Molecular Markers in Assessment of Genetic Diversity

David L. Biate and *Arvind K. Misra*

ABSTRACT

Genetic diversity is defined as the variation of genes within species. It represents the fundamental entity of biodiversity. Over the years, several molecular markers have been developed and utilized to assess genetic diversity, conserve and manage genetic resources. This paper is a brief review of the most commonly used molecular markers in plant genetic diversity studies and an attempt to broadly outline their strength and limitations in the field of plant genetic resource management.

Key words: *Molecular markers, genetic diversity, conservation, management.*

Introduction

Continued alteration of the environment by man has resulted in unprecedented loss of biodiversity. Many species have become extinct or are near extinction due to destruction and fragmentation of their habitats. Consequent erosion of genetic diversity leads to reduced resilience to environmental changes and altered ecosystem processes. The evolutionary potential and survivability of a species is greatly influenced by the extent and distribution of genetic variation present within species. Therefore, for developing effective conservation strategies, assessment of genetic diversity is of fundamental importance (Holsinger and Gottlieb, 1991; Newton *et al.*, 1999).

Rapid advances in the field of molecular genetics over the last few years have opened new avenues for studying plant diversity

at the level of single base pair differences giving us better insight into their underlying evolutionary and adaptation processes and thus enhancing our capabilities for managing plant genetic diversity (Karp, 2002).

Molecular genetic techniques such as RFLP (restriction fragment length polymorphism), PCR (polymerase chain reaction) based RAPD (randomly amplified polymorphic DNA), AFLP (amplified fragment length polymorphism), PCR-RFLPs, microsatellites (SSRs) has been effectively used in both *ex situ* and *in situ* plant genetic resource management and conservation. Information derived from the use of these techniques has been successfully applied by workers involved in the conservation of plant genetic resources to identify a series of conservation units, which include species, evolutionarily significant units (ESUs), management units (MUs), action units (AUs), and family nets (FNs), thereby, developing appropriate and effective strategies for their conservation and management (Wan *et al.*, 2004). Several of these molecular markers have been widely used in studies of genetic diversity and variation in some forest trees (Butcher *et al.*, 2000; Wang and Szmidt, 1994; Szmidt *et al.*, 1996; Chantragoon *et al.*, 1997; Chauhan and Misra, 2002; Maguire *et al.*, 2002). Further, application of these techniques has helped genebanks curators in identifying duplicates, evolving better sampling strategies, multiplication, management and the development of core collections (Karp *et al.*, 1997; Bataillon *et al.*, 1996; van Hintum *et al.*, 2000; van Treuren *et al.*, 2001).

Molecular Markers

Genetic markers that reveal polymorphisms at the DNA level are known as molecular markers. Molecular markers are increasingly becoming the preferred tools over morphological and protein (biochemical) markers for scientists working in the field of plant genetic resource conservation and management. This in part is because of their ability to detect genetic diversity at a resolution much higher than the other methods. Further, molecular markers are not subjected to environmental influences; tests can be carried out at any time during plant development and that, they exist in potentially unlimited numbers, covering the entire genome.

Restriction/Hybridization Based Markers

Restriction fragment length polymorphism (RFLP)

In restriction fragment length polymorphism (RFLP) the total DNA isolated is digested using an appropriate restriction enzyme. Each restriction enzyme, under appropriate conditions, cuts different DNA molecules into fragments of different lengths. The generated fragments are separated by gel electrophoresis and transferred to a membrane by the process termed Southern Blotting. In this method, the gel is first denatured in a basic solution and placed in a tray. A porous nylon or nitrocellulose membrane is laid over the gel. All the DNA restriction fragments in the gel get transferred by capillary action to the membrane. All fragments retain the same pattern on the membrane as on the gel. The membrane is then incubated with labeled DNA probe. Restriction fragments on the membrane that are complementary to the probe form labeled duplexes. The hybridized fragments are then captured on X-ray film and analysed. Alternatively, dried gel may also be used directly for hybridization and detection.

This technique is highly reproducible, simple to use and has discriminating power upto individual level. The markers generated are co-dominantly inherited and are, therefore useful in estimating heterozygosity. However, this technique is time consuming, difficult to automate and requires large amount of DNA (de Vicente and Fulton, 2003)

Variable Number of Tandem Repeats (VNTRs)

Most eukaryotic genomes contain repeats of short sequences (2–15 bps), dispersed through out the genome (also called as Variable Number of Tandem Repeats — VNTRs). These sequences include minisatellites and microsatellites (Simple Sequence Repeats — SSRs or Short Tandem Repeats — STRs). These sequences can be detected either by using simple hybridization protocols or PCR. When restricted DNA is hybridized with these probes, multilocus patterns are obtained which can resolve variation at the population and individual levels (Beyermen *et al.*, 1992).

