http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 5 (10): 1135-1140, 2002

Molecular Markers in Plant Breeding-I: Concepts and Characterization

S. Farooq and F. Azam
Nuclear Institute for Agriculture and Biology, P.O. Box 128, Jhang Road, Faisalabad, Pakistan

Abstract: Plant breeding is a combination of principles and methods of changing the genetic constitution of a plant to make it more suitable for human needs. Conventional plant breeding has evolved with the passage of time from simple seed saving of the best harvest to the selection of seeds according to the laws of Mendel. With the advent of morphological and biochemical markers, the selection process has hastened and the scope of conventional plant breeding increased many folds during the recent years. However, the process that actually revolutionized the plant breeding in the 20th century was the realization that there exist a widespread polymorphism in natural populations, the degree of which can be assessed by sequencing or making restriction maps: an application of new tools of molecular biology. The first and the foremost molecular markers system i.e., restriction fragment length polymorphism (RFLP) was developed in early 1980. These are co-dominant markers and are available in unlimited number. Another breakthrough was the emergence of polymerase chain reaction (PCR) in 1990. With this technology, a new generation of DNA markers such as randomly amplified polymorphic DNA (RAPDs), sequence characterized amplified regions (SCARs), sequence tagged sites (STS), single polymorphic amplification test (SPLAT), variable number of tendom repeats (VNTRs), amplified fragment length polymorphism (AFLP), DNA amplification fingerprinting (DFA), single strand conformational polymorphism (SSCP), single nucleotide polymorphism (SNPs), micro-satellites or short tandem repeats (STRs), cDNA, DNA micro arrays and rDNA-ITS were introduced into the modern plant breeding systems. The concept of DNA based markers has increased our ability many folds, to follow minute regions of chromosome through opportunities such as map based cloning and Marker-assisted Plant (MAP) Breeding. In MAP breeding, the new ideas and concepts have been introduced which need to be understood thoroughly, before applying these ideas in practical breeding programmes particular in country like Pakistan where application of molecular biological approaches are still in its infancy. In order to achieve this objective, efforts were made to write a series of review articles in which concept of MAP breeding is being described thoroughly yet in a simple way so that student and researchers can build their understanding of plant molecular breeding and application of DNA based markers for the genetic dissection of qualitative and quantitative traits. In the present paper, concepts of MAP breeding and the type and characteristics of DNA markers is presented so that choice of the marker(s) can be made rational and for the defined purposes.

Key words: Molecular markers, DNA markers, genetic markers, marker-assisted plant (MAP) breeding, plant breeding

Introduction

Plant breeding is a combination of principles and methods of changing the genetic constitution of a plant to make it more suitable for human need. It is a combination of science and art, seated in the ability of a breeder to identify differences in the traits of economic importance among plants and to improve these traits with available scientific knowledge. Although plant breeding has been existing since about 10,000 years ago, modern plant breeding methods are based on scientific principles of cytogenetics and genetics which begins only with the rediscovery of Mendel's paper that was originally published in 1866 (Mendel, 1866). Mendel's law of inheritance provided the basis of the vast knowledge that has accumulated in genetics through which, it was established that the units of specific material (genes), which can be transferred from one generation to the next, actually determine the inheritance of traits. Since then, plant breeders have been trying to reorganize these genes to recombine desirable traits in one variety in order to make it more suitable to human needs. Conventional plant breeding that involves crossing of the best plants possessing the most desirable traits (e.g., high yield or disease resistance) has helped in achieving this target to a considerable extent. For this purpose thousands of individual plants are selected and tested for developing a variety that can take up to 15 years for wheat, 18 years for potatoes, and even longer for some other crops.

Although, present day plant breeding is much advanced in terms of hybridization and selection procedures, several problems are still unsolved. Perhaps the most important among these is, that breeders like to introduce into their cultivated plants only the gene(s) of interest while conventional breeding methods rely on the transfer of whole genome. This means that along with the gene of interest, undesirable characters will also be co-inherited, and would be eliminated only through extensive back crossing (BC) followed by rigorous selections. It has been proved experimentally (Murray *et al.*, 1988) that even after BC10 generation, possible recovery of recurrent parent genome is only 90%. Hence, at least 20 generations are required for the recovery of full recurrent parent genome. This makes the procedure considerably laborious, time consuming and lengthy. Also, after meiosis in a heterozygote, the alleles that govern a complex character, dispersed among the segregating progeny thus, the effect of some of these genes cannot be detected in the presence and/or absence of other genes because of their interaction. This means that selection of a recombinant of interest would be difficult unless all the selections are repeatedly tested and stabilized in the field that is again a time consuming process.

