

Application of Molecular Diagnostics to Plant Genetic Resources Conservation

LI RUGANG

Biotechnology Research Center, Chinese Academy of Agricultural Sciences, Beijing 100081

(Received March 18, 1994; revised May 20, 1994)

ABSTRACT

The importance of plant genetic resources to past, present, and future crop agriculture has been recognized by scientists. Conservation of genetic resources is actually to maintain genetic diversity. The quality of any collection should not be based predominantly on size or use, but rather on how well the *ex-situ* collection represents, conserves, and makes accessible the genetic diversity of taxon. Curators have to make balance within range of limited funding between the relatively short-term needs of users and the long-term maintenance of maximum genetic diversity within minimum number of accessions. Addressing these challenges requires effective conservation and utilization of plant genetic resources. Part of the solution to this challenge may lie in the use of powerful, yet relatively simple and inexpensive molecular techniques to generate information to better organize the useful genetic variation within a collection. In this mini-review, three molecular techniques (RFLPs, AFLPs, and SSLPs) and their uses in the assessment of genetic diversity are to be described.

1 *Importance of plant genetic resources*

The importance of plant genetic resources to past, present, and future crop agriculture has been recognized by scientists. In 1970, a disease called southern corn leaf blight swept through cornfield from the Southeastern United States into the Great Plains. This epidemic costed farmers 15 percent of the corn crop that year (about 700 million bushels). The reason it happened was that nearly all the corn cultivars planted were genetically susceptible to the fungus that caused the blight. To prevent similar epidemics from recurring in later years, the seed companies and breeders reverted to hybrids that lacked the susceptibility but otherwise had the desirable genetic makeup of the affected corn.

The 1970's epidemic reminded people of how vulnerable modern agriculture has become. It reminded people of the mid-19th century Irish potato famine, which happened for similar reasons and costed not just money, but hundreds of thousands of lives. It reminded us that since our vast agricultural resources were the foundation of our property, seeds and plants were the national treasure.

2 *Major problems to plant genetic resources conservation*

Conservation of genetic resources is actually to maintain genetic diversity. Genetic

diversity is the heritable portion of observable variation and the raw materials on which natural selection has acted to create earth's vast array of organisms. People have to eat food and dress fiber which rely on intensive agriculture. Intensive agriculture benefits from genetic uniformity in crops created by artificial selection. But genetic uniformity increases the potential for crop vulnerability to new pests and stresses. Genetic diversity gives us the sustained ability to develop new plant varieties resistant to those pests, diseases, and environmental stresses.

There is a common recognition that wild ancestors and relatives of cultivated plants are the keys of genetic diversity. Maintaining old or little used varieties and races from the wild and weedy germplasms can preserve genes for disease resistance as well as other traits. But the amount of land where plants grow wild continues to shrink, and many plant species and varieties are disappearing forever. Each passing year reduces the genetic diversity of our planet at a rate far exceeding that governing the evolutionary process (Kresovich et al., 1992). In the highly populated world of our future, some of these plants may help people make difference between abundance and scarcity. Therefore, several developed countries such as USA and Germany, put a lot of energy on *ex-situ* collection of wild, weedy relatives, landraces and cultivars of selected taxon. The quality of any collection should not be based predominantly on size or use, but rather on how well the *ex-situ* collection represents, conserves, and makes accessible the genetic diversity of the taxon (Stalker et al., 1989). With the increase of collection size and the desires of more accessible and better evaluated germplasms, funding shortage becomes an important limited factor. Curators have to make balance within range of limited funding between the relatively short-term needs of users, e. g. the sources of disease resistance and long-term maintenance of maximum genetic diversity within minimum number of accessions. Addressing these challenges requires effective conservation and utilization of plant genetic resources.

Part of the solution to this challenge may lie in the use of powerful, yet relatively simple and inexpensive molecular techniques (Avisé, 1989) to generate information to better organize the useful genetic variation within a collection. The molecular techniques, closely linked with other classical methods, might be employed for the purposes of organizing and characterizing genetic diversity within a collection. Kresovich et al. (1993) listed four genetic parameters of characterizing a collection as following. 1 identity: whether the accession in the gene bank is catalogued correctly and "true to type"; 2 relationship: the degree of relatedness among individual genotypes in an accession or among accessions of a collection; 3 structure: the amount of genetic variation present and how it is partitioned among plants and accessions; and 4 location: the location of a desired gene/gene complex in a specific accession and also the physical location of a desired DNA sequence on a particular chromosome in an individual plant.

