MOLECULAR MARKER TECHNIQUES AND THEIR UTILITY IN TREE BREEDING

M. G. MALLIKARJUNA*, P. DHANANJAYA, SANGH CHANDRAMOHAN AND FARZANA JABEEN
Division of Genetics, Indian Agricultural Research Institute, New Delhi - 110 012
1Division of Seed science and Technology, Indian Agricultural Research Institute, New Delhi - 110 012
2National Bureau of Plant Genetic Resources, New Delhi - 110 012
3Department of Genetics and Plant Breeding, Acharya N. G. Ranga Agricultural University, Hyderabad - 500 030
E-mail: mgrpatil@gmail.com

INTRODUCTION

Much of the understanding of genetic makeup of organism was possible through genomics along with availability of genomic sequences for an increasing number of species and high throughput approaches. This will help to understand the function of genes in terms of their relationship to the phenotype. The relationship between complex trait variation and molecular diversity of genes can be studied based on a genomic approach but the identification of genes responsible for the variation remains a slow and time consuming process, especially in long lived organisms such as forest trees (Vendramin and Morgante, 2006). The quantification of diversity in tree species provides a basis for conservation and utilization of existing variability. It can be specified by three levels viz., morphological, biochemical and nucleic acid sequence level. The genetic diversity has an effect on the higher order of biodiversity. The genetic diversity analysis completed its long journey by using morphological traits mainly and physiological traits to some extent, but these approaches has disadvantages such as non-availability of sufficient number of morphological markers, lack of sufficient variability among the highly heritable traits, time consuming and influenced by G x E interactions (Karp et al., 1997; Rao, 2004).

Majority of the traits in forest species is governed by QTL, but the identification of QTLs in forest tree species is at a slower pace (O’Malley, 1992). The problems encountered in tree breeding are the large size of the genome, scarcity of multigenerational pedigrees, and long generation times (Tulsieram et al., 1992), this is combined with the non-availability of adequate number of morphological markers in conventional tree breeding. These problem can be overcome by some extent through the use biochemical markers where the products controlled by genes. However, for analysis of quantitative variation, the number of useful protein markers is a limiting factor. Hence there is a need to identify markers which can be effectively used for dissection of quantitative traits.

Hence at this juncture, DNA based techniques appear to have a potential to identify the polymorphism revealed by differences in DNA sequence and they can be used at all developmental stages of plants covering whole genome variability. Further, they are unaffected by environment and the measurement of genetic diversity at DNA level provides actual difference at the level of nucleic acid sequences. Along with the traditional and statistical methods the modern biotechnological tools are found to be very significant and reliable approaches. So with this context we are highlighting the important molecular markers system which can be utilized in tree breeding program with available evidences.

Tree breeding and molecular markers

With the advent of developments in molecular biology DNA based differences have become the markers of choice for nucleic acid based analysis of genetic variation. These DNA based markers are found to be useful both in basic studies...
(cloning of genes and phylogenetic studies) and applied research (marker assisted breeding). Molecular markers offer specific advantage in assessment of genetic diversity and also in trait specific crop improvement. Use of markers in applied breeding program range from facilitating appropriate choice of parents for hybridization, for mapping and tagging of gene blocks associated with economically important traits and DNA fingerprinting etc (Haussmann et al., 1999; Gupta and Varshney, 2000).

Paragon properties of a DNA marker viz., highly polymorphic nature, co-dominant inheritance (determination of homozygous and heterozygous states of diploid organisms), frequent occurrence in genome, even distribution throughout the genome, selective neutral behavior (no pleiotropic effects), easy access and fast assay, high reproducibility, easy exchange of data between laboratories and their development at reasonable cost. It is extremely difficult to find a molecular marker meeting all the above criteria. However, depending upon type of study, a marker system can be identified that would fulfill a few of the above characteristics (Weising et al., 1995). Now there are various number of marker systems are available to detect genetic polymorphism at nucleic acid level.

