within a microsatellite array during DNA synthesis, it can result in the gain or loss of one, or more, repeat units depending on whether the newly synthesized DNA chain loops out or the template chain loops out, respectively. The relative propensity for either chain to loop out seems to depend in part on the sequences making up the array, and in part on whether the event occurs on the leading (continuous DNA synthesis) or lagging (discontinuous DNA synthesis) strand (Freudenreich et al., 1997). SSR allelic differences are, therefore, the results of variable numbers of repeat units within the microsatellite structure. 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 that ranges between 10 and 100. These markers often present high levels of inter- and intra-specific polymorphism, particularly when tandem repeats number is ten or greater (Queller et al., 1993). PCR reactions for SSRs is run in the presence of forward and reverse primers that anneal at the 5` and 3` ends of the template DNA, respectively. PCR fragments are usually separated on polyacrylamide gels in combination with AgNO3 staining autoradiography or fluorescent detection systems. Agarose gels (usually 2.5- 3%) with EtBr can also be used when differences in allele size among samples is larger than 10 bp. However, the establishment of microsatellite primers from scratch for a new species presents a considerable technical challenge. Several protocols have been developed (Bruford et al., 1996; McDonald and Potts, 1997; Hammond et al., 1998; Schlotterer, 1998) and details of the methodologies are reviewed by different authors (e.g., Chambers and MacAvoy, 2000; Zane et al., 2002; Squirrell et al., 2003). A review by Zane et al. (2002) describes some of the technical advances that have been made in recent years to facilitate microsatellite development. They cover a range of methods for obtaining sequences rich in microsatellite repeats (some of which can be undertaken in a matter of days), and also highlight the availability of companies who will undertake the construction of enriched microsatellite libraries as a commercial service. The development of microsatellite markers involves several distinct steps from obtaining the library to developing a working set of primers that can amplify polymorphic microsatellite loci. These include:

(1) Microsatellite library construction.
(2) Identification of unique microsatellite loci.
(3) Identifying a suitable area for primer design.
(4) Obtaining a PCR product.
(5) Evaluation and interpretation of banding patterns.
(6) Assessing PCR products for polymorphism (Roder et al., 1998).

SSR primers are developed by cloning random segments of DNA from the target species. These are inserted into a cloning vector, which is in turn, implanted into *Escherichia coli* bacteria for replication. Colonies are then developed, and screened with single or mixed simple sequence oligonucleotide probes that will hybridize to a microsatellite repeat, if present on the DNA segment. If positive clones for microsatellite are obtained from this procedure, the DNA is sequenced and PCR primers are chosen from

sequences flanking such regions to determine a specific locus. This process involves significant trial and error on the part of researchers, as microsatellite repeat sequences must be predicted and primers that are randomly isolated may not display polymorphism (Queller et al., 1993; Jarne and Lagoda, 1996). The next step is to select the best candidate markers and then to optimize conditions for their amplification. Optimization of microsatellite systems involves a more or less comprehensive survey of PCR conditions for amplification of candidate loci. The objective here is to adequately balance the often conflicting requirements for high specificity and high intensity of amplification products. Thus, the issue of signal strength and purity remains the primary focus. Other considerations include obtaining products from various loci with non-overlapping ranges of allele sizes, which can be amplified with similar efficiency under a standard set of conditions and enables multiplexing for high throughput analysis (Schlotterer, 1998).Microsatellite loci are more common in some organisms than in others, and screening may produce few useful loci in some species (Cooper, 1995). The efficiency of microsatellite marker development depends on the abundance of repeats in the target species and the ease with which these repeats can be developed into informative markers. When researchers are isolating plant microsatellites, about 30% of the sequenced clones, on average, can be lost due to the absence of unique microsatellites. Of those sequences that contain unique microsatellites, a number of the clones in a library can contain identical sequences (and hence there is a level of redundancy) and/or chimeric sequences (i.e., one of the flanking regions matches that of another clone). At each stage of SSR development, therefore, there is the potential to 'lose' loci, and hence the number of loci that will finally constitute the working primer set will be a fraction of the original number of clones sequenced (Squirrell et al., 2003). The conversion of microsatellite-containing sequences into useful markers can be quite difficult, especially in species with large genomes (Smith and Devey, 1994; Kostia et al., 1995; Roder et al., 1995; Pfeiffer et al., 1997; Song et al., 2002). The low conversion rates of primer pairs to useful markers in these species are due to the high level of repetitive DNA sequences in their genomes. The recovery rate for useful SSR primers is generally low due to different reasons:

(a) The primer may not amplify any PCR product.

(b) The primer may produce very complex, weak or nonspecific amplification patterns.

(c) The amplification product may not be polymorphic.