This technique is highly reproducible, generates polymorphic patterns, requires very little DNA and can be easily automated. However, they are expensive and VTNRs are difficult to detect.

Polymerase Chain Reaction (PCR) based markers

Polymerase Chain Reaction is a technique that amplifies a specific region of DNA *in vitro* using a thermo stable polymerase enzyme. PCR is quick, sensitive and versatile and has evolved into an indispensable tool over the last few decades for biodiversity assessment and management. Some of the widely used PCR-based molecular marker techniques in genetic diversity studies are described below.

- PCR with arbitrary primers
- Arbitrary Primed PCR (AP-PCR)
- Random Amplified Polymorphic DNA (RAPD)
- DNA Amplification Fingerprinting (DAF)

These techniques make use of short single 'arbitrary' PCR primer for amplifying unknown stretch of DNA in a PCR reaction. The DNA fragments generated are then separated and detected by gel electrophoresis. The three techniques differ from each other primarily in primer length (AP-PCR ~20bp, RAPD ~10bp, and DAF ~6–8 bp respectively) and conditions of stringency.

These techniques require small amounts of DNA and are quick, simple and efficient. Arbitrary primers are easily purchased, with no need for initial genetic or genomic information. However, non-reproducibility of results and unwanted amplification of contaminating DNAs make these techniques less reliable.

Site-targeted PCR

- Sequence Tagged Sites (STSs)
- Sequence Characterized Amplified Regions (SCARs) or Sequence Tagged Amplified Region (STAR)

Unlike PCR with arbitrary primers, in Sequence Tagged Sites (STSs) a pair of PCR primers designed for amplification of specific DNA region/locus is used. The basic principle of STSs and SCARs is similar; their only difference is that, in SCARs, PCR primer is designed based on the sequence information of a band generated through RAPD experiment. Genus specific PCR primers were designed for early and unambiguous identification of *Frankia*, both in isolated form as well as *in situ* (Simonet *et al.*, 1991). Many workers in the field of actinorhizal symbioses have since used these primers.

STS and SCARs have a particular advantage over RAPDs in that they are codominant, that is, they can distinguish between homozygotes and heterozygotes. They are more reliable reproducible, because they use longer primer sequences. However, their only drawback is that prior sequence information is required for this approach.

Expressed Sequence Tags (ESTs)

ESTs are based on DNA sequences from cDNA library of a plant. This technique requires prior sequence information and generation of cDNA clones, which are used to design primers that can detect unique, expressed region of the genome. They are useful since they represent coding regions of DNA and are good for mapping and detection of quantitative trait loci (QTLs) linked genes. However, designing useful primers can be expensive.

Cleaved Amplified Polymorphic Sequence (CAPS)

In CAPS or PCR-RFLP, amplification is carried out using specific sequence based PCR primers designed using information available in databank of genomic or cDNA sequences or cloned RAPD bands. The amplified products are then digested with specific restriction enzymes (usually a four base cutters) capable of detecting accumulated variations in the DNA sequence. The profile obtained on a gel can then be analysed using appropriate software. Using this approach, PCR-RFLProfile were developed for selection of superior genotypes of *Alnus nepalensis* (D.Don) that supports higher nitrogenase activity (Chauhan and Misra, 2002) and for distinguishing different strains of *Frankia,* the micro-symbionts of actinorhizal plants (Varghese *et al.*, 2003).

CAPS is a robust assay because specific long primers can be designed, allows the detection of polymorphisms that may be undetectable from PCR products and markers are co-dominant. However, designing specific primers can be laborious and expensive.

Amplified Fragment Length Polymorphism (AFLP)

AFLP is the combination of restriction digestion and PCR techniques. In AFLP, DNA is first digested with a frequent cutter restriction enzyme and one rare cutter restriction enzyme. Synthetic oligonucle-

otide adaptors for each restriction site are ligated to the digested DNA. Primers are then designed to anneal to the oligonucleotide adaptors, restriction sites and few bases of the original DNA fragment. Normally two separate rounds of PCR are carried out. In the first round of PCR (pre-selective) amplification is carried using primers complimentary to the adaptors, restriction sites and a single base of the template DNA. In the second round, usually pre-selective amplification products undergo another PCR run with selective primers complimentary to two additional bases of the template DNA. Denaturing polyacrylamide gel or capillary gel electrophoresis separates the resulting fragments.