The transfer of recessive genes through classical breeding is even more lengthy as this requires additional generation of selfing after every back crossing and thus not suitable especially for hybrid maize production where turnover time for a hybrid is very short. Some of the characters like complex disease resistance reaction, biotic stresses, mineral deficiencies/toxicity that show continuous variation and do not fit into Mendelian ratios are most difficult to detect and transfer through conventional plant breeding. These characters are controlled by multiple loci known as Quantitative Trait Loci (QTL) and have very strong genetic components but can not be measured by individually recognizable loci under normal conditions of measurement. Thus for quantitatively inherited characters, slow pace of sorting and selection of genetically stable crop varieties is a major problem of conventional plant breeding.

Development of markers systems: As the plant breeding progressed, different approaches were used to solve the problem faced by the plant breeders. One of such approaches is the development of marker systems that was initiated with the mutations at the loci controlling plant morphology (Stadler, 1929). The variations represented through these mutations were observed as an altered plant phenotypes that ranges from pigment differences and gross changes in development (such as vernalization habit or dwarf vs. tall habit) to disease resistance response. Nevertheless, morphological markers (Worland *et al.* 1987) have not been used extensively in practical plant breeding

because of the limited availability of different mutants and those available were not neutral in their effect on agronomic phenotype. These markers were also not considered reliable because phenotypic identification could be misleading due to complex genotype and environment interaction that governs the trait of interest. The dominant phenotypes do express but at a very low frequency in a particular species thus, making their effective utilization in plant breeding difficult.

To find solution to some of these problems, protein and isozyme markers were developed (Market and Moller, 1959). It was no later than 1980 (medina-Finho, 1980) that a tight genetic linkage was established between a nematode resistance gene and an *Aps* isozyme allele in tomato which opens the avenue of tagging gene(s) of agronomic importance (Tanksley and Rick, 1980). The effect of isozymes and other proteins on plant's phenotype is usually neutral and both of them are often expressed co-dominantly making the discrimination possible between homozygote and heterozygote. However, due to limited number of protein and isozyme markers and because of the requirement of a different protocol for each isozyme system, their utilization is also very limited in plant breeding programmes. A new marker system that is restriction fragment length polymorphism (RFLP) was thus developed in early 1980 (Botstein *et al.*, 1980). RFLP markers are co-dominant and available in unlimited number because only a small (1000 nucleotide base pair) fragment is used for cloning from genomes that may contain a billion or more base pairs linearly arranged along the chromosomes. Another breakthrough (Williams *et al.*, 1990) was the emergence of polymerase chain reaction (PCR). With this technology, a new generation of DNA markers such as randomly amplified polymorphic DNA (RAPD), sequence characterized amplified regions (SCARs), sequence tagged sites (STS), single polymorphic amplification test (SPLAT), variable number of tendom repeats (VNTRs), amplified fragment length polymorphism (AFLP), DNA amplification fingerprinting (DFA), single strand conformational polymorphism (SSCP), single nucleotide polymorphism (SNPs), micro-satellites or short tandem repeats (STRs), cDNA, DNA micro arrays and rDNA ITS were introduced into the modern plant breeding systems. Some of these markers are relatively simple, easy to use, automatable, often co-dominant, near infinite in number and are comparatively faster to assay. Frequent application of such markers systems transformed the classical/conventional plant breeding programmes into MAP breeding or plant molecular breeding. However, before initiating large-scale utilization of markers in plant breeding programmes, it is imperative to have clear concepts of gene and genetic marker, their characteristics and purposes so that the technique can be used effectively.

What are gene and genetic markers?: If we consider a specific group of plants of a particular species and a set of characteristics of these plants then this set of characteristics will be defined as a "trait" of that group provided, each plant possesses one of these characteristics. The state of trait will be known as "phenotype" and the genetic information possessed by each plant will be known as "genotype". This will be considered as entirety of its genetic information or a part of it. The expression of a phenotype is governed by different loci that are composed of a set of genes containing both coding and non-coding DNA sequences. In marker-assisted breeding, the term "gene" is frequently used to denote a defined segment of the DNA of an individual as a unit of transmission and not only in the narrow sense of a "functional gene. Gillet (1996) described a locus as a set of transmission homologous genes. In most of the tree species and in humans, all the nuclear loci (except those located on X or Y-chromosome) are diploid while many agricultural crops are polyploid. Accordingly, a trait will be considered a "genetic trait" only, if any two individuals possess the same genotype and also have the same phenotype irrespective of the environment in which they exist (Gillet, 1996; Bergmann *et al.*, 1989). The relationship between genotype and phenotype can be confirmed through inheritance analysis. If a trait is inherited successfully and assigned unambiguously to the phenotype or to a set of one or more loci, than it will be known as "genetic marker". This assignment will recognize all the genes involved in a genotype that are responsible for a particular phenotype. Similarly, a trait will be considered as "gene marker" if there is a 1:1 relationship between phenotype and genotype in a way that all the alleles present at each of the involved loci are unambiguously specifiable for each phenotype.