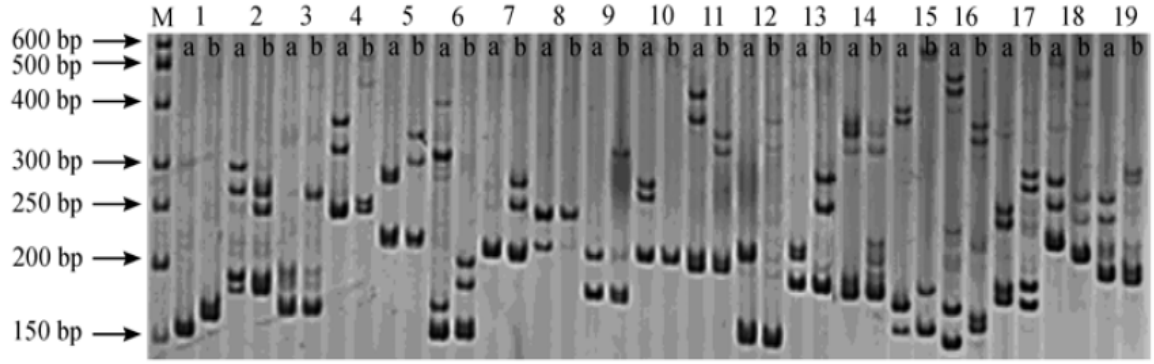
Investigators often prefer to work with loci containing triand tetra-nucleotide repeat arrays rather than dinucleotide arrays because the former frequently give fewer "stutter bands" (multiple near-identical 'ladders' of PCR products which are one or two nucleotides shorter or longer than the full length product. Thus, allele sizing is less error prone using tri- and tetra-nucleotide repeats than di-nucleotide repeats (Diwan and Cregan, 1997). However, this idea must be balanced against practical considerations. Di-nucleotide repeat arrays occur much more frequently than tri- or tetra-nucleotide repeat arrays, and it is easier to run combinatorial screens for them. SSRs are now the marker of choice in most areas of molecular genetics as they are highly polymorphic even between closely related lines require low amount of DNA, can be easily automated for high throughput screening, can be exchanged between laboratories, and are highly transferable between populations(Gupta et al., 1999). For example, a total of 18,828 SSR sequences have been detected in the rice genome (The Rice Genome Mapping project, 2005), of which only 10 - 15% have yet been used, suggesting the high potential available for such marker systems. SSRs are mostly codominant markers, and are indeed excellent for studies of population genetics and mapping (Jarne and Lagoda, 1996; Goldstein and Schlotterer, 1999). The use of fluorescent primers in combination with automatic capillary or gel-based DNA sequencers has got its way in most advanced laboratories and SSR are excellent markers for fluorescent techniques, multiplexing and high throughput analysis.

The major constraint of using SSR markers from genomic libraries is the high development cost and effort required to obtain working primers for a given study species. This has restricted their use to only a few of the agriculturally important crops. A more widespread use of genomic SSRs in plants would also be facilitated if such loci were transferable across species. Recently, a new alternative source of SSRs development from expressed sequence tag (EST) databases has been utilized (Kota et al., 2001; Kantety

et al., 2002; Michalek et al., 2002). With the availability of large numbers of ESTs and other DNA sequence data, development of EST-based SSR markers through data mining has become a fast, efficient, and relatively inexpensive compared with the development of genomic SSRs (Gupta et al., 2003). This is due to the fact that the time-consuming and expensive processes of generating genomic libraries and sequencing of large numbers of clones for finding the SSR containing DNA regions are not needed in this approach (Eujayl et al., 2004). However, the development of EST-SSRs is limited to species for which this type of database exists. Furthermore, the EST-SSR markers have been reported to have lower rate of polymorphism compared to the SSR markers derived from genomic libraries (Cho et al., 2000; Scott et al., 2000; Eujayl et al., 2002; Chabane et al., 2005).

Differences in SSR allele size is often difficult to resolve on agarose gels and high resolutions can be achieved through the use of polyacrylamide gels in combination with AgNO3 staining. The cost of polyacrylamide gels is higher than agarose gels and it is not also as rapid as the latter. The establishment and running cost for an automatic DNA sequencer is not affordable for researchers at the national research systems and universities in developing countries. The other technical problem with microsatellites is the fact that it is not always possible to compare data produced by different laboratories, due to the eventuality of inconsistencies in allele size calling. Such inconsistencies are mainly due to the large variety of automatic sequencing machines used, each providing different gel migration, fluorescent dyes, allele calling software's, and PCR reaction. For the later, the enzyme used for DNA synthesis (Taq DNA polymerase) catalyses the addition of an extra base (usually an adenine) at the end of the PCR product. The proportion of fragments with this extra base may vary from none to 100%, inducing one base pair size differences and complicating data analysis. Although biochemical treatments after PCR or modification of PCR primers can circumvent this problem (Brownstein et al., 1996; Ginot et al., 1996), they are seldom used.

Fig 8- Primer pairs for polymorphisms between two tetraploid cotton



(a, *G. barbadense* and b, *G. hirsutum* M-Molecular weight marker, lane 1-19 different SSR primer amplification product.)

Fig 9- Electrophoretic profiles of the 11 cotton genotypes as revealed by SSR : primers C5 (A) and M8 (B).

