Molecular genetic techniques for plant genetic resources

Report of an IPGRI Workshop
9-11 October 1995
Rome, Italy

W.G. Ayad, T. Hodgkin, A. Jaradat *and* **V.R. Rao,** *editors*

International Plant Genetic Resources Institute
IPGRI

The International Plant Genetic Resources Institute (IPGRI) is an autonomous international scientific organization operating under the aegis of the Consultative Group on International Agricultural Research (CGIAR). The international status of IPGRI is conferred under an Establishment Agreement which, by January 1997, had been signed by the Governments of Australia, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Cameroon, Chile, China, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, Greece, Guinea, Hungary, India, Indonesia, Iran, Israel, Italy, Jordan, Kenya, Malaysia, Mauritania, Morocco, Pakistan, Panama, Peru, Poland, Portugal, Romania, Russia, Senegal, Slovak Republic, Sudan, Switzerland, Syria, Tunisia, Turkey, Uganda and Ukraine. IPGRI's mandate is to advance the conservation and use of plant genetic resources for the benefit of present and future generations. IPGRI works in partnership with other organizations, undertaking research, training and the provision of scientific and technical advice and information, and has a particularly strong programme link with the Food and Agriculture Organization of the United Nations. Financial support for the research agenda of IPGRI is provided by the Governments of Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, India, Italy, Japan, the Republic of Korea, Luxembourg, Mexico, the Netherlands, Norway, the Philippines, Spain, Sweden, Switzerland, the UK and the USA, and by the Asian Development Bank, CTA, European Union, IDRC, IFAD, Interamerican Development Bank, UNDP and the World Bank.

Citation:
Ayad, W.G., T. Hodgkin, A. Jaradat and V.R. Rao, editors. 1997. Molecular genetic techniques for plant genetic resources. Report of an IPGRI workshop, 9-11 October 1995, Rome, Italy. International Plant Genetic Resources Institute, Rome, Italy.

ISBN 92-9043-315-9

IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy

Contents

Increasing the use of plant genetic resources

Technology transfer and application in developing countries

Introduction

Masa Iwanaga
International Plant Genetic Resources Institute, 00145 Rome, Italy

The last few years have seen a dramatic increase in the application of molecular genetic methods to problems relevant to the conservation and use of plant genetic resources. Studies on species relationships and evolution, the extent and distribution of diversity in crop and forestry species, accession identity and the detection of novel variants have provided valuable information for those engaged in conservation work. A variety of different methods for detecting and analyzing variation at the molecular level are now available and offer to those involved in plant genetic resources management ways of improving the effectiveness of their work.

At the same time, it has become clear that there are a number of unresolved questions concerning the appropriate use of the various techniques available. These include methodological questions, problems of data analysis and management, the need to use methods capable of dealing with large numbers of accessions, and resource limitations, especially in developing countries.

The objective of the International Plant Genetic Resources Institute (IPGRI) is to strengthen the conservation and use of plant genetic resources worldwide, with special emphasis on the needs of developing countries. Working in partnership with other organizations, it undertakes research and training and seeks to provide scientific and technical advice and information. IPGRI recognizes the considerable potential of molecular methods for improving the conservation and use of plant genetic resources and is currently concerned with strengthening work in this area.

Most of the early work using molecular techniques was carried out in laboratories in developed countries and these still predominate in terms of available equipment, skills and resources. Today, a number of laboratories in developing countries also possess the capacity for undertaking various kinds of molecular analysis and have staff with experience in the use of the relevant techniques. However, resources to undertake molecular genetic work on plant genetic resources conservation in developing countries remain extremely limited.

In November 1987, IBPGR (now IPGRI) held a meeting on the use of molecular techniques in plant genetic resources conservation and, since then, has collaborated on a number of projects in which molecular techniques have been used. These include analysis of diversity in *Musa*, *Vigna* and *Phaseolus* using RFLPs, fingerprinting and RAPDs and the use of RAPDs to assess stability of *in vitro* conservation of *Musa*. Many other research groups and genebanks have also undertaken studies of direct relevance to the conservation and use of plant genetic resources. Over the last three years, a 35-group pan-European project has been in progress to develop molecular tools for screening diversity in plants and animals.

The potential benefits of using molecular techniques are clear and individual genetic resources conservation projects are likely to make increasing use of the different methodologies now available. However, there remain a number of difficulties in the effective application of the different techniques and a number of questions concerning their effective use. In particular, there are difficulties in using appropriate technologies in resource limiting situations. For these reasons, IPGRI decided to host a small international workshop of experts working in different relevant areas of molecular genetics to discuss and identify the key issues. This Workshop was held in October 1995 with the following objectives:

1. To determine how molecular genetic techniques can be applied to improve the conservation and use of plant genetic resources with particular reference to the analysis of genetic diversity.
2. To identify major research and development needs and, in particular, areas of work on which IPGRI should focus.
3. To consider how adequate resources might be mobilized to carry out the work required.

It was attended by 15 invited participants (see List of Participants) as well as by a number of IPGRI staff members and included presented papers, discussions and working group sessions on key issues. In these Proceedings we have included the presented papers and a brief summary of the conclusions of the working groups. These are arranged according to subject in four sections.

I. Techniques for the analysis, characterization and conservation of plant genetic resources
II. Analysis, management and exchange of molecular data
III. Increasing the use of plant genetic resources
IV. Technology transfer and application in developing countries

A major conclusion of the Workshop was that, given the range of techniques now available and the relative novelty of their application in plant genetic resources conservation use, IPGRI should prepare a document which provided users with guidance as to their characteristics and appropriateness in different aspects of genetic resource management. It was, therefore, proposed that a technical bulletin should be drafted which would contain the following information:

- a brief description of each technique; what it does, how it works and what kind of information it provides;
- a description of the relevant plant genetic resources conservation and use issues for which molecular genetic tools can be used; and
- guidance on the appropriate tools for the different conservation and use issues under different circumstances of resources availability, knowledge base and urgency.

A chart outlining one possible decision-making procedure for determining the appropriate procedures for different circumstances was prepared by Angela Karp (see this Proceedings) and modified during the Workshop. The technical bulletin would provide the necessary text for amplifying and explaining this chart.

Some Current Issues in the Conservation and Use of Plant Genetic Resources

Toby Hodgkin
International Plant Genetic Resources Institute, 00145 Rome, Italy

Introduction

Molecular genetic techniques, both on their own and in combination with other biotechnological approaches, are beginning to have a significant impact on plant genetic resources conservation and use. Initially, the molecular techniques were used largely for the analysis of specific genes, for understanding gene action, gene mapping and the development of gene transfer technologies. Studies of phylogeny and species evolution have also been undertaken and have produced a considerable amount of valuable information. More recently, the techniques have been applied to problems of direct relevance for understanding the distribution and extent of genetic variation within and between species.

Most of the early work using molecular techniques was carried out in laboratories in developed countries and these still predominate in terms of available equipment, skills and resources. Today, a number of laboratories in developing countries also possess the capacity for undertaking various kinds of molecular analysis and have staff with experience in the use of the relevant techniques. However, resources to undertake molecular genetic work on plant genetic resources conservation in developing countries remain extremely limited.

In November 1987, IBPGR (now IPGRI) held a meeting on the use of molecular techniques in plant genetic resources conservation and, since then, has collaborated in a number of projects in which molecular techniques have been used. These include analysis of diversity in *Musa*, *Vigna* and *Phaseolus* using RFLPs, fingerprinting and RAPDs and the use of RAPDs to assess the stability of *in vitro* conservation of *Musa*. Many other research groups and genebanks have also undertaken studies of direct relevance to the conservation and use of plant genetic resources. Over the last three years, a 35-group pan-European project has been in progress to develop molecular tools for screening diversity in plants and animals.

The potential benefits of using molecular techniques are becoming clearer and individual genetic resources conservation projects are likely to make increasing use of the different methodologies now available. However, there remain a number of difficulties in the effective application of the different techniques and a number of questions concerning their effective use. In particular, there are difficulties in using appropriate technologies in resource limited situations. In this paper, I outline some current problems in the conservation and use of plant genetic resources where molecular techniques might be useful.