Types of molecular markers

Jeffreys et al. (1985b) first coined the term 'DNA finger printing' to describe a multilocus RFLP (Restriction Fragment Length Polymorphism) assay designed to distinguish human beings. The DNA based molecular marker technologies that are currently in use for the purpose of fingerprinting in plants can be broadly grouped into, (i) DNA hybridization based markers also known as ‘first generation markers’ e.g. RFLP and (ii) Polymerase Chain Reaction (PCR) based markers that utilize (a) Primers of sequence that are arbitrary to that of target genome, such as Randomly Amplified Polymorphic DNA (RAPD), DNA amplification finger printing (DAF) etc and (b) primers designed based on known sequence information, such as Simple Sequence Repeats (SSRs), these are also known as second generation markers. Finally the third generation markers include Single Nucleotide Polymorphisms (SNPs), they utilizes advent of genomic revolutions.

With the invention of PCR (Saiki et al., 1985; Mullis et al., 1986 and Mullis and Faloona, 1987) and introduction of thermo stable DNA polymerase (Saiki et al., 1988) during the mid of 1980s, the research programs employing molecular markers and its applications in plant or animal genomics and breeding programs has increased tremendously. Within a large number of available molecular markers in recent years ((Caetano - Anolles and Greshoff, 1997; Mohan et al., 1997) each having a different set of advantages in any particular applications including novel approaches like Breeding without Breeding (Harikrishna et al., 2012).

Restriction fragment length polymorphism (RFLP)

Botstein et al. (1980) first used DNA restriction fragment length polymorphism (RFLP) in human linkage mapping and this laid the foundation for the utilization of DNA polymorphisms as genetic markers. With advent of RFLP markers development, RFLPs were recovered from many loci and extended the linkage maps in Poplar (Bradshaw et al., 1994) and Loblolly pine (Devey et al., 1994).

RFLPs detect differences in the length of specific DNA fragments after digestion of genomic DNA with sequence-specific endonuclease (Botstein et al., 1980). These differences result from base pair changes or other rearrangements (e.g. translocation and inversion) at the recognition site of the restriction enzyme or from internal deletion or insertion events. The restricted fragments are separated according to size by agarose gel electrophoresis and polymorphism is detected by Southern blot analysis (Southern, 1979). The DNA sequences used as hybridization probes are generally pre-selected to contain unique or low copy number DNA sequences are labeled with 32P or biotin (Winnacker, 1987). Such probes can either be cloned DNA from specific genes, cDNAs, random genomic DNAs, or specifically synthesized oligonucleotide (Beckmann and Soller, 1986). The probes are mostly species-specific single locus probes of about 0.5 - 3.0kb in size obtained from cDNA library or a genomic library.

Though the RFLP analysis initiated the era of nucleic acid based molecular markers, it has its own limitations such as, it demands large quantity of high quality DNA, has low genotyping throughput and is very difficult to automate. Classical genotyping involves radioactive methods so its use is limited to specific laboratories. RFLP probes must be physically maintained and it is therefore difficult to share them between laboratories. In addition, the level of RFLP is relatively low and selection for polymorphic parental lines is a limiting step in the development of a complete RFLP map. However RFLP markers are found to be powerful tools for comparative and syntenic mapping (Xu, 2010).

Polymerase chain reaction (PCR) based markers

The discovery of PCR technology (Mullis and Faloona, 1987) lead to development of variety of PCR methods for various molecular biological applications. This revolutionized evolution of several PCR based molecular markers, which immensely simplified the DNA fingerprinting (genotyping) work for Marker Assisted Selection (MAS) or Marker Assisted Breeding (MAB).

In principle, PCR is an in vitro method of nucleic acid synthesis by which a particular segment of DNA can be specifically replicated. Generally, it involves two oligonucleotide primers that flank the DNA fragment to be amplified and repeated cycle of heat denaturation of the DNA, annealing of the primers to their complementary sequences and extension of the annealed primers with thermostable DNA polymerase. These primers anneal to the opposite strands of the target sequence and are oriented so that DNA synthesis by the polymerase proceeds across the region between the primers. Since the extension of products themselves are also complementary and capable of binding primers, successive cycles of amplification essentially double the amount of target DNA synthesized in the previous cycle. The result is an exponential accumulation of the specific target fragment that could be resolved by gel electrophoresis. Some of the most commonly used PCR based molecular markers are briefly discussed here under.