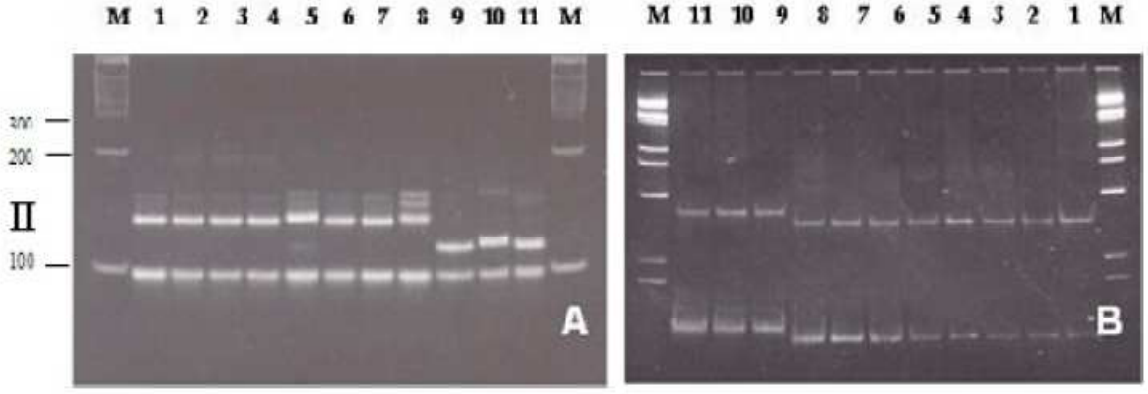
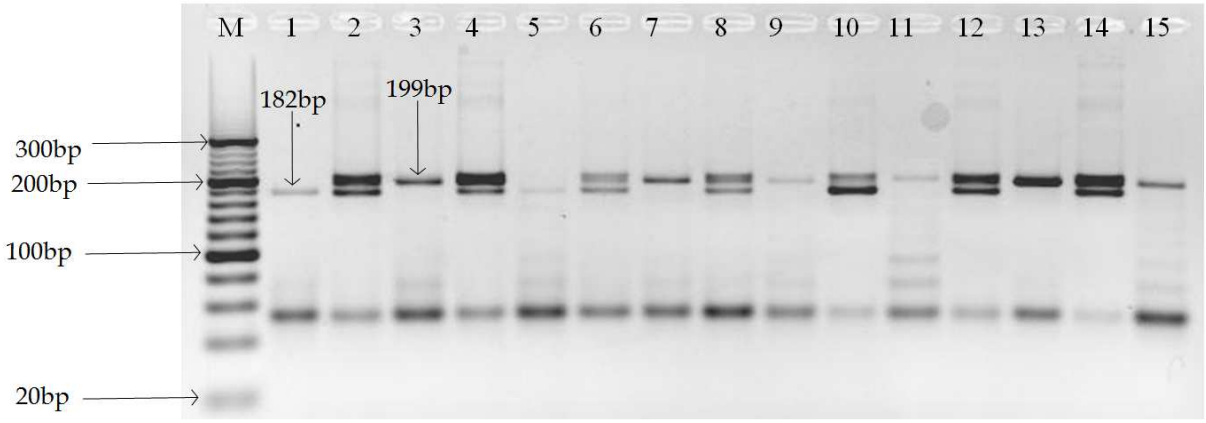


Fig 10 Genetic Purity Testing of hybrids through SSR markers



Homozygous dominant or heterozygous alleles. No fragment is produced from homozygous recessive alleles because amplification is disrupted in both alleles.

6.Single nucleotide polymorphisms (SNPs)

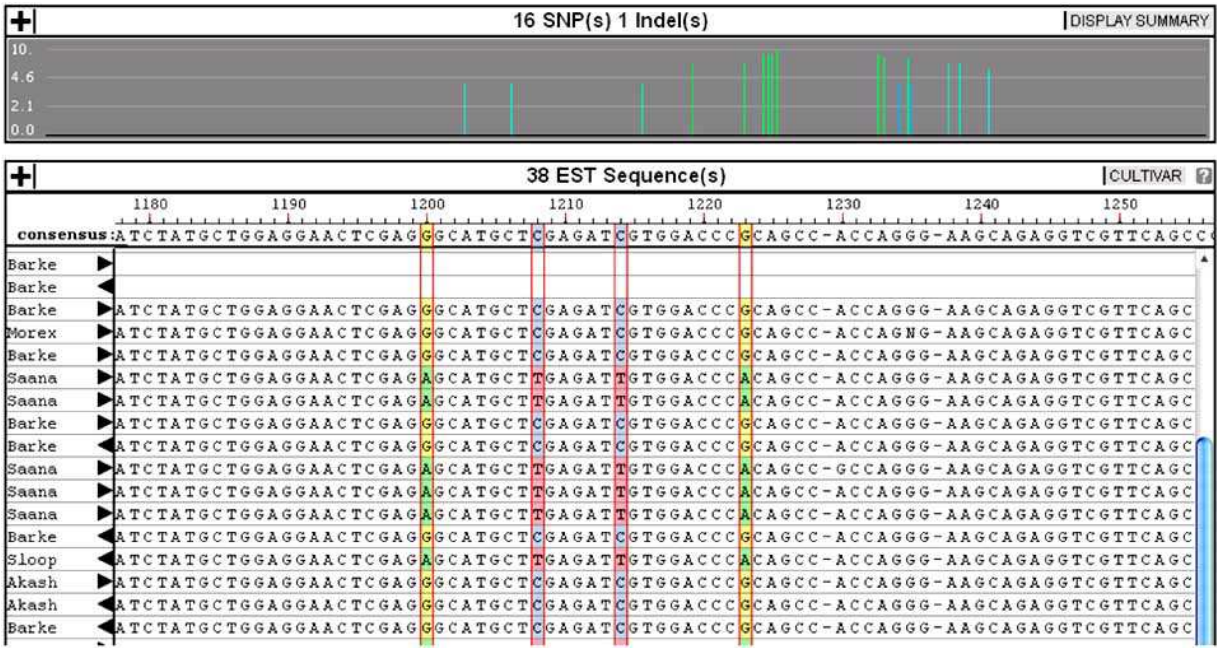
SNPs are a single base change or small insertions and deletions in homologous DNA fragments. In human genome sequencing 10 to 30 million SNPs were found and were the most abundant source of polymorphisms (Collins *et al*., 1998) present both in coding and noncoding regions (Aerts *et al.,* 2002). As a marker SNPs are preferred over other marker systems because they are more frequent, codominant in nature and are sometimes associated with morphological changes (Lindblad-Toh *et al*., 2000). Genomes of higher plants like barley (Kanazin *et al.,* 2002), maize (Tenaillon *et al*., 2001), soybean (Choi *et al*., 2007), sugar beet (Schneider *et al*., 2001), sunflower (Lai *et al*., 2005), rye (Varshney *et al*., 2007) and cotton (Lu *et al.,* 2005; Shaheen *et al*., 2006; Ahmad *et al*., 2007) have also been surveyed for SNPs discovery and characterization. Because SNPs are highly polymorphic, every gene should contain a few SNPs even among strains (Cho *et al*., 1999). MT-sHSP gene is an important gene which helps to tolerate heat shock. The MT-sHSP protects NADH: ubiquinone oxidoreductase of the electron transport chain during heat stress in plants (Herrman *et al*., 1994). SNP markers, combined with QTL data for phenotypic character, can provide a new system of breeding i.e., gene-mediated breeding instead of marker-assisted selection (Lange & Whittacker, 2001). Genetic improvement of crops and agricultural productivity will be enhanced by the availability of rapidly developing genetic resources and tools, including high-density genetic maps (Lacape *et al*., 2005). Polyploid genomes are more difficult to analyze for SNPs than diploids. The ratio of SNP alleles varies in polyploidy genomes (Adams *et al*., 2003). SNPs are now the dominant marker used in biomedical applications due to the availability of the human genome sequence and knowledge of allelic variation derived from the Hap Map project. The ability to screen large numbers of individuals for a range of SNP variants enables the prediction of susceptibility to a wide range of diseases and opens the door to the use of personalized medicine based on the patients genotype. SNPs are becoming increasingly used in animal breeding, with particular success being derived from the bovine Hap Map project. It is expected that in crop genetics, SNPs will co-exist with other marker systems for several years. However, with the development of new technologies to increase throughput and reduce the cost of SNP development, along with further genome sequencing, the use of SNPs will become more widespread.