The usefulness of a marker is completely dependent on its characteristics. However, some questions are I) why the DNA markers were invented, ii) what is new that the DNA makers can accomplish, and iii) why so many types of DNA markers are being invented every other day? To answer these questions, it is essential to know pre-DNA markers and their characteristics including mode of inheritance/transmission, mode of gene action and level of genetic variability that can be detected through these marker systems.

Mode of inheritance could either be "bi-parental" which means all nuclear genes in a plant will be independently inherited, or uni-parental which means only maternal nuclear and organelle gene or paternal organellar gene will be inherited independently. Mode of transmission for nuclear genes could either be diploid or haploid depending upon the phase of the material, number of loci coding the phenotype and number of alleles per locus. The mode of inheritance will also depend upon the gene action that could either be co-dominant, which means, both the alleles present at a diploid or polyploid locus are always scorable hence, heterozygote will always be recognizable as such. Thus, a gene marker will always show co-dominant inheritance. The gene action could also be dominant, which means that the dominant allele will mask the presence of the recessive allele in such a way that the heterozygote (dominant/recessive) will not be distinguishable from the homozygote (dominant/dominant). Any such locus will be called as "genetic marker" like the loci showing "epistasis" that is, the expression of an allele at one locus will mask the expression of an allele at another locus.

The level of genetic variability could be a measure of genetic differences that can be detected with a particular marker system within a sub-population. Generally, different accessions of a particular species will show mixed profiles upon analyses with a particular isozyme system. This means some of the accessions will show identical profiles, some show nearly identical profile and still some others will show completely different profiles. In extreme cases however, the profiles could either be monomorphic i.e. all individual will have same phenotype, or polymorphic i.e. each individual will have a unique phenotype. The differences in profiles exhibited by the accessions could either be qualitative (different types) or quantitative (difference in the frequency distributions of the same types). The accessions would be completely undifferentiated if the frequency distribution of some bands is equal in all accessions. On the other hand if the bands present in each accession are completely distinct, accessions would be completely differentiated from each other (Gregorious, 1988).

Pre-DNA markers and their characteristics: Isozymes were the most frequently used marker system before the advent of DNA markers. These are electrophoretically separable variants of an enzyme (Bergmann *et al.*, 1989) coded by genes at one or several loci. Enzyme molecules are direct products of genes, and thus of DNA, and play essential roles in the primary and secondary metabolism of the organism. These molecules are composed of chains of amino acids as determined by the DNA sequences of the coding genes. Differences in the total electrostatic charges of amino acid sequences indicate the existence of difference in the DNA sequence. Allozymes generally differ due to the substitution of single amino acid of different charge at a locus. Such changes in amino acid composition often alter the charge or less often the conformation of the enzyme. This leads to a change in electrophoretic mobility of the enzymes, and thus provides an

extremely useful method of evaluating genetic differences among groups of individual plants and accessions. The ability to observe allelic variation at different isozyme loci revolutionized the researches in the fields of biochemical genetics, population genetics and evolution. Allozymic polymorphism has been used in plants to examine genetic processes at every stage of the life cycle and to ascertain genetic diversity in all major crops as well as in many other species.

Allozymes generally exhibit Mendelian inheritance and are transmitted by one or only a few nuclear loci accompanied by co-dominant gene action (except for null allele). This ensures the identification of both the genes at a locus. Thus, all the allozymes can be used as "gene marker". Complete penetrance and absence of pleiotropic and epistatic interactions could be considered added advantages. The most significant of all the properties of the isozymes is the simple genetic base of most of the identified polymorphism. It is because of these characteristics that isozyme markers were used in almost every discipline of plant breeding, which has been thoroughly documented (Tanksley and Orton, 1983). Gregorious and Bergmann (1995) found some additional characteristics of certain isozymes among different populations that led to a general assumption that isozyme markers are selectively "neutral". Farooq and Sayyed (1999a and 1999b) observed qualitative and quantitative difference in the profiles of isozyme peroxidase exhibited by cotton (*Gossipium hirsutum*) growing under normal and heavily stressed environment. This means that isozyme markers do have the theoretical ability to respond to contrasting environments. In the stressed environments for example, changes occurring in the plants can only be detected by a "non-neutral marker system. Such reported effects of environment on the properties of some of the isozymes and the fact that every isozyme requires a new protocol could be some of the possible reasons that these are not the markers of choice in modern plant breeding programmes.