This approach is highly reproducible, allows a quick scan of the whole genome for polymorphism, no prior sequence information or probe generation is needed. On the flip side, AFLP are technically demanding and markers generated display dominance. For all their advantages and disadvantages, comparative studies have shown that AFLP is one of the best techniques for detecting polymorphisms.

Internal Transcribed Spacers (ITSs) and Inter Genic Spacers (IGSs)

ESTs utilise coding regions of DNA. Many a times, scientists target unique genes for the purpose. It is here that the problem arises if no sequence information is available for the target region. Not only this, coding regions tolerate fewer mutations because of the selection pressure. The ITSs and IGSs are flanked by coding regions of DNA. While coding regions tend to be more conserved, the ITSs and IGSs can accumulate more variations. Utilising the sequence information of the flanking genes, the unknown regions of ITSs and IGSs can be amplified. Using primers located in the 16S and 23S rRNA genes, we have successfully amplified the *rrn* operon ITS (Chauhan and Misra, 2002; Varghese *et al.*, 2002). The ribosomal RNA genes are highly conserved across organisms; therefore it is possible to design universal primers. We have primers that can amplify the 16S rRNA genes of almost all prokaryotes, as well as that of the eukaryotic mitochondria and chloroplasts (Simonet *et al.*, 1991; Bosco *et al.*, 1992). When *rrn* ITS is amplified, the size of the amplified fragment varies according to the size of the ITS. This size variation can be easily detected by gel electrophoresis.

DNA Sequencing

All markers are derived from polymorphisms in the nucleotide sequence of a particular DNA segment. DNA sequencing represents the most fundamental measure of diversity as it can detect these polymorphisms from within the nucleotide sequences themselves. There are two methods for DNA sequencing — Maxam-Gilbert and Sanger (dideoxy sequencing or chain termination) method. Of the two, the Sanger method is more widely used, as it is easier to automate. Basically in DNA sequencing, the DNA is first broken up into fragments. Each fragment generated is then subsequently cloned and each cloned fragment is used as template to generate a set of fragments differing in length from each other by a single base. These fragments are then separated by gel electrophoresis. In the process known as 'base calling' the base at the end of each fragment is identified. The original sequence of As, Ts, Cs and Gs is recreated for each short piece generated in the first step. The short sequences are then assembled into one long sequence. We have so far contributed 26 nucleotides sequences to the global data bases (GenBank and EMBL).

DNA sequencing results are highly reproducible and informative. However, DNA sequencing is an expensive affair and also technically demanding, making this technology out of the reach of many researchers.

DNA Chips

The underlying principle of DNA chips or Microarray is base pairing or hybridization of mRNA molecules with complimentary DNA sequences from which it originated. DNA chip is first prepared by immobilizing target DNAs onto glass slide or nylon membrane. The DNA chip is then incubated with a hybridization solution containing fluorescent-labeled cDNAs. The cDNAs are detected using laser technology and the data is analyzed using computational methods. Microarray technique allows scientists to generate expression profiles of large number of genes in a single experiment. They are also useful for identification of gene mutation sequences and detection of large number of specific DNA markers. The technique is however limited by its cost factor and the high level of technical expertise required.

Conclusion

Molecular markers differ in the kind and quality of information they provide, in their level of discrimination, in their reproducibility and financial requirements. Within all these constraints, it is therefore important that the right technique be chosen for a given study, to maximize their potential gains. When choosing a particular technique, one should have a clear knowledge of the kind of questions to be addressed and the objectives of the study, the level of diversity information needed, expected level of variation, accessibility of probes and primers sets, the time constraints of the specific project and the level of operational and financial investment available (Karp *et al.*, 1997). It is also important to consider whether a single technique can provide all the required information, if not, one can opt for combination of different techniques.

Endless potentials exist for exploration of plant genetic diversity, phylogeny and evolution, in the 21st century, especially due to the fast advancing field of molecular genetics where new techniques are being continuously developed and existing techniques improved. While we scale the dizzying heights of these technologies, it would be illusionary to believe that molecular genetics alone can provide answers to all the questions and vice versa. Success of future conservation strategies will largely depend on our ability to network different biological disciplines such as ecology and taxonomy together which will provide us better understanding of the diversity available and the ways to enhance our management capabilities.

References

Bataillon, M.T., L.D. Jacques and D.J. Schoen. 1996. Neutral genetic markers and conservation genetics: Simulated germplasm collections. *Genetics* 144:409–417.