In Silico SNP Discovery

The dramatic increase in the number of DNA sequences submitted to databases makes the electronic mining of SNPs possible without the need for sequencing. The identification of sequence polymorphisms in assembled sequence data is relatively simple; the challenge of *in silico* SNP discovery is not SNP identification, but rather the ability to distinguish real polymorphisms from the abundant sequencing errors. Current Sanger sequencing produces errors as frequent as one error every one hundred base pairs,

whilst some of the next generation technologies are even less accurate with errors as frequent as one in every 25 bp. Several sources of sequence error need to be addressed during *in silico* SNP identification. The most abundant error in Sanger sequencing is incorrect base calling, due to the requirement to obtain the greatest sequence length. These errors are usually identified by the relatively low quality scores for these nucleotides. Further errors are due to the intrinsically high error rate of the reverse transcription and PCR processes used for the generation of cDNA libraries and these errors are not reflected by poor sequence quality scores. A number of methods used to identify SNPs in aligned sequence data rely onsequence trace file analysis to filter out sequence errors by their dubious trace quality. The major drawback to this approach is that the sequence trace files required are rarely available for large sequence datasets collated from a variety of sources. In cases where trace files are unavailable, two complementary approaches have been adopted to differentiate between sequence errors and true polymorphisms: (1) assessing redundancy of the polymorphism in an alignment, and (2) assessing co-segregation of SNPs to define a haplotype.

Fig11-Auto SNP database showing the overview of the SNPs in this assembly and the aligned sequences with the SNPs highlighted.



7. DNA Microarrays

This is a powerful, versatile and economical technique for screening of genetic aberrations. The process lies in miniaturization, automation and parallelism permitting large-scale and genome-wide acquisition of quantitative biological information from multiple samples. DNA microarrays are currently fabricated and assayed by two main approaches involving either *in situ* synthesis of oligonucleotides ('oligonucleotide microarrays') or deposition of pre-synthesized DNA fragments ('cDNA microarrays') on solid surfaces. To date, the main applications of microarrays are in comprehensive, simultaneous gene expression monitoring and in DNA variation analyses for the identification and genotyping of mutations and polymorphisms. Its application in plant science, microarrays are being utilized to examine a range of biological issues including the circadian clock, plant defence, environmental stress responses, fruit ripening, phytochrome A signalling, seed development and nitrate assimilation.

