

Molecular Analysis of Genetic Diversity: Advantages and Limitations

D. J. Fairbanks
W. R. Andersen

Abstract—Genetic diversity in plant materials used for revegetation of disturbed sites is important to allow natural selection and adaptation to occur. Several methods for estimation of genetic diversity using molecular markers are available. Protein markers, such as isozymes and seed storage proteins, have been used extensively, but their use has declined in recent years in favor of DNA markers. Use of protein markers is limited by lack of polymorphisms, number of markers available, and the need to use different detection methods with isozymes. DNA markers, including RFLP and RAPD analyses, are reliable methods that are widely used. The major limitation is the cost of labor and supplies.

Genetic diversity is an essential part of nature. It provides populations with adaptive plasticity allowing them to change genetically in response to environmental fluctuations. Without genetic variation, natural selection is ineffective. Genetic uniformity, on the other hand, is often the rule for cultivated species. Large-scale marketing typically requires a product with predictable characteristics including such factors as high yield, predictable seed or fruit size, known nutritional aspects, harvestability, and characterized disease and pest resistance. Numerous well-tested varieties are available commercially to fill specific environmental niches and to meet specific market demands. Variety registration and seed certification have been instituted to insure that varieties are genetically uniform, well characterized, and predictable. Nonetheless, genetic uniformity in cultivated species has its costs. Widespread use of a single variety or a few related varieties can select for emergence of virulent races of pathogens. Once a virulent race emerges, it can infect plants in epidemic proportions since the uniform plants are uniformly susceptible. The 1970 southern corn leaf blight epidemic in the United States, and the 1976 stem rust epidemic in Northwest Mexico's wheat crop are examples.

Plant species used in arid land revegetation straddle the fence between native species and cultivated species. On one hand, plants used in revegetation are often seeded, cultivated, irrigated, and harvested for seed much like their field-crop counterparts. On the other hand, the environments in which they are ultimately planted are much more varied than prepared fields. However, there

is little question of the need for genetic diversity. The diverse environments faced in revegetation are very different from the more uniform conditions of cultivated fields. Nor is there any market demand for uniform products to deter the use of diverse material, as is the case with cultivated plants. Consequently, in selecting and developing materials for land restoration, genetic diversity should be among our primary concerns. It serves as a buffer against uniform susceptibility to pathogens and insects. It also provides genetic plasticity, allowing natural selection to work on plants from a single seed source in a variety of environments. Genetically diverse seed sources can be self-adapting, where those plants most adapted to the environment survive and reproduce while the less adapted are eliminated.

If we accept the premise that genetic diversity in plant materials used for revegetation is essential, then there is a clear need to assess, at least to some degree, the amount of genetic diversity in the materials we use. Unfortunately, little is known about the genetic diversity of seed currently used in revegetation. Most seed is collected from wild populations that may or may not be genetically diverse. Genetic uniformity is probably a major contributing factor to the frequent problem of dieback in large-scale revegetation projects. As the demand for seed from certain species commonly used in revegetation increases, seed sources will probably tend to move toward commercial seed production and away from collection in wild populations. Also, stricter regulation of seed collection on federal land may further contribute to this shift. Already, the demand for four-wing saltbush (*Atriplex canescens*) seed is sufficient to warrant commercial seed production. Will commercial seed production reduce or increase genetic diversity? The answer probably depends on the amount of attention aimed at insuring a high degree of genetic diversity in the seed crop. Uniform characteristics, such as high seed yield, harvestability, uniform plant height, and coordinated maturity are all desirable in a commercial seed production scheme, but selection for these characteristics may compromise the overall genetic diversity that is essential for successful revegetation. Several ways of increasing genetic diversity have been proposed. Mixtures of seed collected from distant sites and hybridization of genetically diverse materials are simple, yet effective ways to increase the diversity of a seed source.

Assessment of Genetic Diversity

Morphological diversity is a good indication of genetic diversity, but at times it can be misleading. Morphological

In: Roundy, Bruce A.; McArthur, E. Durant; Haley, Jennifer S.; Mann, David K., comps. 1995. Proceedings: wildland shrub and arid land restoration symposium; 1993 October 19-21; Las Vegas, NV. Gen. Tech. Rep. INT-GTR-315. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

D. J. Fairbanks is Associate Professor and W. R. Andersen is Professor in the Department of Botany and Range Science, Brigham Young University, Provo, UT 84602.

traits are subject to environmental variation, as many a botanist has observed when plants are grown in a uniform garden. Also, morphological uniformity is not necessarily indicative of genetic uniformity, and morphological diversity for one or two traits can be due to very little genetic diversity.