Phylogeny and evolution

A knowledge of the evolution of crop species and of the relationship between crop species and their wild relatives has been of considerable value for the conservation and use of plant genetic resources. Wild relatives of crop species contain many useful characters which breeders have introduced into crop cultivars. These introductions have tended to be most effective when the wild species are close relatives of the target

crop species or are even directly ancestral to it (Prescott Allen and Prescott Allen 1983; Oldfield 1989). Characters introduced to crop cultivars have included disease resistance (Ross 1978), tolerance of stresses such as salinity (Rick *et al.* 1987) and improved nutritional quality (Levy and Feldman 1987).

Harlan and de Wet (1971) provided a useful approach to classifying crop species and their wild relatives using the concepts of primary, secondary and tertiary genepools. The primary genepool includes all those taxa or species which can cross freely with the crop, the secondary, those species which cross with difficulty, giving few fertile seeds, and the tertiary genepool includes those species which can only be crossed using artificial techniques such as embryo rescue. This is a very practical approach based essentially on species biology and has proved to be of considerable value in allocating resources to conservation work.

Molecular techniques have allowed species relationships to be investigated in much more detail. The classic study by Bonierbale *et al.* (1988) revealed the striking similarities between *Lycopersicon esculentum* and *Solanum tuberosum* in genome organization. This work has now been extended to many other species groups particularly involving members of the Graminae including wheat, barley, rice , maize and sorghum. These studies can provide substantial information about species evolution and the origins of our major crop plants (Gepts 1995). They also provide information on the wild relatives of crop species which should be the subject of conservation work as in the case of the Brassicaceae where a much clearer picture of the relationships between the species is now available (Warwick and Black 1993). There are a number of crop and forestry genepools where such studies would undoubtedly help conservation decision making such as *Manihot* spp., *Prunus* spp. or the Sapotaceae.

The extent and distribution of diversity

A knowledge of the amount, the extent and the distribution of genetic variation is central to the development of effective conservation strategies. The amount of variation can be very different between species and between different populations of a species and there can also be large differences in the distribution of particular characters or groups of characters.

Differences between cultivated species and their wild relatives can be substantial and this has been attributed to phenomena such as the "founder effect" and amphidiploidy which are frequently involved in the evolution of a crop. Allard (1992) found that *Hordeum spontaneum* had an average of 5.15 alleles per loci for 20 isozyme and restriction fragment loci while barley landraces from the same geographic area had an average of 2.75 alleles per locus. Even greater differences have been found for other crops such as tomato (Miller and Tanksley 1990) and *Phaseolus* (Gepts *et al.* 1986; Debouck *et al.* 1989).

Just as there are substantial differences in the amounts of diversity present in different species, so there may be large differences between different groups within species. The occurrence of much larger amounts of variation in traditional cultivars or landraces than in modern cultivars is a common observation. Allard (1992) noted that barley cultivars from California possessed 1.44 alleles per locus in comparison with the 2.75 alleles per locus found in landraces from the Middle East. Different cultivar groups may also possess substantially different amounts of variation as has been found in summer and autumn maturing cauliflowers or in different cabbage types of the *Brassica oleracea* group.

Differences between populations within a species in the total amount of diversity appear to be greater in self-pollinated species than in cross- pollinated species. Schoen and Brown (1991) surveyed data from a number of autogamous and allogamous species and showed that, while there was variation in allelic richness for both groups, the variation between populations of autogamous species was much greater. The benefits of locating populations of autogamous species with high levels of allelic richness would seem to be substantial if one wishes to maximize the diversity conserved. In allogamous species, the emphasis would be on locating sites with different ecogeographic characteristics.

Genetic diversity within a species is not distributed uniformly throughout the range of environments in which it occurs. In fact, current evidence suggests that geographic distribution can account for much of the observed variation in wild plant species (Hamrick and Godt 1990). In crops, geographic distribution patterns also reflect the effect of human selection in particular environments as well as the history of crop development in difference locations.

There are numerous examples of differences in the distribution of particular characters, especially disease resistance and stress tolerance (e.g. Qualset 1975). Country of origin was found to be the most important factor in explaining differences in a number of spike characters in durum wheat (Spagnoletti Zeuli 1987) and has also been shown to be effective in accounting for differences in the distribution of isozymes in soybean (Perry 1991). Using isozyme data from oak populations found in different parts of Europe, Zanetto *et al.* (1994) have described differences in both the amount and of isozyme diversity and the distribution of specific alleles.

Marshall and Brown (1975) have argued that conservation strategies should take account not only of the absolute distribution of a particular allele or group of alleles but also of their frequency. They suggested that four classes of alleles could be recognized:

- common, widely distributed
- common, locally distributed
- rare, widely distributed
- rare, locally distributed

They argued that collecting and conserving the first class presents no problem and that such alleles are likely to be included in small samples collected from a few populations. The conservation of rare, widely distributed alleles will depend on the total resources allocated to conservation and will be relatively insensitive to the particular strategy chosen. The inclusion of rare locally distributed alleles will be very dependent on stochastic effects and such alleles will only be included if sample sizes are very large. However, the conservation of alleles which are common in specific locations but are not widely distributed will depend on the particular sampling strategy chosen and on having methods to locate such variants. They also argued that this class of alleles is particularly important because they will include alleles of adaptive significance for the populations that possess them. Allard (1992) has also emphasized the importance of targeting conservation work on alleles which are common in specific populations.

Conservation actions will always be limited and must be carefully planned to maximize the amount of useful diversity conserved. Ecogeographic and agroecological survey methods can be combined with the use of geographic information systems (GIS) and landscape analysis to provide data on species distribution and the environments in which they occur. However, the information has a low genetic content and ways in which such techniques could be combined with genetic surveys of a manageable and useful kind are urgently needed both for *ex situ* based collecting programmes and for identifying sites for *in situ* conservation.

Genetic erosion

Although it is generally accepted that significant amounts of genetic erosion have occurred and are still occurring due mainly to the destruction of ecosystems and habitats by human activities, there is surprisingly little data on the precise amount and extent. Certainly, for the major food commodities, there has been a dramatic increase in the use of a small number of highly selected uniform crop cultivars and this has been associated with a reduction in the number of traditional cultivars grown by farmers which are usually genetically highly variable. However, while there is clear evidence for a reduction in the number of cultivars grown (particularly in developed countries) and for a reduction in the area on which traditional cultivars or landraces are grown, the extent to which allelic diversity has been lost in particular crops has not been established.

Similarly, for both wild relatives and for forestry species, there is an increasing threat to existing populations from urbanization, road development, change of land use to agriculture and many other factors. These processes lead to fragmentation of existing populations or even to loss of whole populations and the effects of this process on intraspecific diversity have seldom been described. Methodologies are needed which can be used to monitor changes in diversity in the context both of total diversity and of adaptive diversity and which can be applied over significant periods of time. It is an interesting fact that, apart from the work on the Composite Crosses of barley (Allard 1992), we have virtually no genetic data on changes in crop diversity over time.

In situ conservation

There is currently increasing interest in the use of *in situ* conservation both for wild relatives of crop species and for crops themselves. In the case of forestry species, *in situ* conservation has been the traditional approach because of the obvious drawbacks of using *ex situ* methods except where linked to use or where specific emergency conditions require it. However in both forestry and crop species the specific issues and problems involved in conserving within taxa diversity have seldom been addressed. *In situ* conservation is based on specific locations in which the target taxa or species exists within an ecosystem or agroecosystem containing other species. The maintenance of the whole ecosystem is an essential part of an *in situ* conservation strategy but there may well be conflicts between the requirements for maintaining diversity of a particular target species and maintaining the ecosystem as a whole.

In situ conservation is the method of first choice for forest species and the wild relatives of crops. Once populations which should be conserved have been located, they need to be managed and monitored to ensure that they continue to survive within the ecosystems in which they occur. In particular, the genetic diversity present in the populations needs to be monitored over time to determine whether genetic shift or drift is occurring and whether management practices should be modified. In conserving useful species we are often interested in maintaining a much greater amplitude of variation than is necessary for species survival, since characteristics such as stress resistance may only be found at the margins of species distribution.

In situ conservation is also becoming increasingly important for crop plants. It is now recognized that many farmers continue to grow and use traditional varieties or landraces because these fill a need not met by modern cultivars (Brush 1995). The continued maintenance of these landraces ("on farm" conservation) provides an effective basis for sustainable development and has been recognized by the Convention

on Biological Diversity (CBD) to be of key importance. It also provides a continuing resource of adapted germplasm for plant breeders and other users.