3 *Molecular diagnostic techniques and use in assessment of plant genetic diversity*

The best measure of genetic diversity in the context of plant genetic resource conservation is the number of alleles per locus at qualitative marker loci such those governing simple inherited morphological traits, or enzyme, or DNA markers. This parameter

provides a direct measure of the variation at each locus within and between populations. The morphology-based phenotypic assessment of genetic diversity within populations and species is influenced by a multiplicity of factors including life cycle, generation time, outcrosser or inbreeder, means of pollen dispersal, geographic range and ecological niche. Therefore, morphology can not exactly reflect the inherent genetic variation and the relationships among accessions. In late 1960s, allozyme as a genetic marker was applied to studies of genetic variation and correlation this measure of diversity with species characteristics. The advantage in allozyme analysis is that large numbers of samples can be processed rapidly. Allozymes have since provided the most abundant source of information about genetic diversity in natural populations. The frequency of alleles within population can be deduced. The genetic diversity measured by allozyme analysis at the population level accurately reflects diversity at the species level (Hamrich et al., 1990). However, the others have pointed out that electrophoretic analysis of proteins suffers certain limitations, e. g. lack of diversity within some species, less number of loci resolved, posttranslation modification of enzyme and poorer resolution than molecular techniques (Hillis et al., 1990).

The science and technology associated with Human Genome Initiative (HGI) research provide unprecedented impact, in both the short- and long-term, on plant genetic resources conservation (Kresovich et al., 1993). Since DNA is the medium of heredity, DNA directly reflects the relatedness within a collection. The characterization of individual identity at the DNA level has been referred to as 'fingerprinting', 'profiling', 'typing', 'genotyping' or 'identity testing'. The concept of DNA fingerprinting was founded on the observation by Wyman and White (1980) of a polymorphic DNA locus characterized by a number of variable length restriction fragments. Three major fingerprinting assays have since been developed.

3.1 Restriction Fragments Length Polymorphisms (RFLPs)

RFLP relies on the detection of what is termed restriction fragment length polymorphism and is the conventional DNA fingerprinting. The basis of RFLP is based on widespread polymorphisms for the location of restriction sites among individuals within species. Two types of restriction polymorphism are generally observed. The most common one is the result of the creation and abolishment of a restriction site and is doubtless the result of single base changes; the second one is deletion and/or insertion. RFLP analysis involves DNA cleavage by restriction enzyme, electrophoresis of resulting fragments, southern transfer of separated fragments to a membrane support, radioactive labelling of suitable probes, hybridization of probes to membrane-supported fragments and print detection as a banding pattern on X-ray film. Variant banding patterns, or fingerprints, on a X-ray film are the result of polymorphic fragments detected by a particular probe. The probe DNA may hybridize to multiple tandem-repetitive or hypervariable minisatellites (Jeffreys et al., 1985a, b; Vassart et al., 1987) and produce complex fingerprint banding patterns. Alternatively, probes may be locus-specific for individual hypervariable loci (Nakamura et al., 1987) which produce simpler patterns by detecting alleles from single or even multiple loci.

Engles (1981) discussed the problem of estimating genetic diversity from restriction data and clearly showed an application of his equations for estimating the degree of genetic divergence. If a genetic map was constructed based on RFLP markers uniformly

distributed over the genome, then employment of a set of probes to evaluate genetic diversity within collection would minimize sampling error. RFLPs have since been used as genetic markers, and extensive RFLP marker linkage maps have been generated in several important crops and fruit species such as tomato, maize, lettuce, potato and brassica. RFLPs have widely applied in assessment of plant genetic diversity (Helentjaris et al., 1985; Figdore et al., 1988). Several reviews have been published about the application of RFLPs in plant breeding and assessment of plant genetic diversity (Beckmann et al., 1986; Tanksley et al., 1989). The advantage of RFLPs is that the unlimited number of loci can be examined, and the large number of alleles is likely to be found for most loci. However, RFLP analysis suffers from dangerous radioactive operation and laborious work and is done only in well-equipped laboratories. Therefore, application of RFLPs in assessment of genetic diversity is operationally more difficult, time-consuming and expensive.

3.2 Amplification Fragments Length Polymorphisms (AFLPs)

The polymerase chain reaction (PCR) has revolutionized the field of molecular biology. PCR is an in vitro methods for enzymatic synthesis of specific DNA sequences. Depending on principle of PCR, three research laboratories independently developed three strategies to arbitrarily amplify genomic DNA regions. They were termed random amplified polymorphic DNA (RAPD) (Williams et al., 1990), arbitrarily primed PCR (AP-PCR) (Welsh et al., 1990), and DNA amplification fingerprinting (DAF) (Caetano-Anolles et al., 1991), respectively. These strategies do not require prior knowledge of DNA sequence, but rather use one or more arbitrary oligonucleotide primers to target specific but unknown sites in the genome, many of which are polymorphic. Therefore, these techniques were referred to be multiple arbitrary amplicon profiling (MAAP) (Caetano-Anolles et al., 1992) and produce amplification fragment length polymorphism (AFLPs).