Molecular markers, particularly DNA genetic markers, are valuable in that they are not subject to environmental influences, and represent underlying genetic diversity. They provide a reasonable estimation of genetic diversity provided a sufficient number of markers have been analyzed in a sufficient number of individuals to adequately represent the true genetic diversity of a population. Because of time and resource limitations, this ideal is not always possible. In fact, the greatest obstacle to using molecular markers for estimation of genetic diversity is the lack of facilities, time, and financial resources. The remainder of this paper will be devoted to a discussion of several of the various molecular marker analysis methods that are available, focusing on their advantages and limitations.

Protein Markers

Isozymes—Isozymes have been used for decades to estimate genetic diversity (Tanksley and Orton 1983). They represent native enzymes that have slightly different amino acid sequences contributing to different electrophoretic mobilities. Typically, proteins are isolated from a plant tissue, dissolved in a buffer solution, then separated electrophoretically in starch gels. The reason for using starch gels instead of polyacrylamide, which provides superior resolution, is the ability to slice a thick starch gel into several thin slabs in order to test several isozymes with a single electrophoretic step. Each gel slice is incubated in the substrate for the enzyme being tested. The active enzyme converts the substrate into the product at the site in the gel where the enzyme is located. The product is usually detected by chemical methods that cause a color change.

Differences in electrophoretic mobility represent underlying differences in the DNA that coded for the enzyme. Each enzyme represents one locus. After several enzymes have been analyzed, similarities and differences in enzyme mobility can be compared among the genotypes tested. The data can then be subjected to various forms of statistical analysis, usually some form of hierarchical cluster analysis, to determine the relative genetic distances.

Isozyme analysis is effective and has been used in many studies. In general, it is not expensive and requires little in terms of equipment and expertise. However, its use has declined significantly in recent years as DNA methods have replaced it. It has two primary limitations. First, since it involves enzymes that presumably are subject to natural selection favoring the most adaptive types, the degree of polymorphism (genetic difference) is low compared to other methods. It is not uncommon to compare a variety of isozymes among different populations and find that there are no genetic differences in the isozymes even though there may be obvious morphological differences. Isozymes are often more valuable for comparing

related species than for assessing genetic diversity within a species. Second, each isozyme requires a different method for detection, meaning that a variety of chemicals must be purchased and a variety of techniques applied to obtain data from several isozymes. DNA methods use a single procedure for all DNA markers, providing a significant advantage over isozyme analysis when large numbers of markers are compared.

Seed Storage Proteins—Seed storage proteins have also been used to compare genetic diversity (Fairbanks and others 1990). Seed storage proteins are deposited in relatively large quantities in mature seeds and typically remain stable until the seed germinates. Their purpose is to provide the germinating seedling with a source of amino acids until it is able to begin photosynthesis and synthesize its own amino acids from photosynthate precursors. Since seed storage proteins are an amino acid source, rather than enzymes or structural proteins, alterations in their amino acid sequences are not as critical to the plant. Thus, they are not as subject to natural selection as their isozyme counterparts and may have higher levels of genetic polymorphism.

Seed storage proteins can be easily extracted from seeds and analyzed electrophoretically using polyacrylamide gel electrophoresis. Polymorphisms are detected as differences in electrophoretic mobilities and can be analyzed statistically just as isozymes are. Additional polymorphisms can be detected by subjecting the proteins to proteolytic cleavage using enzymes such as trypsin, chymotrypsin, and *Staphylococcus aureus* V8 protease, or chemicals such as cyanogen bromide that cleave polypeptides at specific amino acid sequences. The proteolytic fragments can then be separated electrophoretically to reveal additional genetic polymorphism.

The advantages of seed protein analysis are ease and economy. Relatively little is required in terms of laboratory equipment and the procedure is rapid. The one disadvantage is a distinct lack of available markers. The number of seed storage proteins that can be resolved by electrophoresis in any particular species is usually less than 20 and typically many of them are not polymorphic. While seed protein markers are useful, they usually must be augmented with other types of markers, such as isozymes or DNA markers for reliable estimates of genetic diversity.

DNA Markers

Restriction Fragment Length Polymorphism

(RFLP) Analysis—RFLP analysis is now a standard, well-tested procedure for estimation of genetic diversity. It has been used in thousands of studies in a variety of species. Virtually every university has more than one laboratory with RFLP capabilities. RFLP analysis is based on the ability of restriction enzymes (also called restriction endonucleases) to cleave DNA at specific target nucleotide sequences consisting usually of a four or six nucleotide pair target. For example, one of the more commonly used enzymes, *Eco* RI, cleaves DNA at the target sequence GAATTC. Each time the enzyme encounters this sequence, it cleaves the DNA molecule at that point.