How to choose a suitable marker system?: Studies conducted during the last decade of the 20th century reported numerous DNA markers (Farooq, 2001) that have been and are still being utilized in plant breeding programmes with well-defined objectives nevertheless, isozyme markers are still being widely used as gene markers (Gillet, 1999). This is because I) they are inexpensive compared to DNA markers, ii) protocols for most of the enzyme and protein are very well-established, iii) role of isozymes in metabolism are very well known and iv) the typical level of variation (both qualitative and quantitative) makes them suitable for a number of other purposes (Farooq and Sayyed, 1999a,b). The choice of markers is thus objective dependent. For example if differentiation were to be made between two similar accessions of a species, molecular markers like RAPD, RFLP, or AFLP would do the job (Farooq *et al.*, 1996). However, if closely related varieties differing only in few environmentally induced physiological characters are to be discriminated, the non-neutral marker system such as DNA micro-array or the isozyme markers would be a better choice. Since utilization of DNA markers, especially the DNA micro array, is expensive and requires a lot of skill, most laboratories (especially those in the developing countries) would prefer to use isozyme markers. In order to assess the qualitative and quantitative differences produced in a crop before and after the attack of pests or diseases, isozyme would again be preferred (Farooq and Sayyed, 1999a,b). There are however, certain cases where even isozyme analyses do not solve the problem. For example, it was difficult to discriminate between a commercial rice cultivar Basmati 370 and its two dwarf mutants through isozyme analysis due to low level of polymorphism. This problem was finally solved through the use of molecular (RAPD) markers (Farooq, 2001).

The production techniques, advantages and disadvantages of DNA markers: Going through the literature on the development and utilization of molecular markers in plant breeding, four techniques appeared prominent. These include I) restriction Fragment length polymorphism (RFLP), ii) random amplified polymorphic DNAs (RAPDs), iii) amplified fragment length polymorphism (AFLP), and iv) single sequence repeats (SSRs) or micro-satellites. In addition, Internal Transcribed Spacer (ITS) and cDNA markers have also made important contribution.

The information for different genes and gene systems in higher plants is stored in the DNA sequences of the nuclear chromosomes and the organelle genome. The DNA content of plant genome differs significantly between species, some of which possess as little as 0.5 pg DNA or 0.5 x10 base pairs such as rice, while another cereal has 17 pg per 1C nucleus or 1.6 x 10 base pairs. Plants are able to replicate their DNA with high accuracy but due to mechanisms like mutations, changes occur in simple base pairs of DNA. In addition, changes also occur due to inversions, translocations, deletions or transposition. Thus there exists a tremendous amount of DNA variation in natural population of plants. One of the ways to access this variation is by using a special class of enzymes called restriction enzymes. These enzymes have the ability to recognize and cleave target sites (called restriction sites) made up of specific base sequences in the DNA. A large piece of DNA will thus be reduced to a series of smaller fragments of defined sizes by digestion with a restriction enzyme. Such differences in fragment size arising as a consequence of restriction enzyme digestion of a nucleotide or total DNA are called Restriction Fragment Length Polymorphism (RFLP).

Restriction fragment length polymorphism (RFLP): This was the first DNA marker (Botstein *et al.*, 1980), which resulted from the difference in the sequences of nucleotide of different plants. Such differences arise due to mutations occurring over time and are detected as variation (polymorphism in the length of a restriction fragment). The most common mutation is substitution of a single nucleotide that eliminates a restriction site in which substitution occurred. It can also change the length of a DNA fragment and detected as a discrete marker that is directly representative of an individual's genotype. It is also possible to detect rearrangements in the DNA sequences intervening between two restriction sites, or two priming sites such as insertion and/or transposition, deletion or error in DNA replication that can generate unique variation termed as allele. A battery of restriction enzymes, which displays multiple alleles per enzyme, is used to generate profiles for identification of an individual that is also known as DNA fingerprinting.

RFLP analysis of plants involves several steps. These includes extraction of DNA from plant, digestion of DNA with restriction enzymes, size fractionation on a gel using electrophoresis, transfer of DNA fragments on to a hybridization membrane, cloning of the individual fragment into a plasmid, labeling of cloned DNA sequences with ^{32}P and hybridization of labeled single stranded probe to its single stranded DNA counterpart on the filter. Un-hybridized probe is then washed off, and the filter exposed to x-ray film to obtain an autoradiogram. The bands visible on an autoradiogram represent the restriction fragments of the digested DNA that contain the sequences homologous to the cloned sequences used as probe. RFLP analysis requires high quality DNA. It is an extremely reliable method for DNA typing and for determining genetic relationship or disease linkage. The biggest advantage of RFLP is that this technique is robust and readily transferable between laboratories. Nevertheless, the wider application of RFLP is restricted due to several limitations. These limitations include I) requirement of 50-200 microgram DNA requiring large scale extraction which is tedious and laborious, ii) overall lengthy procedure and iii) involvement of radioactivity.

Amplified fragment length polymorphism (AFLP): AFLPs are fragments of DNA that have been amplified using directed primers from restriction digestion of genomic DNA (Metthes *et al.*, 1998; Karp *et al.*, 1997). AFLP refers to molecular markers obtained by

selective PCR amplification of restriction fragments. The technique generates fingerprints of any DNA, regardless of its source, and without prior knowledge of DNA sequence. Scoring AFLP data is simple and easy because polymorphism are identified in the form of presence/absence of data rather than determination of sizes at various loci. The technique involves three steps i.e., I) restriction enzyme digestion, ii) ligation of adapters and iii) selective amplification of restriction fragments based on recognition of unique nucleotides flanking the restriction site. By varying the number of these additional nucleotides that extend beyond the restriction sites into the unknown sequences, it is possible to control the proportion of the ligated fragments that could be amplified. In general, 75-150 fragments are amplified with each primer combination. As each fragment represents a unique site, the proportion of the genome assayed with each primer combination is thus, much higher than with any other DNA analysis approach.