Supporting "on farm" conservation provides conservation workers with new challenges. An understanding of the social, economic and cultural aspects of traditional farming systems becomes essential, as does support for traditional seed supply systems and farmer-based breeding and experimentation (Hodgkin *et al.* 1993). "On farm" conservation is carried out by farmers and communities when it meets their needs and interests and can only be sustained when it continues to do so. We need to know why farmers grow landraces, when they grow landraces and how they maintain them. At the same time, this approach raises major biological and genetic questions for the conservationist. Data is needed on the amounts of variation maintained and on the way farmers manipulate this variation over time.

Ex situ conservation

The objective of *ex situ* conservation is to maintain the accession without change as regards its genetic constitution (Frankel and Soulé 1981). The methods used are designed to minimize the possibility of change occurring through mutation, selection, random drift or contamination. For many crop species and their wild relatives, *ex situ* conservation can be carried out using seeds of the species which can be stored for long periods (up to hundreds of years) at low temperatures and low humidities. However, there are a number of clonally propagated crops such as potato or banana where this is not possible and a number of species which produce seeds which cannot be stored and are therefore termed recalcitrant. These last two groups of species can only be maintained *ex situ* in field genebanks as growing plants or *in vitro* using tissue culture or cryopreservation.

There are a number of problems associated with *ex situ* conservation which can affect the genetic integrity of an accession, or which present the genebank manager with difficult decisions. For material conserved as seeds, the periodic need to regenerate a sample because of declining seed viability in storage is one of the most important (Breese 1989; Rao 1991). The material must be grown in ways that ensure that genetic drift or shift are minimized using large enough populations and good seed production conditions. Available facilities in genebanks often limit the number of plants of an accession that can be grown and this is particularly so for allogamous species. The effect of different practices has not been studied in sufficient detail to provide managers with good guidance on the consequences of using smaller than desired population sizes or sub-optimum growing conditions but some studies in collaboration with IPGRI are now in progress.

Another significant issue in the management of *ex situ* collections has been the identification of duplicate accessions. Duplicates occurring within a single genebank are a waste of resources and there is considerable pressure to find ways of identifying and eliminating them. A more general problem concerns the extent to which similar accessions such as cultivars with the same Another name are, in fact, true duplicates. They may be found to differ with respect to isozyme or molecular markers while appearing to be morphologically identical. The identification of two accessions as identical on the basis of a few specific marker genes such as isozymes or RFLPs is clearly as questionable as the decision that two accessions that users consider identical are in fact different on the basis of a single molecular marker.

The problems of ensuring that the identity of accessions is maintained is equally important in the case of field genebanks where material may be subject to frequent replanting and to all the other problems of maintaining the crop in the field. For this

reason, there has been a substantial effort by IPGRI and other organizations to develop *in vitro* techniques for conservation (Withers 1994). It is important that any technique developed does not affect the genetic stability of the material conserved. Usually organized cultures, such as shoots, meristems or embryos are preferred for such work since unorganized tissues such as callus are more vulnerable to somaclonal variation. Nonetheless, the possibility of such variation occurring still exists and ways of monitoring material to confirm its genetic integrity are, therefore, needed. The same is true of cryopreservation (e.g. Harding 1991) which offers the best long term hope for conservation in vitro but where methods of monitoring stability are an essential component of the development of successful procedures.

Use

IPGRI's mandate deliberately includes reference to the use of plant genetic resources. Conservation is not seen as an end in itself and we are concerned to ensure that genetic resources are available to the user community. This community includes not only the plant breeders and agronomists who are involved in the production of new cultivars, it also includes the communities and farmers involved in the *in situ* maintenance of traditional crop cultivars and those who participate in *in situ* maintenance of forest genetic resources and wild crop relatives.

Traditional crop varieties managed by communities as part of their normal farming practices are, of course, being used as they are conserved. In this situation, use and conservation are intimately linked. However, it is hoped that these materials can also be made available to a wide range of other users both in the formal and in the informal sector. Crop relatives conserved *in situ* need to be made available to users and ways of accessing these resources are also important.

There are two somewhat connected approaches to improving the use of genetic resources. The first involves the development of strategies that enable searches for desired accessions or characteristics to be carried out efficiently in large sets of accessions or in large numbers of *in situ* conserved populations. This is exemplified by the development of core collections which aim to represent the diversity spectrum of a large genebank collection in a more limited set of accessions (usually about 10% of the total collection) as defined by Frankel (1984). Core collections have been established for a number of crops using a variety of approaches but, in each case, the larger collection has been divided into groups on the basis of taxonomic and ecogeographic features and samples have been obtained from each of the groups to constitute the core collection (Hodgkin *et al.* 1995). The objective is to define groups in which within group variation is less than between group variation and thus allow users to sample specific groups and simplify their search for specific characters. This general approach can be extended to sampling both within and between populations as our knowledge of the distribution of variation becomes more complete and as our ways of stratifying groups of accessions becomes more effective.

The second approach is concerned with the way one can improve the identification of accessions or plants with specific characters. Many characters are rather difficult to identify and are controlled by a number of genes which can interact in more or less complex ways. Some characters may only be detectable in mature plants under specific conditions and may be difficult to screen (disease resistance or cold tolerance of fruits etc.). Linked marker loci, particularly molecular markers, have considerable potential in screening for such characters and there is much interest in developing such approaches for breeding programmes. The improved definition of the control of

expression of such characters can make a substantial contribution as has been shown in the case of time to flower by Summerfield and colleagues (Summerfield *et al.* 1995).

Concluding remarks

Despite their diverse nature, the various issues listed above have some features in common. Firstly, they all involve genetic aspects of conservation biology - that is their solution depends on an improved use of genetic information about species, populations, accessions or characters. Secondly, what is required is not so much complex technical solutions but rather, simple easily applied methods that can be used with large numbers of samples in a variety of different situations and on a wide range of species.

Molecular genetic techniques have added a powerful new dimension to genetic studies and have already begun to be applied in many of the situations described above. However, the techniques available are changing rapidly and the quantity and quality of the information obtained from using them is also changing. There are exciting possibilities for using such techniques to improve conservation and use of genetic resources but there are also a number of questions which need to be considered as we begin to apply the techniques. These include:

The relative appropriateness of different techniques for different issues.

- The extent to which different techniques can be applied in situations where large numbers of samples need to be examined.
- The costs of different methods and the availability of the materials needed for their use.
- The dangers of information overload and the need to decide how much data is needed or useful.
- The relationship between molecular data and information on characters of interest to users.
- The degree to which different techniques give different kinds of information or even reflect different aspects of diversity.

No doubt there will be a number of other questions that need to be added to this list and considered during this Workshop and I look forward to the discussions of the next three days.

References

Gepts, P. and F.A. Bliss. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. 40:469-478.

Allard, R.W. 1992. Predictive methods for germplasm identification. Pp. 119-146 *in* Plant Breeding in the 1990s (H.T. Stalker and J.P. Murphy, eds.). CAB International, Wallingford, Oxon, UK.

Anon. 1995. Turkey designated by GEF as ideal site for landmark *in situ* conservation project. Diversity 11:64-67.

Bonierbale, M.W., R.L. Plaisted and S.D. Tanksley. 1988. MFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103.

Breese, E.L. 1989. Regeneration and multiplication of germplasm resources in seed genebanks: the scientific background. International Board for Plant Genetic Resources, Rome, Italy, 69p.

Brush, S.B. 1995. *In situ* conservation of landraces in centers of crop diversity. Crop Sci. 35:346-354.

Debouck, D.G., A. Maquet and C.E. Posso. 1989. Biochemical evidence for two different gene pools in lima beans, *Phaseolus lunatus* L. Ann. Rept. Bean Improvement Coop. 32:58-59.

Frankel, O.H. and M.E. Soule. 1981. Conservation and Evolution. Cambridge University Press, Cambridge, UK.

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. *In* Genetic Manipulation: Impact on Man and Society (W. Arber, K. Llimensee, W.J. Peacock and P. Starlinger, eds.). Cambridge University Press. Cambridge, UK.

Gepts, P. 1995. Genetic markers and core collections. Pp. 127-146 *in* Core Collections of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th. J.L. van Hintum and E.A.V. Morales, eds.). Wiley-Sayce Publishers, UK.