When cellular DNA is subjected to restriction enzyme digestion, the enzyme cleaves the DNA into thousands of fragments, each defined by the target sequence. Differences in nucleotide sequences at target sites, or in the number of nucleotides between target sites generate fragments of different sizes. These different-sized fragments are called restriction fragment length polymorphisms or RFLPs. They can be separated from each other by gel electrophoresis, and then detected by hybridization of blotted DNA fragments with a labeled probe.

RFLP markers have been used extensively for genome mapping and for analysis of genetic diversity (see Tanksley 1993 and references cited therein). They have also been used in plant breeding programs for marker-assisted selection. RFLP markers tend to be inherited as simple Mendelian codominant alleles, and are not influenced in any way by environmental factors. Therefore, their heritabilities are invariably 100%.

There are three major advantages to RFLP analysis over protein markers. First, the number of polymorphic markers that can be generated is virtually limitless. No laboratory has ever exhausted the supply of potential RFLPs. Second, the procedure is the same for every marker, so once a laboratory is set up for RFLP analysis, no major changes need to be made for analyzing different markers. Third, the level of polymorphism is greater in RFLPs than in protein markers. Since RFLPs do not necessarily represent sequences that are expressed as genes (as proteins do), their DNA sequence is not as subject to preservation by natural selection, resulting in higher levels of polymorphism. Consequently, the amount of useful information obtained per marker tends to be higher with RFLPs than with protein markers.

There are disadvantages, however. RFLP analysis is more complicated and expensive than protein analysis. It requires more time and expertise. Until recently, RFLP analysis also required the use of radiolabeled DNA, which imposed the added safety precautions needed in working with radioactive materials. However, recent advances in recombinant DNA technology have made reliable non-radioactive labels available that work well for RFLP analysis. Another significant disadvantage for RFLPs is probe availability. For major field crops, sets of RFLP probes were once sold for as much as several hundred-thousand dollars. Now they are often available for the asking. Each time an RFLP program is initiated in a new species, a new set of probes often needs to be developed. In a few cases, probes from a related species may be used if they are available. For instance, most tomato probes can be used in potato. Probe generation by cloning and testing can be laborious and expensive. If probes are not readily available, RFLP programs may require a substantial initial investment of time and money before reliable data are available. Nonetheless, RFLP analysis remains a reliable and much-used technique for analysis of genetic diversity.

Random Amplified Polymorphic DNA (RAPD)

Analysis—RAPD analysis is a relatively new method designed to do essentially the same thing as RFLP analysis does with less effort and expense (Williams and others 1990). It is similar to the polymerase chain reaction, but relies on a single decanucleotide primer to amplify DNA

fragments that can be used for genetic analysis. Its use in genetic diversity analysis has been explained in detail in a number of sources (Fairbanks and others 1993; Joshi and Nguyen 1993; Dweikat and others 1993; Vierling and Nguyen 1992; Heusden and Bachman 1992; Williams and others 1990; Andersen and Fairbanks 1990).

The advantages to RAPD analysis are numerous. RAPD markers are typically inherited as simple dominant Mendelian alleles and for this reason can be used for most genetic marker applications. The procedure is streamlined compared to RFLP analysis. RFLP analysis requires digestion of DNA with a restriction enzyme, separation of fragments on an agarose gel, blotting onto a filter, probe hybridization, and fragment detection, a procedure that typically requires several days. RAPD analysis requires amplification of DNA in a thermocycler and separation of fragments on an agarose gel, a procedure that requires about six hours. The cost of RAPD analysis is about the same for supplies as RFLP, but less for labor. About 60% of the cost comes from a single reaction component, the enzyme Taq polymerase (Fairbanks and others 1993). There is no need for probe development since the same set of random primers may be used for all species. The only adaptation required is a few initial experiments to determine the ideal DNA and magnesium concentrations for the species in question. These many advantages have made RAPD analysis quite popular in the past few years.

In spite of the numerous advantages, there are some disadvantages. Alterations in DNA, the concentrations of template DNA, polymerase, primers, and magnesium can cause different fragments to be produced. For this reason, RAPD results are not always repeatable when conducted in different laboratories. Generally, it is best to replicate experiments and use appropriate controls when comparing one experiment with another. Also, there is often a tendency to produce non-specific amplification as a background smear of DNA, or fragments of DNA that are not present in high amounts and appear as faint bands in the gel. Often it is a matter of guesswork to determine which DNA fragments truly represent genetic polymorphism and which polymorphisms among less visible bands represent artifacts of the procedure. In spite of these limitations, most RAPD studies confirm the results of previous studies done using other types of molecular markers and morphological or agronomic traits, indicating that the RAPD analyses were reliable (Transue and others in press). Also RAPD markers mapped to a particular site on a chromosome typically map to the same site when converted to RFLPs, a further indication of reliability (Williams and others 1990).