Random amplified polymorphic DNA (RAPD)

Williams et al. and Welsh and Mc Clelland in 1990 independently described the utilization of single arbitrary
oligonucleotide primer in a low stringency PCR for simultaneous amplification of several discrete DNA fragments, referred as RAPD and arbitrarily primed PCR (AP-PCR) respectively. The difference between these techniques mainly attributed to number of nucleotides in primer sequence. RAPD is a dominant multiloci marker system and allow the detection of polymorphism without any prior knowledge of the nucleotide sequence (Hunt and Page, 1992). In this technique, a single short oligonucleotide primer is allowed to bind at many loci and amplify random sequences from the DNA template. The primers are generally random sequence of 9-10 nucleotides long and generate an average of 2-10 bands. However this technique is often criticized for its lack of reproducibility over time as well as between laboratories (Jones et al., 1997). The RAPD markers are well exploited in tree genetics and breeding by various workers, in Olive (Gemas et al., 2004), Morus sp (Awasthi et al., 2004), Neem (Azadirachta indica) (Deshwal et al., 2005; Bhatt et al., 2011), Eucalyptus sp (Grattapaglia et al., 1994), Larch (Arcade et al., 2000) etc.

**Amplification fragment length polymorphism (AFLP™)**

AFLP is a new multiplex PCR based method which combines the reliability of the RFLP and ease of the PCR. This approach, given by Zabeau and Vos (1993), is based on selective PCR amplification of restriction fragments from a total double digest of genomic DNA under high stringency conditions and can be used for DNA of any origin or complexity. The fingerprints are produced without prior knowledge of the genome using a limited set of generic primers. AFLP technique is reliable since stringent reaction conditions are used for primer annealing. The advantage of AFLP analysis is that it allows the reliable identification of over 50 loci in a single reaction. Though the AFLP analysis is not popular as much as other PCR based markers like RAPD, it was employed in genetic studies of tree species like Larch (Arcade et al., 2000) and Neem (Singh et al., 2002).

**Minisatellites or variable number tandem repeats (VNTRs)**

Eukaryotic genomes contain some regions of highly repetitive DNA, which have higher mutation rates (estimated mutation rate of minisatellite loci was 2 X 10⁻³ per meiosis) than single copy DNA. Minisatellite markers were first discovered in human beings by Wyman and White (1980) and Jeffreys et al. (1985a, b).

Minisatellite repeats are ranges from 10 to 100 base pairs, occurring in tandem arrays -of up to 1000 distributed throughout the genome (Jeffreys et al., 1985a). There may be up to 1000 minisatellite loci in human genome (Jarman and Wells, 1989; Wahls et al., 1990). Minisatellite markers have been used for a variety of purposes in several tree species including fingerprinting in rubber tree (Hevea brasiliensis) (Besse et al., 1993) and diversity analysis and linkage map construction in pedunculate oak (Quercus robur L.) (Barreneche et al., 1998).

**Simple sequence repeats (SSR) or microsatellites**

Simple sequence repeats also termed as microsatellites, sequence-tagged microsatellite sites (STMS) or short tandem repeats (STRs) are tandem repeats of short nucleotide sequence motifs (di-, tri-, tetra - and penta-nucleotide tandem repeats) that can detect high levels of polymorphism at multiple loci. These are an ideal genetic markers, would provide the specificity and the rapidity of PCR with more information per locus examined. Jeffreys et al. (1985a) suggested that PCR primers from the conserved flanking regions of VNTR loci be developed, thereby allowing PCR amplification of the entire VNTR locus. The resulting PCR products would vary in size according to the repeated DNA units in the VNTR alleles present. This approach was extended to a different type of VNTR locus at which the repetitive DNA units are only 2 to 5 base pairs in length rather than repeat units in the range of 11 to 60 base pairs in length. These workers suggested that high levels of polymorphism existed as a result of variation in the number of such short repeat units as a result they are found to be abundant and evenly spread throughout eukaryotic genomes.

SSRs have been characterized in many tree genomes such as Eucalyptus sps (Brondani et al., 1998), olive (Olea europaea L.) (Rallo et al., 2000), rubber tree (Hevea brasiliensis) (Feng et al., 2009) etc. Because of their merits such as abundance, even genomic distribution and high level of polymorphism, SSRs are considered to be the markers of choice for DNA fingerprinting and genetic analysis studies in various plant species and other living organisms.