Microarray technology is a hybridization-based method combining miniaturization and the application of fluorescent dyes for labelling. The latter facilitates the combination of two differently labeled samples in a single hybridization experiment and thus the use of competitive hybridization to reduce experimental error. In this way relative expression levels of large numbers of genes can be determined simultaneously with a high degree of sensitivity. Today, two fundamentally different microarray-based technologies are

available. Both are suitable for large-scale expression analyses. A photolithographic method for high-density spatial synthesis of oligonucleotides was introduced by Fodor and colleagues.With this method arrays can be produced containing up to a few hundred thousand distinct elements (Fodor *et al.*, 1991). As oligonucleotide arrays allow highly sensitive detection of DNA mismatches, they are well suited for DNA variation analysis as well. Manufacturing such arrays requires, however, prior sequence knowledge as well as complicated design and production methodologies (Lipshutz *et al.*, 1999). The alternative method, in which pre-synthesized nucleic acids are mechanically deposited onto a solid surface, allows a more flexible design for the fabrication of microarrays (Duggan *et al.*, 1999). In most cases PCR-amplified cDNA clones are used and the resulting arrays are referred

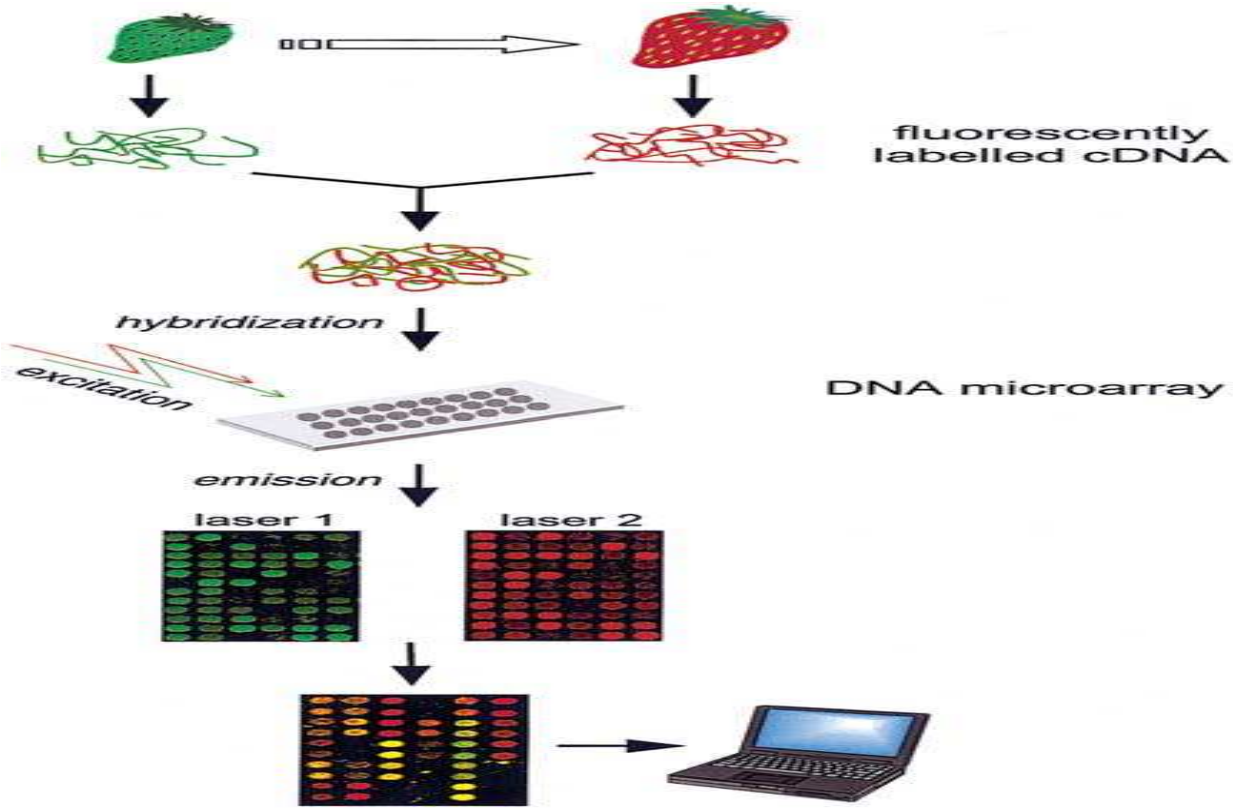
to as cDNA microarrays. However, this technology can also be used to manufacture oligonucleotide arrays.

Basic principle of microarray technology

The specificity of microarray technology relies on the selective and differential hybridization of nucleic acids. Earlier methods, such as DNA and RNA gel blot analysis, use a unique, labeled nucleic acid molecule in solution. This so-called probe is hybridized to the complex mixture under study, such as a total RNA sample, that has been attached to a solid support. Information obtained from such experiments relates to the abundance of one single polynucleotide of interest. Array-based methods such as oligonucleotide arrays and cDNA arrays use the reverse strategy (Figure), where complex mixtures of labelled polynucleotides (such as cDNA derived from mRNA) are hybridized with large numbers of individual elements (e.g. unique PCR products in cDNA microarrays), attached to a solid surface. In this way information on the abundance of many polynucleotide species is gained in parallel. Labelling with fluorescent dyes possessing different excitation and emission characteristics allows the simultaneous hybridisation of two samples on a single array. The strength of fluorescence emission at the two wavelengths represents the amount of a specific polynucleotide from each sample bound to the array. In this manner a single experiment provides quantitative hybridization data for hundreds to thousands of probes. For expression studies using cDNA microarrays this approach of combining two differently labeled samples (reference and test sample) is common practice. For each gene the corresponding amount of signal in both samples can then be quantified in parallel and expression ratios obtained. This strategy, to use expression ratios instead of absolute expression levels, for the analysis of changes in gene expression, has been shown to be a very powerful one and has helped overcome a large source of experimental variation. Assuming the influence of the different dyes on the hybridization characteristics of the labelled molecules to be identical, the initial ratios between specific, differently labelled mRNA molecules should be maintained upon hybridization to the array. As a result, ratios between the two samples for each gene will then be independent of the amount of mRNA hybridized (Vorst *et al.*, 2001).