MAAP techniques target multiple and arbitrary amplicons but differ in primer length, amplification stringency, and procedure used to resolve DNA patterns (Williams et al., 1990; Welsh et al., 1990; Caetano-Anolles et al., 1991). Although exact genetic basis revealed by MAAP techniques are still not clear, amplification fragment length polymorphisms (AFLPs) detected can be used easily as genetic markers and show wide applications. Recently, several reviews have been published about the application of AFLPs (Williams et al., 1993; Newbury et al., 1993; Tingey et al., 1993; Caetano-Anolles, 1993). AFLPs have been used in genotypical fingerprinting and identification, determining genetic relatedness and genetic variation (Williams et al., 1993), reconstructing phylogeny and performing genomic mapping and linkage analysis.

So far, most applications of MAAP techniques are reported using RAPD markers. The technical simplicity, low cost, nonisotope operation and the frequency of identification of polymorphisms of RAPD analysis make it viewed as the current state-of-the-art for molecular diagnostics in plant gene banks. However, as a dominant marker, AFLPs can not tell us the number of alleles per locus which is important to assess genetic diversity of a collection. Also, because of the enormous amplification of very low initial quantities of DNA sequence, all PCR-based are stated to be prone to artifacts caused by contamination of the reaction mixture by foreign DNA (Newbury et al., 1993).

3.3 Simple Sequence Length Polymorphisms (SSLPs)

A substantial portion of the genome of higher eukaryotes consists of interspersed noncoding repetitive DNA element. These DNA sequences with "variable numbers of tandem repeats", termed VNTRs (Nakamura et al., 1987), include those loci called hypervariable regions (HVRs), minisatellite (Jefferys et al., 1985a) and microsatellite (Litt et al., 1989; Weber et al., 1989; Tautz 1989). The size, number and pattern of repeat unit exhibit high degree of polymorphism among unrelated individual genomes. Minisatellites consist of tens to hundreds of repeat unit which has sequence of 11~60 base pairs, and produce regions up to 20kb in length. Microsatellites consist of around 10~50 copies of motifs from 1 to 6 bp that generate repeating regions less than 100bp. The VNTR loci are flanked by conserved endonuclease restriction sites. Thus, the length of the restriction fragment produced by this type of genetic locus is proportional to the number of core units contained.

Microsatellite (Litt et al., 1989), also termed simple sequence repeat (SSR) by Jacob et al. (1991), short tandem repeat by Edwards et al. (1991) was first demonstrated by human geneticists in 1989. Microsatellite regions are highly interspersed throughout eukaryotic genome. The best known of these DNA sequences is the repeat (dT-dG)_n, where n=10~60, that is present at over 50 000 loci in humans (Litt et al., 1989). According to Edwards et al. (1991), the combined frequency in the human genome of tri- and tetrameric SSR with n=7 or greater is estimated to be 400 000 or one SSR per 10 kbp. In a (TG)_n repeat in the human cardiac muscle actin gene locus, Litt et al., (1989) detected 12 length variant (alleles) in only 37 individuals. Therefore, these tandemly repetitive sequences represent a rich source of highly polymorphic, multiallelic, stable, widely dispersed genetic marker (Hudson et al., 1992). Hybridization of restriction-digested genomic DNA against a polymorphic minisatellite "core" sequence (like M13, 33.6 or 33.15) or a simple repetitive sequence motif (like (CA)₈, (GATA)₄ and (GTG)₅) often results in the simultaneous detection of several extremely variable loci and creates "DNA fingerprinting" in human, fungi, and plant (Vassart et al., 1987; Rogstad et al., 1988; Weising et al., 1991; Thomas et al., 1993; Poulsen et al., 1993).