**Inter-simple sequence repeats (ISSR)**

ISSR-PCR (Zietkiewiez et al., 1994) is another method which relies on two primers for PCR. It involves the amplification of regions between adjacent inversely oriented microsatellites, containing a primer binding sites. This can be undertaken for any species that contains sufficient number and has genome wide coverage of SSR motifs. It does not require any genomic sequence information. This technique amplifies large number of DNA fragments per reaction, representing multiple loci across the genome and therefore is an ideal method for fingerprinting varieties. ISSRs have been successfully used to estimate the extent of genetic diversity at inter and intra-specific level across the tree species which include Larch (Arcade et al., 2000), Jatropha (Senthil Kumar, 2009), Pongamia (Rout et al., 2009) etc.

These are multilocus markers, more robust than RAPD and are usually dominant in nature (Reddy et al., 2002). Rakoczy-Trojanowska and Bolibok (2004) reported that ISSRs were suitable for phylogenetic studies, genetic diversity evaluation and cultivar identification.

**Single nucleotide polymorphisms (SNPs)**

Single Nucleotide Polymorphisms (SNPs) are the new generation DNA markers, describes polymorphisms caused by point mutations that give rise to different alleles, containing alternative bases at a given nucleotide position within a locus, most commonly used for the assessment of polymorphism in the genome. These are considered as potentially the best type of genetic marker for genome mapping as well as fingerprinting because of their abundance in the genome, co-dominant inheritance, stability and simplicity compared to microsatellite markers (Rafalski, 2002 and Nasu et al., 2002) and their potential association with disease and adaptive traits (Gonzalez-Martinez et al., 2006). Development and application of SNP markers is gaining momentum especially after the release of genome sequence information in various plant species. The association analysis and SNP genotyping analysis...
was successfully utilized in species like Pinus (Neale, 2007) and Vitis (Vezzulli et al., 2008).

Organelle based DNA markers

Besides nucleus of the cell, the mitochondria and chloroplast possess the nucleic acids. Generally the chloroplasts are inherited maternally in angiosperms but in conifers it is paternally inherited to the next generations. Additionally in Pinaceae, Taxodiaceae and Cuptassaceae families various workers reported the maternal inheritance of mitochondria and paternal in heritance of chloroplast (Ali et al., 1991; Neale et al., 1986; Neale et al., 1989; Neale and Sederoff, 1989 and Szmidt and Hallgren, 1987). The paternal inheritance of chloroplast can be well utilized in studies of pollen dispersal. For exploitation of polymorphism at chloroplast DNA level chloroplast can be well utilized in studies of pollen dispersal. For exploitation of polymorphism at chloroplast DNA level the SSRs and RFLP-PCR markers are employed by various researchers viz., Burg et al. (1993) in Oak, Jack et al. (1995) in Oil palm, Lashermes et al. (1996) in coffee, Provan et al. (1999, 2001) in Torrey pine and Schlogl et al. (2007) in Araucaria sp.

Genotyping by sequencing

Although abundant diversity is a challenge to assays that rely on scoring fixed positions, it is advantageous to direct sequencing approaches because sequencing efficiency for genotyping scales directly with genetic diversity. Next-generation sequencing (NGS) technologies have been recently used for whole genome sequencing and re-sequencing for exploring within-species diversity and performing genome-wide association studies in various plant species. Genotyping by sequencing (GBS) is a simple highly-multiplexed system for constructing reduced representation libraries for the Illumina next-generation sequencing platform developed in the Buckler lab by Rob Elshire. It generates large numbers of single nucleotide polymorphisms (SNPs) for use in genetic analysis. Future application of GBS to breeding, conservation and global species and population surveys may allow plant breeders to conduct genomic selection on a novel germplasm or species without first having to develop any prior molecular tools, or conservation biologists to determine population structure without prior knowledge of the genome or diversity in the species (Elshire et al., 2011). At present there are various advanced sequence platform (SoLi D, Helicos and Nanopore sequencing) are available with high throughput and accuracy for genotyping.

CONCLUSION

The availability of various genotyping tools paved their way in analysis of genetic diversity at molecular level in tree species and this provides greater insight for quantification of existing diversity and their phylogenetic relationships with respect to evolutionary clock. This technology has been employed in various tree species, the results and accuracy are encouraging in breeding programme but there is a need to extend these technologies to tap the potential variations in underutilized tree species for economically important traits. Further they can be utilized for construction of genetic linkage maps, establishment of marker trait association and isolation genes for economically important traits are the future thrust areas in the tree breeding programmes.

REFERENCES