Hamrick, J.L. and M.J.W. Godt. 1990. Allozyme diversity in plant species. Pp. 43-63 *in* Plant Population Genetics, Breeding and Genetic Resources (A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir, eds.). Sinauer Associates Inc., Sunderland, Massachusetts, USA.

Harding, K. 1991. Molecular stability of the ribosomal RNA genes in *Solanum tuberosum* plants recovered from slow growth and cryopreservation. Euphytica 55:141-146.

Harlan, J.R. and J.M.J. de Wet. 1971. Toward a rational classification of cultivated plants. Taxon 20:509-517.

Hodgkin, T., A.H.D. Brown, Th. J.L. van Hintum and E.A.V. Morales. 1995. Future directions. Pp. 253-259 *in* Core Collections of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley and Sons, UK.

Hodgkin, T. and D.G. Debouck. 1992. Some possible applications of molecular genetics in the conservation of wild species for crop improvement. Pp. 153-182 *in* Conservation of Plant Genes, DNA Banking and *in vitro* Biotechnology (R.P. Adams and J.E. Adams, eds.). Academic Press, Inc., San Diego, California.

Hodgkin, T., V. Ramanatha Rao and K. Riley. 1993. Current issues in conserving crop landraces *in situ*. Paper presented at On-Farm Conservation Workshop, Bogor Indonesia, Dec. 6-8, 1993.

Levy, A.A. and M. Feldman. 1987. Increase in grain protein percentage in high-yielding common wheat breeding lines by genes from wild tetraploid wheat. Euphytica 36:353-359.

Marshall, D.R. and A.H.D. Brown. 1975. Optimum sampling strategies in genetic conservation. *In* Genetic Resources for Today and Tomorrow (O.H. Frankel and J.G. Hawkes, eds.). Cambridge University Press. Cambridge, UK.

Miller, J.C. and S.D. Tanksley. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. 80:437-448.

Oldfield, M.L. 1989. The value of conserving genetic resources. Sinauer Associates Inc., Sunderland, Massachusetts, USA. 379p.

Perry, M.C., McIntosh, M.S. and Stoner, A.K. 1991. Geographical patterns of variation in the USDA soybean germplasm collection II. Allozyme frequencies. Crop Sci. 31:1356-60.

Ramanatha Rao, V. 1991. Problems and methodologies for management and retention of genetic diversity in germplasm collections. Pp 61-68 *in* Proceedings of ATSAF/IBPGR Workshop on Conservation of Plant Genetic Resources (B. Becker, ed.). ATSAF/IBPGR, Bonn.

Rick, C.M., J.W. DeVerna, R.T. Chetelat and M.A. Stevens. 1987. Potential contributions of wide crosses to improvement of processing tomatoes. Acta Hort. 200:45-55.

Ross, H. 1978. Wild species and primitive cultivars as ancestors of potato varieties. Pp. 237-245 *in* Broadening the Genetic Base of Crops. PUDOC, Wageningen, The Netherlands.

Schoen, D.J. and A.H.D. Brown. 1991. Intraspecific variation in population gene diversity and effective population size correlates with the mating system. Proc. Nat. Acad. of Sci., USA 88:4494-97.

Spagnoletti Zeuli, P.L. and C.O. Qualset. 1987. Geographical diversity for quantitative spike characters in a world collection of durum wheat. Crop Sci. 27:235-41.

Summerfield, R.J., R.J. Lawn, R.H. Ellis, A. Qi, E.H. Roberts, S. Shanmugasundaram, P.M. Chay, J.B. Brouwer, J.L. Rose, S.J. Yeates and S. Sandover. 1995. Towards the reliable prediction of time to flowering in six annual crops. IPGRI/CSIRO publication, August 1995.

Warwick, S.I. and L.D. Black. 1993. Molecular relationships in subtribe Brassicinae (Cruciferae, tribe Brassiceae). Can. J.Bot. 71:906-918.

Withers, L.A. 1994. New technologies for the conservation of plant genetic resources. Pp. 429-435 *in* Proceedings of the International Crop Science Congress. Ames, USA. Crop Science Society of America.

Zanetto, A. *et al.* 1994. Geographic variation of inter-specific differentiation between *Quercus Robur* L., and *Quercus Petraea* (Matt.) Liebl. Forest Genet. 1(2):111-123.

Techniques for the analysis, characterization and conservation of plant genetic resources

Molecular techniques in the analysis of the extent and distribution of genetic diversity

Angela Karp and Keith J. Edwards
IACR-Long Ashton Research Station, Department of Agricultural Sciences, University of Bristol, Bristol BS18 9AF, UK

Introduction

Recent advances in molecular biology, principally in the development of the polymerase chain reaction (PCR) for amplifying DNA, DNA sequencing and data analysis, have resulted in powerful techniques which can be used for the screening, characterization and evaluation of genetic diversity. The extensive number of research articles currently appearing in the literature, describing the use of these techniques in a wide range of plant species and diversity problems, is testimony to their increasing impact in this field. Nevertheless, there are still many problems to be addressed before universal strategies for their wide-spread use can be recommended. Comparative studies in which different approaches have been contrasted in specific germplasms, some of which are discussed at this meeting, provide extremely valuable insights into the relative strengths and weaknesses of the different technologies.

It is still difficult, however, to extrapolate these experiences into the broadest context of diversity screening, where, in practice, a whole range of additional factors may need to be taken into account relating to: the specific questions being addressed; the capital investment and skills available; the level of polymorphisms anticipated and; the 'starting position' of the plant in question in terms of whether previous molecular investigations have been carried out. A further complication is that the techniques sample the genome differently and provide different information with respect to the diversity questions being addressed. Whilst it may be unrealistic to identify screening strategies that will suit all the potential systems under study, it should be possible to identify a clear rationale for selecting the appropriate screening strategy for any given system. In this short review we examine the choice of molecular screening technologies available for analyzing the extent and distribution of genetic diversity in natural populations and *in situ* resources. The merits and limitations of each technique are discussed with a view to identifying criteria for their recommendation. In terms of making general recommendations, a major criterion is how quickly the techniques can be adapted to work on any new system and this will thus form the underlying theme of the present discussion.

Molecular genetic screening strategies

A whole range of different techniques can be used to detect polymorphisms at the DNA level. In fact the seemingly bewildering array of possible approaches is among the first problems faced by newcomers considering the application of these techniques to their own system. In reality, however, this wide array falls into three broad categories with respect to basic strategy: (A) Non-PCR based approaches; (B) PCR Arbitrary priming; and (C) Targeted-PCR and sequencing.

Non-PCR based screening techniques

Restriction fragment length polymorphism (RFLP) analysis was the first technology developed which enabled the detection of polymorphisms at the sequence level. The approach involves digesting DNA with restriction enzymes, separating the resultant DNA fragments by gel electrophoresis, blotting the fragments to a filter and hybridizing probes to the separated fragments. A probe is a short sequence of oligonucleotides which share homology and are thus able to hybridize, with a corresponding sequence or sequences in the genomic DNA. The sequence may be known (e.g. a cloned gene) or unknown (e.g. from random cDNA or genomic DNA clone). Specific probe/enzyme combinations give highly reproducible patterns for a given individual but variation in the restriction patterns between individuals can arise when mutations in the DNA sequence result in changed restriction sites. RFLP analysis is used extensively in the construction of genetic maps and has been successfully applied to genetic diversity assessments, particularly in cultivated plants (e.g. Castagna *et al.* 1994; Deu *et al.* 1994; Jack *et al.* 1995) but also in populations and wild accessions (e.g. Besse *et al.* 1994; Laurent *et al.* 1994; Bark and Harvey 1995). As a technique for diversity studies, there are three important advantages which should be considered. The first is that RFLPs are highly reproducible between laboratories and the diversity profiles generated can be reliably transferred. The second is that RFLPs are co-dominant markers, enabling heterozygotes to be distinguished from homozygotes. The third advantage is that no sequence-specific information is required and, provided suitable probes are available, the approach can be applied immediately for diversity screening in any system.