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 co-dominant at others. Thus, they act both as gene and genetic markers. Level of variability is abundant as each and every pattern is unique. The major advantage of the AFLP technique is generation of a large number of polymorphism. The fact that no sequence information is required and that the PCR technique is fast with a high multiplex ratio (Rafalski *et al.*, 1996) makes the AFLP very attractive choice.

The problems associated with AFLPs are of three types and all are related with practical handling, data generation and analysis. These problems are not unique to AFLP technology but also associated with other marker systems. An ideal marker should have sufficient variation for the problem under study, be reliable and simple to generate and interpret. Unfortunately, neither AFLP nor other DNA markers exhibit these qualities. Thus a specific technique or techniques selected on the basis of objectives be utilized collectively to achieve the best results (Karp *et al.*, 1997; Harris, 1999).

Micro-satellites or short tandem repeats (STRs): These are ideal genetic markers for detecting differences between and within species. It consists of tandomly repeated 2-7 base pair units, distributed widely throughout the genome. These are heritable, useful to monitor gene flow, excellent for parentage determination and forensics and ideally suitable for high throughput analysis via multiplexing with highly reproducible profiles. Variation in the number of tandomly repeated units results in highly polymorphic banding patterns. Profiles are generated by PCR amplification of unique loci using discriminatory primers sets. Micro-satellites may even be used across species and genus boundaries.

Micro-satellites are not limited to the nuclear genome. They occur in chloroplast as well as in mitochondrial genome (Soranzo *et al.*, 1999) as a repetition of guanine and cytosine. Nuclear micro-satellites are of bi-parental inheritance and possess few loci with many alleles per locus. They are co-dominant with exception of null alleles at some loci, can detect large variation within population and low variation between populations, are non-coding and thus confer stability to the genome. Chloroplast micro-satellites (cp) on the other hand are of uni-parental inheritance with single locus (many alleles per locus), are non-coding and detects low variation within population and large variation between populations. Their genome does not recombine due to its paternal inheritance in conifers (Cato and Richardson, 1996; Vendramin and Ziegenhagen, 1997; Sperisen *et al.*, 1998) and maternal inheritance in angiosperms (Dumolin *et al.*, 1995). For this reason, cp. STR variants accumulated in a uni-parental chloroplast lineage can provide information about the history of tree population. Mitochondrial micro-satellites are also of uni-parental (maternal), single locus inheritance with many alleles per locus. They are also non-coding and can detect low variation within population but large variation between regions. Chloroplast micro-satellite variants are supposed to be generated in a stepwise manner by addition or deletion of a single repeat unit. Under such a stepwise mutation model (Valdas *et al.*, 1993), micro-satellite variants with small repeat length differences are more closely related than alleles with larger length differences (Jarne and Lagoda, 1996).

Ribosomal DNA internal transcribed spacers (rDNA-ITS): The Ribosomal DNA (rDNA) is a multigene family with nuclear copies in eukaryotes that are arranged in tandem arrays in nucleolar organizer regions (NORs) generally at more than one chromosomal location. Each unit within a single array consists of genes coding for small (18S) and large (28S) rRNA subunits. The 5.8 S nuclear rDNA gene lies embedded between these genes but separated by two internal transcribed spacers ITS1 and ITS2. The copy numbers of 18S-5.8S-28S rRNA genes in diploid genomes of *Quercus cerris*, *Q. ilex*, *Q. petraea*, *Q. pubescens* and *Q. robur* are estimated to be in the number of 1300-4000 (Zolodos *et al.*, 1999). The small subunit is highly conserved and has been used to shed light on deep evolutionary branches, e.g. for relationship between *Archaebacteria* and *Eubacteria*, while more conserved domains within the 28S region have been used to cover evolutionary time through the Paleozoic and Mesozoic eras. The faster evolving ITS regions has been employed for population and congeneric phylogenies (Bayer *et al.*, 1996). The smallest rDNA gene of the cluster 5.8S is too short to provide a robust phylogenetic signal. Internal transcribed spacers represent bi-parental nuclear mode of inheritance through several loci with several alleles per locus, are co-dominant, exhibit high level of variability even within a single individual and composed of non-coding sequences.