Two fundamentally different approaches are currently utilized in microarray fabrication. The printing-type technologies are based on the deposition of minute (sub-nanolitre) quantities of a DNA solution onto a solid surface (carrier). These fall into two distinct categories: contact printing (various methods for mechanical deposition) and non-contact printing (liquid delivery). Photolithographic techniques, on the other hand, can be used to synthesize oligonucleotides directly on the carrier. Oligonucleotide arrays are well suited for the detailed analysis of DNA variation as they allow the detection of single nucleotide mismatches during hybridization.These analyses can include both the discovery of novel DNA variants and the determination of known variants, for example in large-scale genotyping. Sequence variations, such as single nucleotide polymorphisms (SNPs), can serve as genetic markers. Several different oligonucleotide array designs, which are composed of probes complementary to sub-sequences of a target, can be used to determine the identity and abundance of the target sequence. When oligonucleotide arrays are used to detect known polymorphisms, such as SNPs, instead of a tiling array, another design has been applied, combining two tilling arrays, termed variant detector arrays. In such a genotyping array, each SNP is tested by two VDAs corresponding to the two alternative alleles.

Fig-12. Scheme of a typical cDNA microarray assay for gene expression analysis



In this example, mRNA levels are compared between the green and red stages of fruit development. First, mRNA is isolated from each tissue and reverse-transcribed in the presence of different fluorescent dyes resulting in labelled cDNA. Next, the two cDNA populations are mixed and hybridized to a cDNA microarray. Each array element contains DNA representing a different gene. The specific cDNAs from both populations, representing individual transcripts, will hybridize specifically with the probe on the corresponding array element. After hybridization, the microarray is scanned with a confocal laser device for fluorescence emission at two wavelengths after independent excitation of the two dyes. The relative abundance of mRNA from each gene in green vs. red fruit is reflected by the ratio green/red as measured by the fluorescence emitted from the corresponding array element. Image analysis software is used to determine fluorescence intensities that allow the quantitative comparison between the two stages of fruit development for all genes on the array.

Current Application of molecular markers

Molecular markers have already shown their applications in a variety of ways in several plant species (Gupta and Varshney 2004). The development of Genic Molecular Markers now permits a targeted approach for detection of nucleotide diversity in genes controlling agronomic traits in plant populations. Some main areas of plant breeding and genetics, where the implementation of Genic Molecular Markers will prove quite useful, are discussed below.

Trait Identification and Mapping

One of the main applications of molecular markers in plant breeding is their use as diagnostic markers for the trait in the selection. However, use of Random Molecular Markers as a diagnostic tool entails the risk of losing the linkage through genetic recombination. Even in case of Genic Molecular Markers, the gene-targeted markers where polymorphism was discovered through one allele analysis without any further specification of the polymorphic sequence motif are threatened by the same way (Rafalski and Tingey, 1993). In contrast to Random Molecular Markers or Gene-Targeted Markers, Functional markers , Direct functional markers or Indirect functional markers allow reliable application of markers in populations without prior mapping and the use of markers in mapped populations without risk of information loss owing to recombination.

Functional Diversity Analysis

Characterization of genetic variation within natural populations and among breeding lines is crucial for effective conservation and exploitation of genetic resources for crop improvement programmes.

Molecular markers have proven useful for assessment of genetic variation in germplasm collections (Hausmann et al. 2004; Maccaferri et al. 2006). Evaluation of germplam with Genic Molecular Markers might enhance the role of genetic markers by assaying the variation in transcribed and known function genes, although there may be a higher probability of bias owing to selection. While using the genic SSR markers for diversity studies, the expansion and contraction of SSR repeats in genes of known function can be tested for association with phenotypic variation or, more desirably, biological function (Ayers et al. 1997).

The presence of SSRs in the transcripts of genes suggests that they might have a role in gene expression or function; however, it is yet to be determined whether any unusual phenotypic variation might be associated with the length of SSRs in coding regions as was reported for several diseases in human (Cummings and Zoghbi 2000). Similarly, the use of SNP markers for diversity studies may correlate the SNPs of coding *vs.* noncoding regions of the gene with the trait variation. The variation associated with deleterious characters, however, is less likely to be represented in the germplasm collections of crop species than among natural populations because undesirable mutations are commonly culled from breeding populations (Cho et al. 2000). Several studies involving Genic Molecular Markers, especially genic SSRs, have been found useful for estimating genetic relationship on one hand (see Gupta et al. 2003 Gupta and Rustgi 2004, Varshney et al. 2005) while at the same time these have provided opportunities to examine functional diversity in relation to adaptive variation (Eujayl et al. 2001, Russell et al. 2004). It seems likely that with the development of more GMMs in major crop species, genetic diversity studies will become more meaningful by a shift in emphasis from the evaluation of anonymous diversity to functional genetic diversity in the near future. Nevertheless, use of the neutral Random Molecular markers will remain useful in situations where: (i) Genic Molecular Markers would not be available, and (ii) to address some specific objectives e.g. neutral grouping of germplasm.