There are serious limitations, however, with the RFLP strategy with respect to wide-scale usage at the population level and particularly with regard to its potential for immediate application to any system. Firstly, a good supply of probes is needed that can reliably detect variation at the below species level. It may be possible to utilize (heterologous) probes from other related species, a possibility very much strengthened by syntenic relationships between related genera. If this is not possible, probes must be isolated from cDNA or genomic DNA libraries from the species in question, which requires additional skills and considerable investment of time. In all cases, it will be necessary to select suitable probe/enzyme combinations before the actual diversity screening work can begin. There are few ways of speeding up the necessary pre-screening work, although prior knowledge of sequence composition can help in the choice of restriction enzymes. RFLPs are time-consuming and they are not amenable to automation without considerable capital investment. Once probe/enzyme combinations have been selected, throughput will depend on the number of gels that can be run each day in the laboratory in question (see Table 1). To this must be added the factor of DNA extraction. RFLP analysis requires relatively large quantities of good quality DNA (e.g. 10μg per digestion). For some plant systems, where extraction is problematic because of the presence of polyphenols or polysaccharides which complex with the DNA, or where only very limited amounts of source material are available, this feature alone may preclude the choice of RFLP analysis for diversity screening.

Even in those systems where all the above considerations are optimal, the main problem faced may simply be that insufficient polymorphisms are detectable at the below species level by RFLP analysis. There are, however, certain classes of sequence which may overcome this problem. Interspersed among the genomes of higher organisms are highly variable regions which are comprised of repeats of short simple sequences. These are known as "microsatellites", where the basic repeat unit is around two to eight base pairs in length and "minisatellites" for longer repeat units of 16 to 100 base pairs. Although their function is still not entirely clear, many are located in

centromeric regions, telomeric regions and in the roots of chromatin loops, whilst others are thought to play a role in pairing and synapsis of chromosomes. Because these regions are hypervariable, RFLP analysis with probes for micro- and minisatellites gives multilocus patterns which have resolved variation at the levels of populations and individuals. The variation usually results from changes in the copy number of the basic repeat and is often referred to as Variable Numbers of Tandem Repeats (VNTRs). Because of the very high levels of polymorphisms that they detect, VNTRs are recognized as powerful tools, particularly for fingerprinting and cultivar identification in plants (e.g. Beyermann *et al.* 1992; Vosman *et al.* 1992). They have also been used for studying within and between population variation, for ecological studies (e.g. Alberte *et al.* 1994; Wolff *et al.* 1994) and for estimating genetic distances (e.g. Lynch 1990; Antonius and Nyborn 1994).

When considering the choice of this approach for any new system under study, in addition to all the factors listed above for RFLP with standard probes, a particular aspect should be considered which relates to the nature of VNTRs and the quality of the data generated. As with other RFLPs the success of multilocus fingerprinting is dependent on the probe/enzyme combination used which has to be tested out each time a new species is studied. VNTR loci are co-dominant, but because they are multilocus probes, the complexity of the patterns obtained, in combination with the presence of an infinite number of alleles at each locus, means that alternative statistical procedures are required for the use of VNTR data in classical population genetic models and makes the accuracy of analyzing relationships using VNTR data very problematic (Lynch 1990; Scribner *et al.* 1994). The problem can be reduced by selecting single locus VNTRs, but this increases the pre-screening and selection process considerably. Taking all aspects of non PCR-based screening approaches into consideration, it is difficult to envisage that this would be the preferred choice today, given that alternative strategies are now available. When combined with PCR amplification of a specific locus, however, both VNTRs and standard RFLP probes have much to offer, as will be described later.

Table 1. Comparison of the different molecular screening techniques

Characteristics	RFLPs	RAPDs	Seq Tag SSRs	AFLPs	PCR-Seq
Development costs ($ per probe)	Medium (100)	Low (none)	High (500)	Low (none)	High (500)
Level of Polymorphism	Low - Medium	Medium	High	Medium	Medium
Automation?	No	Yes/No	Yes/No	Yes/No	Yes
Cost of Automation	High	Medium	High	High	High
Reliability	High	Low	High	Medium	High
Level of Skill Required	Low	Low	Low - Medium	Medium	High
Cost ($ per assay)	High (2.00)	Low (1.00)	Low (1.00)	Medium (1.50)	High (2.00)
Radioactivity	Yes/No	No	Yes/No	Yes/No	Yes/No
Samples/ day (research)	20	50	50	50	20

PCR arbitrary priming techniques

With the advent of PCR, a number of techniques became available for the screening of genetic diversity. These require no prior sequence-specific information and can, therefore, be applied directly to any organism. The techniques are based on the use of a single 'arbitrary' primer, which may be purchased from commercial companies, in a PCR reaction on genomic DNA and result in the amplification of several discrete DNA products. Each of these products will be derived from a region of the genome that contains two short segments with some homology to the primer, which are on opposite strands, and sufficiently close together for the amplification to work. A number of closely related techniques based on this principle were developed at the same time and are collectively referred to as multiple arbitrary amplicon profiling (MAAP) (Caetano-Annollés 1994). The most commonly used is RAPD (Randomly Amplified Polymorphic DNA) analysis in which the primers are usually 10-mer or 20-mers and in which the amplification products are separated on agarose gels in the presence of ethidium bromide and visualized under ultraviolet light (Williams *et al.* 1990). AP-PCR (Arbitrary primed PCR) (Welsh and McClelland 1990) and DAF (DNA Amplification Fingerprinting) (Caetano-Annollés *et al.* 1991) differ from RAPDs principally in primer length, the stringency conditions and the method of separation and detection of the fragments. In all cases, polymorphisms are detected as the presence or absence of bands and result from sequence differences in one or both of the primer binding sites. For simplicity, only RAPDs will be further referred to in this discussion.

The enormous attraction of RAPDs is that there is no requirement for DNA probes, nor for any sequence information for the design of specific primers. The procedure involves no blotting or hybridizing steps. The technique is, therefore, quick, simple and efficient and only requires the purchase of a thermocycling machine and agarose gel apparatus to set up in a laboratory for any new system under study. It requires small amounts of DNA (10ng per reaction) and sample throughput can be quite high (see Table 1). The procedure can also be made automatic with extremely high throughput. RAPDs have also been proved to detect higher levels of polymorphism compared with RFLPs in cases where the two techniques have been applied to the same material. They have been extensively used for screening diversity, particularly at intraspecific levels, including many population studies (Hadrys *et al.* 1992). Unfortunately, the approach has serious limitations.

The first concerns the nature of the data generated. RAPDs are dominant markers such that the homozygote conditions are the only genotypes discernible as presence or absence of the band. In addition, the presence of a band of apparently identical molecular weight in RAPD gels of different individuals cannot be taken as evidence that the two individuals have the same band, although this assumption is commonly made. Further complications are that single RAPD bands can be comprised of several co-migrating amplification products and, as in the case of DNA fingerprinting, there can be uncertainty in assigning markers to specific loci in the absence of preliminary pedigree analysis (Clark and Lanigan 1993). Lynch and Milligan (1994) have recently discussed these limitations of RAPD for population genetic analysis and state 'provided there is only a single amplifiable allele per locus, this does not prevent the estimation of allele frequencies necessary for population-genetic-analysis, but it does reduce the accuracy of such estimation relative to analysis with codominant markers.' Although completely unbiased estimators for RAPDs do not appear to be possible, they suggest several steps which will ensure that the bias is negligible. In their article, they derive estimators for: gene and genotype frequencies; within and between population heterozygosities; degree of inbreeding; population subdivision and degree of individual relatedness. One important conclusion from their study is that to achieve

the same degree of statistical power using RAPDs (or any other dominant marker system), compared with co-dominant markers, two to ten times more individuals need to be sampled per locus and further, to avoid bias in parameter estimation, the marker alleles for most of these loci should be in relatively low frequency.

The use of RAPDs for determining the distribution and extent of variation is challenged even further when the second general problem of RAPDs is considered concerning the robustness of the data generated. RAPDs are notoriously prone to user-error in that, unless the most consistent of conditions is strictly adhered to, the RAPD profiles obtained can vary considerably between different runs of the same sample. Even though careful practice quickly overcomes this problem, RAPD profiles are difficult to reproduce between laboratories, which may have different PCR machines or use different sources of polymerase and associated buffers. Even within a laboratory, the time saved by the direct application of RAPDs is often lost in achieving consistency and in confirming the reproducibility of the results obtained. As PCR machines are being improved all the time and new thermostable polymerases continue to appear on the market, it is predictable that any particular data from RAPD profiles will have a transient life. It is our feeling that this aspect of RAPDs cannot be over-emphasised and that, together with the statistical qualifications outlined above, these disadvantages of this strategy seriously outweigh the apparent advantages which might otherwise make this the procedure of choice.