Complementary DNA (cDNA): The adaptation of individuals as well as populations to the environment is not well understood because of the lack of knowledge of the genes involved in these processes. The highly developed and preferentially used marker systems such as RAPDs, AFLPs, nuclear SSRs, rDNA ITS, mainly represent repetitive regions of the nuclear genome, with the rDNA markers being further confined to the nucleolar organizer regions (NOR). These neutral markers are ideal for assessing the genetic diversity as well as evolutionary relationship of populations because the distribution of their variation is presumably not influenced by selective forces. None of the markers systems presently available are reported to have close linkage with adaptive characters. Thus, there was a dire need to develop additional "non-neutral" DNA marker system representing genes involved in adaptive processes that could pinpoint differences appearing in populations on the basis of selective/adaptive features. The recently developed DNA micro-array could facilitate such work (Schena *et al.*, 1995) and can thus help identify the genes involved in adaptation process. However, this technology requires the availability of cDNA sequences for use as probes in testing differential expression of genes anticipated under various environmental conditions (Gill *et al.*, 1997). Since complementary DNA sequences are the representative of mRNA, it is possible to pinpoint functional differences between alleles of a locus through cDNA. Complementary DNA represents bi-parental mode of inheritance

through one to few loci with few alleles per locus, is co-dominant, and exhibits low variation within and low differentiation between populations.

Allele specific PCR (AS-PCR): This refers to amplification of specific alleles or DNA sequence variants at the same locus. Specificity is achieved by designing one or both PCR primers so that they partially overlap the site of sequence difference between the amplified alleles. Many variants of this technique have been described under different names such as RAPD, AP-PCR, and DAF (DNA amplification finger printing) etc. All these are based upon polymerase chain reaction (PCR), which involves randomly synthesized short oligo-nucleotide sequence as primers. These primers are specific to the ends of a given sequence in the DNA of a plant. Genetic differences between individuals can be detected when the size of the segment of DNA bracketed by the primers is different. The bracketed region can be isolated and purified, so that restriction mapping and sequencing is possible allowing more refined analysis of genetic differences. These markers are known as randomly amplified polymorphic DNA (RAPD) markers that do not involve southern analysis and radioactive labeling. Randomly amplified polymorphic DNA markers are therefore, cheaper and easier to use than RFLP and AFLP markers. However, RAPDs are very sensitive to laboratory conditions (Farooq *et al*., 1994), are usually dominant markers, are potentially population specific (Farooq *et al*., 1995), are prone to artifacts from contamination, vary in laboratory protocols and show low level of polymorphism. AP-PCR and DAF (Caetano-Anolles *et al*., 1993) are also related with PCR amplification using single random primer. Principle is same but there are differences in the experimental details. Randomly amplified polymorphic DNA bands are detected on agarose gels using ethidium bromide staining while AP-PCR products are frequently analyzed on acrylamide gels and detected by autoradiography. The products of DAF are analyzed on acrylamide gels and detected by sliver staining.

The availability of other DNA marker systems: Sequence characterized amplified regions (SCARs). This refers to PCR amplification of individual RAPD band by using specific primers (24 mer long) based on the DNA sequence of the RAPD bands of interest. SCARs may be dominant (like AS-PCR) or may show size polymorphism (co-dominance). Polymorphism may also be identified by restriction digestion before or after amplification. A more pleasing acronym for these markers would be STARs (sequence tagged amplified regions). In these marker systems, sequencing data is required that can become a limiting factor under specific circumstances.

Sequence tagged sites (STS): This is a general term applied to any unique genome fragment amplified with 1824 base pair primers derived from known sequences or end sequenced RFLP probes. Polymorphism can be recognized as on/off (i.e. where one primer does not work in a variety and no PCR product is produced), or length variants where part of the intervening sequences are deleted or inserted. Though robust, STS are beset with low level of polymorphism that can be enhanced by development of four-base pair cutter and the later converted into a specific polymorphic marker (SPLAT).

Single polymorphic amplified test (SPLAT): If STS do not reveal polymorphism, it is usually converted into SPLAT. Individual STS products from different genotypes are themselves sequenced in search for sequence variation. Any difference revealed may be exploited in the production of internal primer. In general SPLAT are on/off dominant markers.

Variable number of tandem repeats (VNTRs): This technique has generally been developed for mammalian genome but it does show some promise in plants as well. The polymorphism is based on differential number of tandomly repeated and conserved motif, that appear to be distributed over the whole genome. The development of a VNTR marker involves identifying DNA fragments containing the tandem repeat segments usually by blot hybridization. The flanking segments are sequenced in order to design locus specific PCR primer. Alleles can then be detected by gel electrophoresis of PCR products through these locus specific primer pairs. Depending upon the repeat unit length, the tandem repeats are called mini-satellites (approximately 10-50 base pair) or micro satellite (< 5 base pair, simple sequence repeats or SSRs). Variable number of tandem repeats markers possess all advantages of RFLP and RAPD marker with no disadvantage except that they are expensive to produce and sustain (Nakamura *et al*., 1987).

Single-strand conformational polymorphism (SSCP): This relies on secondary and tertiary structural differences between denatured and rapidly cooled amplified DNA fragments that differ slightly in their DNA sequences. Different SSCP alleles are resolved on non-denaturing acrylamide gels, usually at low temperatures. The ability to resolve alleles depends on the conditions of electrophoresis. This again requires DNA sequence data.