Interspecific or Intergeneric Transferability

Perhaps one of the most important features of Molecular Markers is that these markers provide high degree of transferability among distantly related species. In contrast, except RFLPs all other Random Molecular markers are generally constrained in this regard. Transferability of Genic Molecular markers to related species or genera has now been demonstrated in several studies For example, a computational study based on analysis of ~1000 barley Genic Molecular Markers suggested a theoretical transferability of barley markers to wheat (95.2%), rice (70.3%), maize (69.3%), sorghum (65.9%), rye (38.1%) and even to dicot species (~16%). Infact, *in silico* analyses of Genic Molecular Markers of wheat, maize and sorghum with complete rice genome sequence data have provided a larger number of anchoring points among different cereal genomes as well as provided insights into cereal genome evolution (Sorrells et al. 2003, Salse et al. 2004). In some studies, the use of Genic Molecular Markers of major crop species has been shown to enrich the genetic maps of related plant species for which little marker information is available. For example, barley EST-SSR as well as EST-SNP markers have been shown transferable as well as mappable in syntenic regions of rye (Varshney et al. 2004, 2005, 2007). Further, such kind of markers from the related plant species offers the possibility to develop anchor or conserved orthologous sets (COS) for genetic analysis and breeding in different species. In this direction, Rudd et al. (2005) identified a large repository of such COS markers and developed a database called “PlantMarker”.

Expression profile Analysis

Gene expression monitoring currently is the most widespread application of Molecular Markers such as microarrays. Microarray assays may be directly integrated into functional genomic approaches aimed both at assigning function to identified genes, and to studying the organization and control of genetic pathways acting together to make up the functional organism. The rationale behind this approach is that genes showing similarity in expression pattern may be functionally related and under the same genetic control mechanism. At present, both cDNA microarrays and oligonucleotide microarrays are used for gene expression monitoring.

Future of plant molecular markers

It is clear that the genic molecular markers and especially the functional markers are extremely useful source of markers in plant breeding for marker-assisted selection because these markers may represent the genes responsible for expression of target traits. If so, there will not be any recombination between the markers and the trait, thus representing perfect indirect selection tools. While low level of polymorphism is an inherent feature of the genic molecular markers, it is compensated by their higher interspecific transferability as well as capacity to sample the functional diversity in the germplasm. These features make the development and application of the genic molecular markers more attractive for plant breeding and genetics.

With more DNA sequence data being generated continuously, the trend is towards cross-referencing genes and genomes using sequence and map-based tools. Because polymorphism is a major limitation for many species, SSR and SNP based Genic Molecular Markers will be valuable tools for plant geneticists and breeders. In the longer term, development of allele-specific, functional markers for the genes controlling agronomic traits will be important for advancing the science of plant breeding. In this context genic SSR and SNP markers together with other types of markers that target functional polymorphisms within genes will be developed in near future for major crop species. The choice of the most appropriate marker system, however, needs to be decided on a case-by-case basis and will depend on many issues including the availability of technology platforms, costs for marker development, species transferability, information content and ease of documentation.

CONCLUSIONS

Literature review indicated that since the advent of RFLP markers, a range of other markers has been introduced during the last two decades of the 20th century to fulfill various demands of the breeding programmes for crop improvement. These markers have been acted as versatile tools and have found their own position in various fields like taxonomy, physiology, embryology, Molecular biology and genetic engineering. Ever since their invention they are being constantly modified for enhanced utility as a means to solving problems and to bring about automation in the genome analysis, gene tagging, phylogenetic analysis, and selection of desirable genotypes etc. It is also evident that molecular markers offer several advantages over traditional phenotypic markers as they provide data that can be analyzed objectively. This gives new dimensions to breeding especially with respect to the time required for developing new and improved crop varieties.

REFERENCES

1. Gupta PK, Roy JK, Prasad M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants., *Curr Sci*; **80**: 524-535 (2001)
2. Doveri S, Lee D, Maheswaran M, Powell W Molecular markers: History, features and applications. In Principles and Practices of Plant Genomics., Volume 1, C.K.a.A.G. Abbott, ed. (*Enfield, USA: Science Publishers*), pp. 23-68(2008)
3. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome., *Nature*; **437**: 1299-1320 (2005**)**
4. Khatkar MS, Zenger KR, Hobbs M A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in Holstein-Friesian cattle., *Genetics*; **176:** 763-772 (2007)
5. Rafalski A Applications of single nucleotide polymorphisms in crop genetics. *Curr Op in Plant Biol*; **5:** 94-100 (2002)
6. Katti MV, Ranjekar PK, Gupta VS Differential distribution of simple sequence repeats in eukaryotic genome sequences., *Mol Biol Evolut*; **18:** 1161-1167(2001)