Beyermann, B., P. Nürnberg, A. Weihe, M. Meixner, J.T. Epplen and T. Bvrner. 1992. Fingerprinting plant genomes with oligo nucleotide probes specific for simple repetitive DNA sequences. *Applied and Environmental Microbiology* 83:691–694.

Bosco, M., M.P. Fernandez, P. Simonet, R. Materassi and P. Normand. 1992. Evidence that some *Frankia* sp. strains are able to cross boundaries between *Alnus* and *Elaeagnus* host specificity groups. *Applied and Environmental Microbiology* 58:1569–1576.

Butcher, P.A., G.F. Moran and R. Bell. 2000. Genetic linkage mapping in *Acacia mangium*. Evaluation of restriction endonucleases, inheritance of RFLP loci

and their conservation across species. *Theoretical and Applied Genetics* **100**:576–583.

Changtragoon, S., A.E. Szmidt and X-R. Wang. 1997. Genetic diversity of rattan in Thailand: I-Preliminary results. In: *International Plant Genetic Resources Institute Workshop: Molecular Techniques in Plant Genetic Resources*. pp. 39–43. Rome, 9–11 October 1995.

Chauhan, V.S. and A.K. Misra. 2002. Development of molecular markers for screening of *Alnus nepalensis* (D.Don) genotypes for the nitrogenase activity of actinorhizal root nodules. *Moecular Genetics Genomics* **267**:303–312.

de Vicente, M.C. and T. Fulton. 2003. Using molecular marker technology in studies on plant genetic diversity. Illus. Nelly Giraldo. International Plant Genetic Resources Institute, Rome, Italy and Institute for Genetic Diversity, Ithaca, New York, USA.

Holsinger, K.E. and L.D. Gottlieb. 1991. Conservation of rare and endangered plants: Principles and prospects. In: *Genetics and conservation of rare plants*, D.A. Falk and K.E. Holsinger (eds.). Oxford University Press. pp. 195–230.

Karp, A., S. Kresovich, K.V. Bhat, W.G. Ayad and T. Hodgkin. 1997. Molecular tools in plant genetic resources conservation: A guide to the technologies. *International Plant Genetic Resources Institute: Technical Bulletin* No. 2.

Karp, A. 2002. The New Genetic Era: Will it Help us in Managing Genetic Diversity? In: *Managing Plant Genetic Diversity*, J.M.M. Engels, V.R. Rao, A.H.D. Brown and M.T. Jackson. (eds.). *International Plant Genetic Resources Institute*. pp. 43–56.

Maguire, T.L., R. Peakall and P. Saenger. 2002. Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (Avicenniaceae) detected by AFLPs and SSRs. *Theoretical Applied Genetics* **104**:388–398.

Newton, A.C., T.R. Allnutt, A.C.M. Gillies, A.J. Lowe and R.A. Ennos. 1999. Molecular phylogeography, intraspecific variation and the conservation of tree species. *Tree* **14(4)**:140–145.

Simonet, P., M-C. Grosjean, A.K. Misra, S. Nazaret, B. Cournoyer and P. Normand. 1991. *Frankia* genus specific characterization by polymerase chain reaction. *Applied and Environmental Microbioliology* **57**:3278–3286.

Szmidt, A.E., S. Changtragoon and X-R. Wang. 1996. Contrasting patterns of genetic diversity in two tropical pines: *Pinus kesiya* (Royle) and *Pinus merkusii* (Jungh et de Vriese). *Theoretical and Applied Genetics* **92**:436–441.

van Hintum, Th. J.L., A.H.D. Brown, C. Spillane and T. Hodgkin. 2000. Core collections of plant genetic resources. *International Plant Genetic Resources Institute: Technical Bulletin* No. 3.

van Treuren, R., L.J.M. van Soest and Th.J.L. van Hintum. 2001. Marker-assisted rationalisation of genetic resource collections: A case study in flax using AFLPs. *Theoretical and Applied Genetics* **103**:144–152.

Varghese, R., V.S. Chauhan and A.K. Misra. 2002. Hypervariable spacers regions are good sites for developing specific PCR-RFLP primers for screening actinorhizal symbionts. *Journal of Biosciences* **28**:437–442.

Wan, Qiu-Hong., Hua. Wu, Tsutomu Fujihara and Sheng-Guo Fang. 2004. Which genetic marker for which conservation genetics issue? *Electrophoresis* **25**:2165–2176.

Wang, X-R. and A.E. Szmidt. 1994. Hybridisation and chloroplast DNA in a *Pinus* species complex from Asia. *Evolution* **48**:1020–1031.