Single nucleotide polymorphism (SNPs): This is the newest and highly automated genotyping technique that can detect changes in nucleotide sequences by one base substitution. Such variation are very common in the genome therefore, SNPs are highly abundant. SNPs use DNA strands embedded in some solid surface like silicon chips or micro spheres, and variation are resolved by lasers linked with computers. Bi-allelic data scored by presence/absence of a particular variant makes the scoring simplified (Anonymous, 2001).

DNA micro arrays: This is a powerful, versatile and economical technique for screening of genetic aberrations. High-density gene sequences are printed onto glass slides. Fluorephore labeled genomic or complimentary DNA (cDNA) is hybridized to slides with fixed signature patterns and resolved using computer driven fluorescent image. This is a sophisticated technique and requires a lot of expertise (Schena *et al*., 1995).

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. These markers have been and are being acted as versatile tools and have found their own position in various fields like taxonomy, physiology, embryology, 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

Anonymous, 2001. Molecular Characterization of Mutant Genes Controlling Important Traits for Seed Crops. IAEA TECDOC-312.D2.RC.790-2, pp: 13.

Bayer, R., D.E. Soltis and P.S. Soltis, 1996. Phylogenetic inferences in *Antennaria* (Asteraceae:Gnaphalieae:Cassiniinae) based on sequences from nuclear ribosomal DNA internal transcribed spacers (ITS). Am. J. Bot., 83: 824-832.

Bergmann, F., H.R. Gregorius and F. Scholz, 1989. Isoenzyme: Indicators of environmental impact on plants or environmentally stable gene markers? In: Genetic Effects of Air Pollutants in Forest Tree Populations (F. Scholz., H. R. Gregorius., and D. Rudin, Eds.), Springer-Verlag Heidelberg, New York, Tokyo, pp: 17-28.

Botstein, D., R. White, M. Skolnick and R.W. Davis, 1980. Construction of genetic linkage map in human, using restriction fragment length polymorphism. Am. J. Hum. Genet., 32: 314-331.

Caetano-Anolles, G. and B.J. Bassam, 1993. DNA amplification fingerprinting using arbitrary oligo-nucleotide primers. Appl. Biochem. Biotechnol., 42: 189-200.

Cato, A.A. and T.E. Richardson, 1996. Inter and intra-specific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. Theor. Appl. Genet., 93: 587-592.

Dumolin, S., B. Demesure and R.J. Petit, 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor. Appl. Genet., 91: 1253-1256.

Farooq, S., T. M. Shah, E. Askari, A. A. Zaidi, M. Arif and N. Iqbal, 1994. Identification of different wheat genotypes through polymorphism based on random amplified polymorphic DNA (RAPD). Pak. J. Bot., 26: 373-382.

Farooq, S., M. Arif and N. Iqbal, 1995. Utilization of RAPD markers for identification of cultivated and wild rice species. Pak. J. Bot., 27: 127-138.

Farooq, S., N. Iqbal and M. Arif, 1996. Detection of Genetic variability in Basmati and non-Basmati rice varieties and their radiation induced mutants through Random Amplified Polymorphic DNA (RAPD). In: Application of DNA based marker mutations in sexually propagated cereal crops (IAEA) TECDOC-1010. Vienna, Austria, pp: 85-96.

Farooq, S. and H. Sayyed, 1999a. Isozyme markers in cotton breeding-II. Inter and intravarietal variation in activity of isozyme peroxidase as affected by area of cotton cultivation. Pak. J. Bot., 31: 347-359.

Farooq, S. and H. Sayyed, 1999b. Isozyme markers in cotton breeding-III. Variation of isozyme peroxidase in different loci of different varieties and their relationship with cotton leaf curl virus disease. Pak. J. Bot., 31: 361-370.

Farooq, S., 2001. Marker-assisted plant breeding. In: New Horizon in DNA Technology (I. A. Khan and A. A. Khan, Eds.), Proc. One-day International Seminar, University of Agriculture, Faisalabad, Pakistan, pp: 73-88.

Gill, R., W.T.C. Hodgman, C.B. Littler, M.D. Oxer, D.S. Montgomery, S. Taylor and P. Sanseau, 1997. A new dynamic tool to perform assembly of expressed sequenced tags (ESTs). Com. Appl. Biosci., 13: 453-457.

Gillet, G.M., 1996. Qualitative inheritance analysis of isoenzymes in haploid gametophyte: Principles and a computerized method. Silvae Genetica, 45: 8-16.

Gillet, G.M., 1999. DNA markers - concepts and characteristics. In: Which DNA Marker for Which Purpose? (Gillet, E.M., Ed.), Institute fur Forstgenetika und Forstpflanzenzuchtung, Unuiversitat Gottingen, Germany, pp: 1-8.