Jeffreys et al. (1988) suggested that the highly informative nature of minisatellite loci can be combined with specificity and rapidity of polymerase chain reaction (PCR) technology. Primers to the conserved flanking regions of minisatellite loci were developed allowing amplification of the entire locus. Resulting PCR products possess electrophoretic mobilities which differ according to the number of core units in the minisatellite allele(s) present. This approach was extended to microsatellite in 1989 (Litt et al., 1989; Weber et al., 1989; Tautz, 1989). The DNA sequences flanking microsatellite are conserved, allowing the selection of PCR primers that will amplify the intervening microsatellite. As initially reported (Litt et al., 1989; Weber et al., 1989; Tautz 1989), the PCR reaction includes only ³²P-labeled nucleotide or one or two ³²P-labeled primers to allow visualization of amplification products via autoradiography after electrophoresis on a standard sequencing gel. Variation in PCR product length is a function of the number of SSR units. This type of genetic marker has been referred to as a simple sequence length polymorphism (SSLP, Tautz 1989) or sequence-tagged microsatellite site (STMS, Bechmann et al., 1990). SSLP can be placed on genetic maps in relation to other SSLP, RFLP, and phenotypic markers. Fluorescence-based automated DNA de-

tection system (Ziegle et al., 1992) greatly increases sensitivity and simplifies use and make SSLP analysis be more readily obvious. Although minisatellite DNA is also highly polymorphic, it is less common than microsatellite and has larger sequence motifs making it less amenable to RCR analysis. The high frequency, high level of allelic variation and uniform distribution of SSR on human genome, combined with PCR amplification and automated detection system make SSLP an ideal genetic marker for linkage analysis and genomic fingerprinting. Transfer of SSLP genetic markers and associated techniques to plant gene banks will improve plant gene bank operations (Kresovich et al., 1993).

The knowledge of microsatellite DNA in plants is increasing. Several reports (Cregan, 1992; Akkaya et al., 1992; Poulsen et al., 1993; Morgante et al., 1993; Zhao et al., 1993; Thomas et al., 1993) have supported the occurrence of microsatellite DNA in plants. The abundance and polymorphism of microsatellite DNA were detected in numerous plant species using oligonucleotides of simple repetitive sequences as probes (Poulsen et al., 1993; Thomas et al., 1993). The studies based on database searches (Cregan, 1992; Morgante et al., 1993) and actual screening of genomic libraries of soybean (Akkaya et al., 1992), rice (Zhao et al., 1993), and grapevine (Thomas et al., 1993) indicated the findings similar to those noted previously with human. Up to 13 alleles in (GA) 22 repeat were detected among 26 *Vitis vinifera* cultivars (Thomas et al., 1993). SSLP markers revealed the high heterozygosity (69%~88%) within individual grapevine cultivars and high genetic variation between cultivars (Thomas et al., 1993). Codominant segregation of SSLP markers was documented in soybean (Akkaya et al., 1992) and rice (Zhao et al., 1993). Based on the published DNA sequence data in 34 plant species, microsatellite DNAs have combined frequency of one every 50 kb, and unlike AC prevailed in animals, AT repeats were by far the most frequently observed class of dinucleotide microsatellite (Morgante et al., 1993). Therefore, SSLP markers are ideal genetic markers for construction of plant genetic maps, filling gaps of existing genetic maps, identification of markers linked to disease and pest resistance genes, and genomic DNA fingerprinting. Also, once identified and published, SSLP markers can be shared the use among entire community.

REFERENCES

- Awise J C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. *TREE*, 4: 279~281
- Akkaya M S, A A Bhagwat, P B Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* 132: 1131~1139
- Beckmann J S, M Soller. 1986. Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica*, 35: 111~124
- Beckmann J S, M Soller. 1990. Toward a unified approach to the genetic mapping of eukaryotes based on sequence-tagged microsatellite sites. *Bio/Technology*, 8: 930~932
- Caetano-Anolles G, B J Bassam, P M Gresshoff. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technology*, 9: 553~557
- Caetano-Anolles G, B J Bassam, P M Gresshoff. 1992. DNA fingerprinting: MAAPing out a RAPD redefinition? *Bio/Technology*, 10: 937
- Caetano-Anolles G. 1993. Amplifying DNA with arbitrary oligonucleotide primers. *PCR Methods and Applications*, 3: 85~94