7. Toth G, Gaspari Z, Jurka J. Microsatellites in different eukaryotic genomes., Survey and analysis. *Genome Res* **10:** 967-981 (2000)
8. Schlotterer C. Evolutionary dynamics of microsatellite DNA., *Chromosoma*; **109**: 365-371 (2000)
9. Subramanian S, Mishra RK, Singh LGenome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol*; **4:** R13 (2003)
10. Moxon ER, Wills C DNA microsatellites: Agents of evolution., *Sci Am*; **280:** 94-99 (1999)
11. Gupta M, Chyi YS, Romeroseverson J, Owen JL. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor Appl Genet*; **89**: 998-1006 (1994)
12. Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* ; **11**: 2453-2465 (2002)
13. Mortimer J, Batley J, Love C, Logan E, Edwards D. Simple Sequence Repeat (SSR) and GC distribution in the Arabidopsis thaliana genome. *J of Plant Biotechnol* **7:** 17-25 (2005)
14. Akbari M, Wenzl P, Caig V, Carlig J, Xia L, Yang S, Uszynski G, Mohler V, ehmensiek A, Howes N, Sharp P, Huttner E, Kilian A Diversity arrays technology (DArT) for high throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.***113**:1409-1420 (2006)
15. Akopyanz N, Bukanov NO, Westblom TU, Berg DE PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen Helicobacter pylori. *Nucleic Acid Res.* **20:**6221-6225 (1992)
16. Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol.* **132**:84–91 (2003)
17. Batley J, Mogg R, Edwards D, O'Sullivan H, Edwards KJ A high-throughput SNuPE assay for genotyping SNPs in the flanking regions of Zea mays sequence tagged simple sequence repeats. *Mol. Breed.* **11**:111–120 (2003)
18. Caudrado A, Schwarzacher T The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosome* **107**:587–594 (1998)
19. Decroocq V, Fave MG, Hagen L, Bordenave L, Decroocq S Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theor. Appl. Genet.* **106**:912–922 (2003)
20. Demeke T, Adams RP, Chibbar R Potential taxonomic use of random amplified polymorphic DNA (RAPD): a case study in Brassica. *Theor. App. Gene*. **84**:567-572 (1992)
21. Ding C, Cantor CR A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight *MS. Proc. Natl. Acad. Sci. USA* **100:**3059–3064. (2003)
22. Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR Medicago truncatula EST-SSRs reveal crossspecies genetic markers for Medicago spp. *Theor. Appl. Genet.* **108**:414–422 (2004)
23. Gao LF, Jing LR, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. *Theor. Appl. Genet.* **108**:1392–1400 (2004)
24. Garg K, Green P, Nickerson DA Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Res*. **9**:1087–1092 (1999)
25. Goldstein DB, Pollock DD Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.* **88**:335–342 (1997)
26. Hancock JM Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein DB, Schlotterer C (eds) Microsatellites: evolution and applications. *Oxford University Press*, Oxford, pp. 1–9 (1999)
27. Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, et al. A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* **148**:479–494 (1998)

28. Hearne CM, Ghosh S, Todd JA Microsatellites for linkage analysis of genetic traits. *Trends Genet.* **8:**288–294 (1992)
29. Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, et al. Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. *Mol. Breed.* **3**:381-390 (1997)
30. Kantety RV, Rota ML, Matthews DE, Sorrells ME Data mining for simple-sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. Plant Mol. Biol. **48**:501–510 (2002)
31. Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas (Lund)* **135**:145–151 (2001)
32. Liu S, Anderson JA Marker assisted evaluation of Fusarium head blight resistant wheat germplasm. *Crop Sci*. **43**:760–766 (2003)
33. Luo C, Deng L, Zeng C High throughput SNP genotyping with two mini-sequencing assays. *Acta Biochim. Biophys. Sin.* **36**:379– 84 (2004)
34. Mansfield ES, Worley JM, McKenzie SE, Surrey S, Rappaport E, Fortina P Nucleic acid detection using non-radioactive labeling methods. *Mol. Cell. Probes* **9**:145-156 (1995)
35. Matsumoto A, Tsumura Y Evaluation of cleaved amplified polymorphic sequence markers. *Theor. Appl. Genet.* 110: 80–91 (2004)
36. Rafalski A The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **2:** 225- 238 (1996)
37. Prins R, Groenewald JZ, Marais GF, Snap JW, Koebner RMD AFLP and STS tagging of Lr19, a gene conferring resistance to leaf rust in wheat. *Theor. Appl. Genet.* **103**:618–624 (2001)
38. Qi X, Stam P, Lindhout P Use of locus-specific AFLP markers to construct a high-density molecular map in barley. *Theor. Appl. Genet.* **96**:376-384 (1998)
39. Qureshi SN, Saha S, Kantety RV, Jenkins JN EST-SSR: a new class of genetic markers in *cotton. J. Cotton Sci*. **8:**112–123 (2004)
40. Rabouam C, Comes AM, Bretagnolle V, Humbert JF, Periquet G, Bigot Y Features of DNA fragments obtained by random amplified polymorphic DNA (RAPD) assays. *Mol. Ecol.* **8:** 493-503 (1999)
41. Rafalski JA Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci.***162:**329–333 (2002)
42. Sobrino B, Briona M, Carracedoa A SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci. International* **154:**181–194 (2005)
43. Song QJ, Fickus EW, Cregan P Characterization of trinucleotide SSR motifs in wheat. *Theor. Appl. Genet.* **104:**286–293 (2002)
44. Southern EM Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. BioI.* 98:503-517 (1975)
45. The Rice Genome Sequencing Project The map-based sequence of the rice genome. *Nature* **436**:793-800 (2005)
46. Thiel T, Michalek RK, Varshney RK, Graner A Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet*. **106**: 411–422 (2003)
47. Vignal A, Milana D, Sancristobala M And Eggenb A A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34:**275-305 (2002)
48. Von Post R, von Post L, Dayteg C, Nilsson M, Forster BP and Tuvesson S A high-throughput DNA extraction method for barley seed. *Euphytica* **130**:255–260 (2003)
49. Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. USA* **101**: 9915-9920 (2004)