Gregorius, H.R., 1988. The meaning of generic variation within and between subpopulations. Theor. Appl. Genet., 76: 947-951.

Gregorius, H.R. and F. Bergmann, 1995. Analysis of isoenzyme genetic profiles observed in tree populations. In: Population genetics and genetic conservation of forest trees Baradat, P. Baradat, W.T. Adams, and G. Muller-Starck, (Eds.), SPB Academic Publishing, Amsterdam, The Netherlands, pp: 79-96.

Harris, S.A., 1999. Molecular approaches to assessing plant diversity. In: Plant conservation biotechnology (E.E. Benson, Ed.), Taylor and Francis, London, pp: 11-24.

Jarne, P. and J.L. Lagoda, 1996. Microsatellites, from molecules to population and back. Tree, 10: 424-429.

Karp, A., S. Kresovich, K.V. Bhat, W.G. Ayand and T. Hodgkin, 1997. Molecular tools in plant genetics resources conservation: a guide to the technologies. IPGRI Tech. Bull. No. 2.

Market, C.L. and F. Moller, 1959. Multiple forms of enzymes: tissue, ontogenetic and species specific pattern. Proc. Nat. Acad. Sci. USA, 45: 753-763.

Metthes, M.C., A. Daly and K.J. Edwards, 1998. Amplified fragment length polymorphism (AFLP). In: Molecular Tools for Screening Biodiversity (A. Karp, A.P.G. Isaac, and D.S. Ingram, Eds.), Chapman and Hall, London, pp: 183-190.

Medina-Filho, H.P., 1980. Linkage of *Aps-1*, Mi and other markers on chromosome 6. Rep. Tomato Genet. Coop., 30: 26-28.

Mendel, G., 1866. Experiments in plant hybridization. In: Proc. Baunn. Natural History Society. Harward University Press, Cambridge.

Murray, M.G., Y. Ma, J. Romero-Severson, D.P. West and J. H. Cramer, 1988. Restriction Fragment Length Polymorphism: what are they and how can breeders use them? In: Proc. 43rd Annual Corn and Sorghum Research Conference (Wilkinson *et al.,* Eds.), American Seed Trade Association. Washington, DC, pp: 72-78.

Nakamura, Y., M. Lappert, P. O'Connell, R. Wolf, T. Holm, M. Culver, C. Martin, E. Fujimotot, M. Hoff, E. Kumlin and R. White, 1987. Variable number of tandom repeats (VNTR) from human gene mapping Science, 235: 1616-1622.

Rafalski, J.A., J.M. Vogel, M. Morgante, W. Powell, C. Andre and S.V. Tingey, 1996. Generating and using DNA markers in plants. In: Non-mammalian Genomic Analysis: A Practical Guide (Biren B., and E. Lai, Eds.). Academic Press, London, pp: 75-134.

Schena, M., D. Shalon, R.W. Davis and O.P. Brown, 1995. Quantitative monitoring of gene expression patterns with a complementary DNA micro array. Science, 270: 467-470.

Soranzo, N., J. Provan and W. Powell, 1999. An example of microsatellite length variation in the mitochondrial genome of conifers. Genome, 42: 158-161.

Sperisen, C., U. Buchler, G. Matyas, M. Anzidei, A. Madaghiele, T. Skroppa, and G.G. Vendramin, 1998. Polymorphic tandem repeats in the chloroplast and mitochondrial genomes of Norway spruce. In: Genetics and Breeding of Norway Spruce (Skroppa T., L. Paule, and D. Gomory, Eds.). Arbora Publisher, Svolen, Slovakia, pp: 15-24.

Stadler, L.J., 1929. Chromosome number-and the mutation rate in *Avena* and *Triticum.* Proc. Natl. Acad. Sci., USA, 15: 876-881.

Tanksley, S.D. and C.M. Rick, 1980. Isozymic linkage map of the tomato; application in genetics and breeding. Theor. Appl. Genet., 57: 161-170.

Tanksley, S.D. and T.J. Orton, 1983. Isozymes in plant genetics and breeding. Elsevier, New York.

Valdas, A.M., M. Slatkin and N.B. Freiner, 1993. Allele frequencies at microsatellite loci: The stepwise mutation model revisited. Genetics, 133: 737-749.

Vendramin, G.G. and B. Ziegenhagen, 1997. Characterization and inheritance of polymorphic plastid microsatellites in Abies. Genome, 40: 857-864.

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

Worland, A.J., M.D. Gale and C.N. Law, 1987. Wheat genetics. In: Wheat Breeding: Its Scientific Principals (Lupton, F. J. H. Ed.), Chapman and Hill, London, UK, pp: 129-172.

Zolodos, V., D. Papes, M. Cerbah, O. Panaud, V. Besendorfer and S. Siljak-Yakovlev, 1999. Molecular- cytogenetic studies of ribosomal genes and heterochromatin reveal conserved genome organization among 11 *Quercus* species. Theor. Appl. Genet., 99: 969-977.