Several refinements of RAPD analysis have been proposed to overcome some of these limitations. Nonspecific amplification can be reduced in some cases by prior digestion of template DNA with restriction enzymes (Riede and others in press; Sharma and others 1992). Artifacts due to variations in magnesium concentration can be reduced by using the Stoffel fragment (a genetically engineered polymerase) in place of Taq polymerase (Sobral and Honeycutt 1993). Additional polymorphisms can sometimes be visualized using denatured RAPD markers and polyacrylamide gel electrophoresis (Dweikat and others

1993). Each of these procedures is somewhat more laborious than standard RAPD analysis.

In order to provide markers that can be reliably transferred from one laboratory to another, Paran and Michelmore (1993) recommended that the nucleotide sequence of selected RAPD markers be determined and this information used to synthesize long primers that are specific for the marker. Then true PCR can be used to amplify the markers avoiding the possibility of artifacts. While this procedure requires a substantial investment of time and resources, it may be worthwhile if markers need to be compared in several different laboratories.

Conclusions

Several valuable methods for molecular marker analysis are currently available for estimation of genetic diversity. DNA marker analysis, including RFLP and RAPD analyses, are now the most frequently used methods due to their advantages over protein markers. Perhaps the greatest limitation for using DNA markers is the cost associated with laboratory set up, labor, and supplies.

References

- Andersen, W.R.; Fairbanks, D.J. 1990. Molecular markers: Important tools for plant genetic resource characterization. *Diversity* 6(3,4):51-53.
- Dweikat, I.; Mackenzie, S.; Levy, M.; Ohm, H. 1993. Pedigree assessment using RAPD-DGGE in cereal crop species. *Theor. Appl. Genet.* 85:497-505.
- Fairbanks, D.J.; Burgener, K.W.; Robison, L.R.; Andersen, W.R.; Ballon, E. 1990. Electrophoretic characterization of quinoa seed proteins. *Plant Breeding* 104:190-195.
- Fairbanks, D.J.; Waldrigues, A.; Ruas, C.F.; Ruas, P.M.; Maughan, P.J.; Robison, L.R.; Andersen, W.R.; Riede, C.R.; Pauley, C.S.; Caetano, L.G.; Arantes, O.M.N.; Fungaro, M.H.P.; Vidotto, M.C.; Jankevicius, S.E. 1993. Efficient characterization of biological diversity using field DNA extraction and random amplified polymorphic DNA markers. *Rev. Brasil. Genet.* 16:11-22.
- Heusden, A.W. van; Bachman, K. 1992. Genotype relationships in *Microseris elegans* (Asteraceae, Lactuceae) revealed by DNA amplification from arbitrary primers (RAPDs). *Pl. Syst. Evol.* 179:221-223.
- Joshi, C.P.; Nguyen, H.T. 1993. Application of the random amplified polymorphic DNA technique for the detection of polymorphism among wild and cultivated tetraploid wheats. *Genome.* 36:602-609.
- Paran, I.; Michelmore, R.W. 1993. Development of reliable PCR-based markers linked to downy mildew resistance in lettuce. *Theor. Appl. Genet.* 85:985.
- Riede, C.R.; Fairbanks, D.J.; Andersen, W.R.; Kehrer, R.L.; Robison, L.R. Enhancement of RAPD analysis by restriction endonuclease digestion of template DNA in wheat. *Plant Breeding* (in press).
- Sharma, J.K.; Gopalakrishna, V.; Dos, B.C. 1992. A simple method for elimination of unspecific amplifications in polymerase chain reaction. *Nucl. Acids Res.* 20: 6117-6118.
- Sobral, B.W.S.; Honeycutt, R.J. 1993. High-output genetic mapping of polyploids using PCR-generated markers. *Theor. Appl. Genet.* 86:105-112.
- Tanksley, S.D. 1993. Mapping polygenes. *Ann. Rev. Genet.* 27:205-233.
- Tanksley, S.D.; Orton, T.J. (ed.) 1983. *Isozymes in Plant Genetics and Breeding*, v. 1, v. 2. Elsevier, Amsterdam. 561 p., 472 p.
- Transue, D.K.; Fairbanks, D.J.; Robison, L.R.; Andersen, W.R. Species identification by RAPD analysis of grain amaranth genetic resources. *Crop Sci.* (in press).
- Vierling, R.; Nguyen, H.T. 1992. Use of RAPD markers to determine the genetic relationships of diploid wheat genotypes. *Theor. Appl. Genet.* 84:835-838.
- Williams, J.K.G.; Kubelic, A.R.; Livak, K.J.; Rafalski, J.A.; Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18:6531-6535.