More recently, Keygene have developed a method which is equally applicable universally, which reveals very high levels of polymorphism and which is highly reproducible. This procedure, termed Amplified Fragment Length Polymorphism (AFLP) (Zabeau and Vos 1992) is essentially intermediate between RFLPs and RAPDs, in that the first step is restriction digestion of the genomic DNA but this is then followed by selective rounds of PCR amplification of the restricted fragments. The fragments are amplified by P^{33}- labelled primers designed to the sequence of the restriction site, plus one to three additional selected nucleotides. Only fragments containing the restriction site sequence plus the additional nucleotides will be amplified and the more selected nucleotides added on to the primer sequence (up to a maximum of three can be added at either site) the fewer the number of fragments amplified by PCR. This selection is necessary to achieve a total number of fragments within the range that can be resolved on a gel (approximately 150 to 200 fragments). The amplified products are normally separated on a sequencing gel and visualized after exposure to X-ray film. Recently, the technique has been automated, using fluorescent labelled primers and, therefore, high throughput can be achieved. Two different types of polymorphisms are detected: (1) point mutation in the restriction sites, or in the selective nucleotides of the primers which result in a signal in one case and absence of a band in the other; and (2) small insertions/deletions within the restriction fragment which results in different size bands.

AFLPs have proven to be extremely proficient in revealing diversity at below the species level and provide an effective means of covering large areas of the genome in a single assay. Although we have classified them under arbitrary priming approaches they can be targeted to specific sequences (e.g. VNTRs) if these are used in the primer design. All the evidence so far indicates that they are as reproducible as RFLPs, thereby overcoming one of the major problems with RAPDs. They require more DNA (1μg per reaction) and are more technically demanding than RAPDs, requiring experience of sequencing gels, and (manually) necessitating the use of radioactivity, but their recent automation and the availability of kits in some species means that the technology can be bought in at a higher level. AFLPs, however, do run into the same problem as RAPDs regarding the type of data generated and the concomitant problems of data analysis for population genetic parameters. Although Keygene are developing means

of identifying heterozygotes, AFLPs are essentially a dominant marker system, the identity of the DNA fragments amplified on the gels is not known, and fragments which migrate to the same molecular weight in the AFLP profile of two different individuals cannot be conclusively interpreted as being the same. Unlike RAPDs, individual bands on an AFLP gel are single DNA fragments (although they may be repeated sequence elements), but the assignment of alleles to loci may be difficult, without pedigree analysis, when levels of heterozygosity are high and the resultant AFLP patterns very complex. In short, AFLPs provide multilocus bi-allelic fingerprints, combining aspects of RAPDs and multilocus VNTRs, and they will thus need to be subject to considerable analysis by statisticians before the applicability of their data to population analyses can be determined.

Targeted PCR and sequencing

The opposite approach to the arbitrary amplicon profiling procedures is to design primers to target specific regions of the genome. The targeted amplified product can be compared on an agarose gel to the corresponding product from another individual but the resolution achievable will only detect differences in length of the fragment resulting from many base pair changes. In order to resolve all the possible sequence differences, it is necessary to sequence the entire fragment, either manually or using an automated DNA sequencer. Once this has been done, sequences from different individuals can be aligned, any differences detected and the data entered and analyzed in statistical packages. This approach is applicable to extremely small samples, down to single pollen grains or tiny leaf fragments.

Although manual sequencing is routine for many laboratories and automated sequencing is affordable for many others, it is still a daunting prospect to have to derive the sequence of several thousands of individuals. Such an approach also requires that the investigators have sufficient molecular biological skills and equipment. A number of gel systems, such as TGGE (thermal gradient gel electrophoresis), DGGE (denaturing gradient gel electrophoresis), single strand conformational polymorphism (SSCP) and heteroduplex formation, provide sensitive detection assays without the need for complete sequencing but, at present, none are routinely used because they require sophisticated gel systems, highly controllable conditions and experienced workers (Hayashi 1992; Reisner *et al.* 1992; White *et al.* 1992). Another possibility for comparing sequences without sequencing, however, which is much easier to carry out, is to combine restriction fragment analysis with targeted PCR. In this PCR-RFLP approach the amplified product is digested with a specific restriction enzyme and the products directly visualized on the agarose gel by ethidium bromide staining (Tragoonrung *et al.* 1992; Ghareyazie *et al.* 1995). The approach is popular since it requires a simply PCR-based assay, the target sequence is known, the fragments can be interpreted and checked for any artifactual data and the differences generated are robust and usable for analysis by digital means. The approach is most informative when the restriction sites are mapped rather than simply detected as RFLPs. Plants possess three different genomes and, therefore, three potential sources of sequences for a PCR-targeted approach: the chloroplast genome (cpDNA), the mitochondrial genome (mtDNA) and the nuclear genome. cpDNA is maternally inherited in most plants. It is highly abundant in leaves and, therefore, amenable to isolation. The entire DNA sequence is known for three species (a liverwort, tobacco and rice) and appears to be highly conserved in terms of size, structure, gene content and order. It has consequently proved to be a very powerful tool for phylogenetic studies but, more recently, an increasing number of examples of intraspecific variation in cpDNA have indicated its potential for below-species diversity studies. cpDNA fragments amplified from

regions of the genome such as the tRNA region have been successfully used to study plant populations using both PCR-RFLP and PCR-sequencing strategies (e.g. Ali *et al.* 1991; McCauley 1994). Because of the conserved nature of the chloroplast genome, the primers used in these studies should have broad range applicability. In contrast to the chloroplast genome, mtDNA in plants has proved to be more limited as a tool for studying diversity. It is less abundant in leaves and more difficult to extract, there is less background knowledge, fewer probes are available and these have been less well characterized. The high rates of structural rearrangements and the relatively low rates of point mutations mean it is of limited use at the interfamilial and interspecific taxonomic levels. Consequently its contribution to phylogenetic studies has been limited. Conversely, the high frequency of rearrangements which can be easily detected as RFLPs mean that mtDNA can be very useful at detecting variation at the intraspecific and population levels. Primer pairs for conserved regions of mtDNA sequences are available and when used in PCR-RFLP analyses have provided very informative studies of population differentiation and diversity (Dong and Wagner 1993; Strauss *et al.* 1993). In terms of the nuclear genome, there is only one sequence to date that has been used extensively as a tool for studying genetic diversity and that is the rDNA gene family which encodes for ribosomal RNA. Ribosomal RNA genes are located at specific chromosomal (*Nor*) loci (usually in only one or two chromosomes) where they are arranged in tandem repeats which can be reiterated up to thousands of times. Interestingly, at each locus, rDNA genes are remarkable for the high degree of sequence conservation between members of the same gene family. Each repeat unit comprises a transcribed region separated from the next repeat by an intergenic spacer (IGS). The transcribed region comprises: an external transcribed spacer (ETS), the 18S gene, an internal transcribed spacer (ITS1), the 5.8S gene, a second internal transcribed spacer (ITS2) and the 26S gene. Certain regions have been highly conserved throughout eukaryotic evolution and, therefore, provide very useful phylogenetic tools, whilst other regions, such as ITS1 and ITS2, are highly variable and can be used to detect polymorphisms at the below species level. Primers pairs have been designed which will enable amplification of all the different regions of the rDNA and primer pairs, particularly for the ITS, have been used to detect variation in studying plant populations (Karvonen 1994). However, in the current literature, there are few reports where rDNA primers have been used for population analyses.