- Cregan P B. 1992. Simple sequence repeat DNA length polymorphisms. *Probe*, **2**(1):18~22
- Edwards A, A Civitello, H A Hammond, C T Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, **49**:746~756
- Engels W R. 1981. Estimating genetic divergence and genetic variability with restriction endonucleases. *Proc. Natl. Acad. Sci. USA*, **78**:6329~6333
- Figdors S S, W C Kennard, K M Song, M K Slocum, T C Osborn. 1988. Assessment of the degree of restriction fragment polymorphism in *Brassica*. *Theor. Appl. Genet.*, **75**:833~840
- Hamrick J L, M J Godt. 1990. Allozyme diversity in plant species. In: Brown A H D, M T Clegg, A L Kahler, B S Weir (eds.), *Plant population genetics, breeding, and genetic resources*. Sinauer Associates, Inc. Sunderland:43~63
- Helenjaris T, G King, M Slocum, C Siedenstrang, S Wegman. 1985. Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. *Plant Mol. Biol.* **5**:109~118
- Hillis D L, C Mortitz (Eds.), 1990. *Molecular Systematics*. Sunderland: Sinauer Associates, Inc.
- Hudson T J, M Engelstein, M K Cee, E C Ho, M J Rubenfield, C P Adams, D E Housman, N C Dracopoli. 1992. Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Gnomics*, **13**:622~629
- Jacob H J, K Lindpaintner, S E Lincoln, et al.. 1991. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell*, **67**:213~224
- Jaffreys A T, V Wilson, S L Thein. 1985a. Hypervariable minisatellite regions in human DNA. *Nature*, **314**:67~73
- Jeffreys A T, V Wilson, S L Thein. 1985b. Individual-specific 'fingerprints' of human DNA. *Nature*, **316**:76~79
- Jeffreys A T, V Wilson, R Neumann, J Keyte. 1988. Amplification of human minisatellites by the polymerase chain reaction: towards fingerprinting of single cells. *Nucleic Acids Res.*, **16**:10953~10971
- Kresovich S, J R McFerson. 1992. Assessment and management of plant genetic diversity: consideration of intra-and interspecific variation. *Field Crops Res.* **29**:185~204
- Kresovich S, W F Lamboy, A K Szewc-McFadden, J R McFerson, P L Forsline. 1993. Molecular diagnostics and plant genetic resources conservation. *AgBiotech News and Info.*, **5**:225~258
- Litt M, J A Luty. 1989. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of human Genetics*, **44**:397~401
- Morgante M, A M Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. *The Plant J.*, **3**(1):175~182
- Nakamura Y, M Leppert, P O Conell, et al.. 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. *Science*, **235**:1616~1622
- Newbury H J, B V Fork-Lloyd. 1993. The use of RAPD for assessing variation in plants. *Plant Growth Regulation*, **12**:43~51
- Poulsen G B, G Kahl, K Weising. 1993. Abundance and polymorphism of simple repetitive DNA sequences in *Brassica napus* L.. *Theor. Appl. Genet.*, **85**:994~1000
- Rogstad S H, J C Patton, B A Schaal. 1988. M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. *Proc. Natl. Acad. Sci. USA*, **85**:9176~9178
- Stalker H T, C Chapman (eds.), 1989. *Scientific management of germplasm: characterization, evaluation and enhancement*. IBPGR, Rome
- Tautz D. 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Res.*, **17**:6463~6471
- Tanksley S D, N D Young, A H Paterson, M W Bonierbale. 1989. RFLP mapping in plant breeding—new tools for an old science. *Bio/technology*, **7**:257~264
- Thomas M R, N S Scott. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). *Theor. Appl. Genet.*, **86**:985~990

- Tingey S V, J P Del Tufo, 1993. Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiol.*, **101**: 349~352
- Vassart G M, M Georges, R Monsieur, H Brocas, A S Lequarre, D Cristophe, 1987. A sequence in M13 phage detects hypervariable minisatellite in human and animal DNA. *Science*, **235**: 683~684
- Weber J L, P E May, 1989. A bundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, **44**: 388~396
- Weising K, J Ramser, D Kaemmer, G Kahi, J T Epplen, 1991. Oligonucleotide fingerprinting in plants and fungi. In: Burke T, G Dolf, A J Jeffreys, R Wolff (eds.), *DNA Fingerprinting: approaches and applications*, Berlin, Germany: Birkhauser Verlag 313~329
- Welsh J, M McClelland, 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.*, **18**: 7213~7218
- Williams C, D A S Clair, 1993. Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. *Genome*, **36**: 619~630
- Williams J G K, M K Hanafey, J A Rafalski, S V Tingey, 1993. Genetic analysis using RAPD markers. *Methods Enzymol.*, **218**: 704~740
- Williams J G K, A R Kubelik, K J Livak, J A Rafalski, S V Tingey, 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, **18**: 6531~6535
- Wyman A R, R White, 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA.*, **77**: 6754~6758
- Zhao X, G Kochert, 1993. Phylogenetic distribution and genetic mapping of a (GGC) microsatellite from rice (*Oryza sativa* L.). *Plant Mol. Biol.*, **21**: 607~614
- Ziegle J S, Y Su, K P Corcoran, L Nie, P E Mayrand, L B Hoff, L J McBride, M N Kronick, S R Diehl, 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics*, **14**: 1026~1031