The advantages of PCR-targeted approaches are in the quality of the data and the information they engender. The fragment in which polymorphisms are studied is of known identity, therefore avoiding the ambiguities of analyzing RAPD and AFLP bands, or random RFLP probes. For population studies, the use of an organellar sequence in complementation with a nuclear sequence can provide particularly illuminating data with respect to mechanisms of differentiation, gene flow and dispersal. In contrast, the origin of RAPD (and AFLP) fragments with respect to the three genomes is generally unknown, although where the origin of the fragments has been studied, there is clear evidence that at least a proportion of RAPD fragments are of cpDNA or mtDNA origin. One very important advantage of the PCR-targeted approach, when used in combination with sequencing, or restriction site analysis, has recently been discussed by Milligan and co-authors (Milligan *et al.* 1994) in relation to conservation issues. They conclude that development of genealogical based analytical methods coupled with studies of DNA sequence variation within and among populations is likely to reveal the most information on demographic processes. RFLPs, RAPDs and AFLPs provide indirect data that is only useful when converted into distance measures. This enables frequency data and distance measures to be determined for each genotype class but does not enable the classes to be ordered or

grouped in any way. Data based on DNA sequences or restriction site mapping, on the other hand, provide the means of both classifying individuals into different classes and also of assessing relationships among the classes. There are many who would argue that this alone makes them the only system of choice.

There are clear disadvantages of the PCR-targeting approach, however. The first is that, unless the frequency of variants is high enough to be easily detected by PCR-RFLP, or other sensitive gel assays, sequencing will be required which, in turn, necessitates investment of adequate resources and experienced researchers. Another problem is that, although the quality of the data is high, because the approach is often resource-intensive (see Table 1) the coverage of the genome is highly restricted, usually to only one sequence and, therefore, to one point of comparison. The major difficulty, which is pertinent to both these issues, lies in the identification of sequences that are reliably variable enough at the population level in any system under study. If investigators are starting completely from scratch with this approach in an entirely new system, they would have to clone potentially suitable target sequences, derive the sequence of these fragments, design primers and test out the fragments amplified by the primer pairs for detecting polymorphism in their particular genotypes. All of this requires an enormous investment of time before serious diversity screening work can begin. For the PCR-targeted strategy to be widely applicable, target sequences need to fit two specific criteria. They must contain regions where the sequence is sufficiently conserved such that primers designed for one organism will amplify the same region in a broad range of taxa. At the same time, they must contain regions where the sequence varies at a rate that is high enough for polymorphisms to occur at the population level. Ideally, this should be at a rate such that PCR-RFLP, or rapid assays such as SSCP and TGGE, would uncover sufficient polymorphism, although complete sequencing is the only method that will detect all the variation present. Fortunately, new investigators selecting this strategy do not have to start entirely from scratch, because regions of cpDNA and mtDNA that fit these criteria have been identified, as described above. At the present time, however, there is a dearth of nuclear genes that fit the bill. Furthermore, the rate at which sequences vary (and, therefore, the success of this strategy) appears to differ between genomes and, at present, the limited number of suitable sequences and the worry that those available may not be variable enough in the system under study, are the main reasons why this approach may not be the choice selected. Additional problems when conserved primers are used for PCR are contamination and the detection of multiple gene copies and pseudogenes for nuclear sequences which were thought to be single copy.

Microsatellites or simple sequence repeats (SSRs) are highly mutable loci and, as mentioned earlier, when used as RFLP probes are variable at the population level and can even distinguish individuals and assign parentage. The problems of using them as multi-locus probes, outlined earlier, arise because the repeat sequence may be present in many different regions of the genome. However, since the flanking sequences at each of these loci may be unique, if SSR loci are cloned and sequenced, primers to the flanking regions can be designed and used to amplify only that single region containing the SSR, which is then referred to as a sequence-tagged microsatellite (or a sequence tagged SSR) (Morgante and Olivieri 1993).

There are several important advantages of choosing sequence-tagged SSRs for population genetic studies. They are usually single loci which, because of their high mutation rate, are often multi-allelic (Saghai-Maroof *et al.* 1994), they are co-dominant markers and they can be detected by a PCR (non-hybridization based) assay. They are very robust tools that can be exchanged between laboratories and their data is highly informative. As with conventional VNTRs, the variation at the SSR locus is caused by

changes in the repeat length. Although many such changes can be resolved on agarose gels, it is common to run SSRs on sequencing gels where single repeat differences can be resolved and all possible alleles detected. The assay is relatively quick (see Table 1), but throughput can be increased by selecting a small number of different SSRs with alleles that have different non-overlapping size ranges and multiplexing either the PCR reactions, or the products of the separate reactions, so that all the alleles of the different loci can be run in a single lane on the gel. Multiplexed SSRs have been automated, in which case throughput can be increased further.

There are, nonetheless, some negative aspects of using sequence tagged SSRs. Although they are co-dominant markers, their mode of evolution is different from normal coding loci. Different SSR alleles are thought to arise by slippage or unequal crossing-over and their rate of mutation, and the possibility of deriving the same length alleles by multiple events, mean that it is difficult to use them to estimate relatedness beyond a few generations. This in turn means that the phylogenetic information cannot be derived from SSRs. Furthermore, the potentially infinite number of alleles possible at SSR loci make computation of allelic frequencies problematic. Both these features have been addressed by statisticians so that for important population genetic parameters such as F_{ST} estimators for SSR loci (R_{ST}) have been derived, but phylogenetic inferences are still limited. Another major problem with choosing this strategy is that unless the investigator is extremely fortunate, sequence-tagged loci will not be available for the system that they wish to study. Although they are ubiquitous (Gupta *et al.* 1994), retrieval of SSRs has not been easy in plants because of their relative low abundance compared with animal genomes. Where they have been isolated, it has often been found that they show limited cross transferability to other genera and even to other species within the same genus. All this means that the investigator wishing to choose SSRs is first faced with having to isolate them. Whilst retrieval strategies have now been devised which work with high efficiency (Edwards *et al.* submitted) this requirement still necessitates a considerable investment of time and requires a certain level of extra skilled expertise and resource. All the SSR loci that the investigator does retrieve will have to be tested for usefulness. When it is considered that many may turn out to be monomorphic, the whole procedure is extremely high cost at present and cannot be applied immediately for most systems under study. It is conceivable that this particular problem will recede as more and more sequence-tagged SSR loci are described in the literature. However, their limited cross-transferability still make them a difficult choice for someone starting from scratch in an entirely new system, particularly when speed is essential and data from the screening work are urgently needed.

Criteria for deciding upon the 'right' screening strategy

A summary of the different characteristics and how they compare for the different molecular screening technologies is given in Table 1. Faced with all these competing strengths and weaknesses, is it possible to draw up a check list of criteria which might help new investigators decide upon the right choice for their particular system ? In many ways the question is premature because more information is required on the technologies and on the statistical appropriateness of the data they generate. Furthermore, the word 'right' has no practical reality as often the choice will be a compromise based on opposing needs. Bearing all this in mind and knowing that the picture may change significantly as more work is done and new procedures are developed, we have tried to present a logical framework of questions and answers

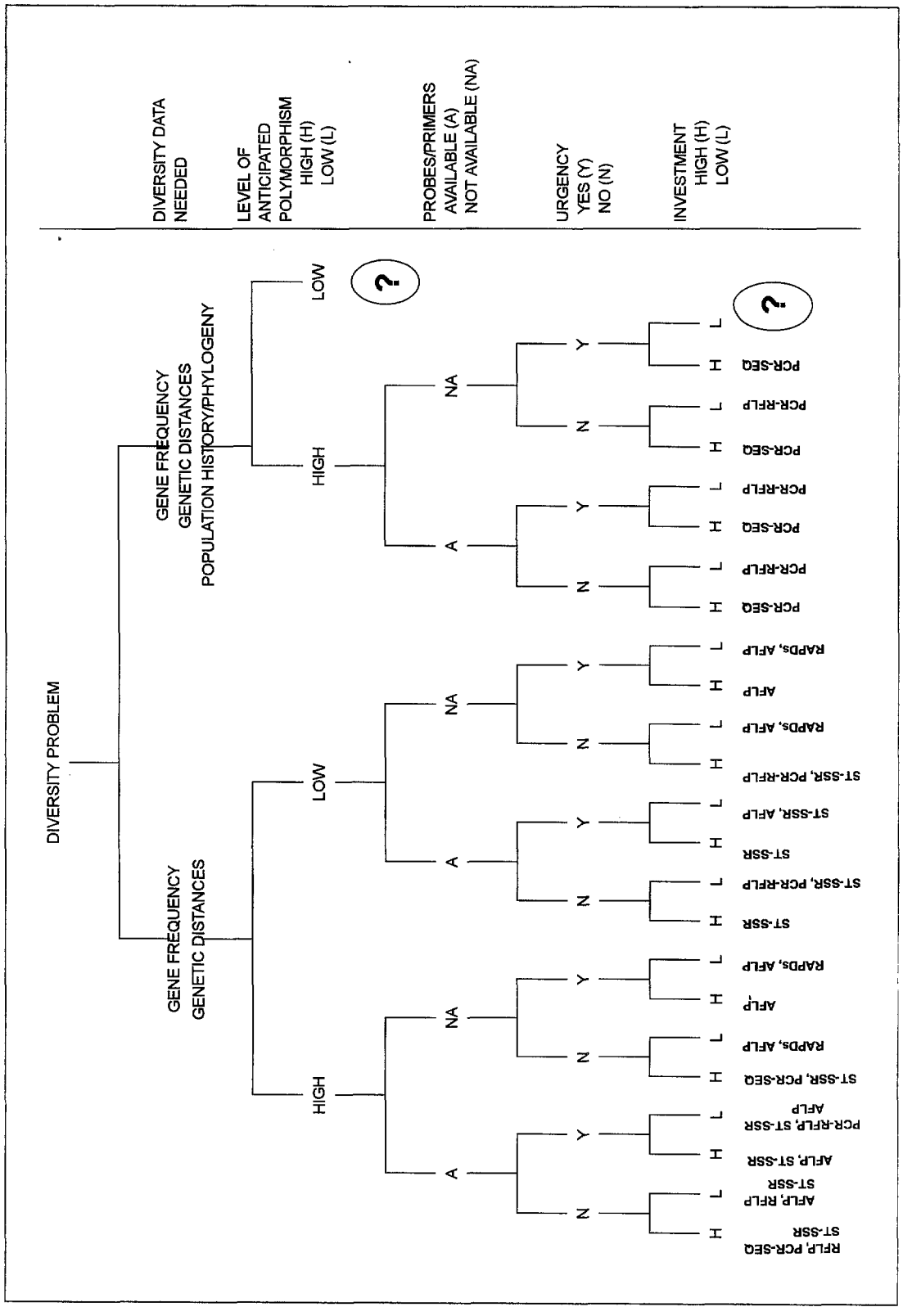

Fig. 1. A rational scheme for making decisions on the choice of molecular genetic screening strategy. It is important to note that the scheme presumes that supply of material is not limited. Where only minute samples are available, only PCR-based techniques are applicable and, of these, AFLP would be excluded from the range of choices. Where a (?) appears on the scheme there is considerable doubt as to the feasibility of the approach and we would recommend re-thinking the question that it is possible to address given these limitations.

which should help any new investigator, faced with the problem of applying these technologies now, decide upon the most appropriate strategy (Fig.1). This scheme will hopefully stimulate discussion and, most certainly, different molecular biologists will have their own opinions on the content of the bottom line of the figure. However, whilst the latter may change, we believe that the series of questions posed will remain valid and may help provide a much-needed framework for making recommendations on the use of these techniques in diversity studies.

Acknowledgements

The authors gratefully acknowledge the participants of the Framework III BIOTECHNOLOGY EU-funded contracts Nos. PL 920295, 920476, 920486, 920373 who collectively interact in the generic project "Molecular Genetic Screening Tools" coordinated by AK and with whom many discussions on this subject matter have been enjoyed.

References

Alberte, R.S., G.K. Suba, G. Procaccini, V. Zimmerman and S. Fain. 1994. Assessment of genetic diversity of seagrass populations using DNA fingerprinting: implications for population stability and management. Proc. Nat. Acad. Sci. USA 91:1049-1053.

Ali, I.F., D.B. Nelae and K.A. Marshall. 1991. Chloroplast DNA restriction fragment length polymorphisms in *Sequoia sempervirens* D. Don Endl., *Pseudotsuga menziesii* (Mirb). Franco, *Claocedrus decurrens* (Tott.) and *Pinus taeda* L. Theor. Appl. Genet. 81: 83-89.

Antonius, K. and H. Nybom. 1994. DNA fingerprinting reveals significant amounts of genetic variation in a wild raspberry *Rubus idaeus* population. Mol. Ecol. 3:177-180.

Bark, O.H. and M.J. Havey. 1995. Similarities and relationships among populations of the bulb onion as estimated by RFLPs. Theor. Appl. Genet. 90:407-414.

Besse, P., M. Seguin, P. Lebrun, M.H. Chevallier, D. Nicolas and C. Lanaud. 1994. Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. Theor. Appl. Genet. 88:199-207.

Beyermann, B., P. Nürnberg, A. Weihe, M. Meixner, J.T. Epplen and T. Börner. 1992. Fingerprinting plant genomes with oligo nucleotide probes specific for simple repetitive DNA sequences. Theor. Appl. Genet. 83:691-694.

Caetano-Anollés, G. 1994. MAAP: a versatile and universal tool for genome analysis. Plant Mol. Biol. 25:1011-1026.

Caetano-Anollés, G., B.J. Bassam, and P.M. Gresshoff. 1991. DNA amplification fingerpinting using very short arbitrary oligonucleotide primers. Bio/Technology 9:553-557.

Castagna, R., G. Maga, M. Perenzin and M. Heun. 1994. RFLP-based genetic relationships of Einkorn wheats. Theor. Appl. Genet. 88:818-823.

Clark, A.G. and C.M.S. Lanigan. 1993. Prospects for estimating nucleotide divergence with RAPDs. Mol. Biol. and Evol. 10:1096-1111.

Deu, M., D. Gonzalez-de-Leon, J.C. Glaszmann, I. Degremont, J. Chantereau, C. Lanaud and P. Hamon. 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor. Appl. Genet. 88:838-844.

Hayashi, K. 1992. PCR-SSCP: A method for detection of mutations. Genetic Analysis: Techniques and Applications 9:73-79.

Dong, J. and D.B. Wagner. 1993. Taxonomic and population differentiation of mitochondrial diversity in *Pinus banksiana* and *Pinus contorta*. Theor. Appl. Genet. 86:573-578.

Edwards, K.J., J.H.A. Barker, A. Daly, C. Jones and A. Karp. 1995. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques (in press).

Gharzeyazie, B., N. Huang, G. Second, J. Bennet and G.S. Kush. 1995. Classification of rice germplasm, I. Analysis using ALP and PCR-based RFLP. Theor. Appl. Genet. 91:218-227.

Gupta, M., Y.S. Chyi, J. Romero-Severson and J.L. Owen. 1994. Amplification of DNA markers from evolutionary diverse genomes using single primers of simple sequence repeats. Theor. Appl. Genet. 89:998-1006.

Hadrys, H., M. Balik and B. Schierwater. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Mol. Ecol. 1:55-63.

Jack, P.L., T.A.F. Dimitrijevic and S. Mayes. 1995. Assessment of nuclear, mitochondrial and chloroplast RFLP markers in oil palm (*Elaeis guineensis Jacq.*). Theor. Appl. Genet. 90:643-649.

Karvonen, P., A.E. Szmidt and O. Savolainene. 1994. Length variation in the internal transcribed spacers of ribosomal DNA in *Picea abies* and related species. Theor. Appl. Genet. 89:969-974.

Laurent, V., A.M. Risterucci and C. Lanaud. 1994. Genetic diversity in cocoa revealed by cDNA probes. Theor. Appl. Genet. 88:193-198.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Mol. Biol. Evol. 7:478-484.

Lynch, M. and B.G. Milligan. 1994. Analysis of population structure with RAPD markers. Mol. Ecol. 3:91-99.

Milligan, B.G. 1991. Chloroplast DNA diversity within and among populations of *Trifolium pratense*. Current Genet. 19:411-416.

Milligan, B.G., J. Leebens-Mack and A.E. Strand. 1994. Conservation genetics: beyond the maintenance of marker diversity. Mol. Ecol. 3:423-435.

Morgante, M. and A.M. Olivieri. 1993. PCR-amplified microsatelilites as markers in plant genetics. Plant J. 3:175-182.

McCauley, D.E. 1994. Contrasting the distribution of chloroplast DNA and allozyme polymorphism among local populations of *Silene alba*: Implications for studies of gene flow in plants. Proc. Nat. Acad. Sci. USA 91:8127-8131.

Riesner, D.G., J. Steger, M. Wiese, M. Wulfert, M. Heibey and K. Henco. 1992. Temperature-gradient electrophoresis for the detection of polymorphic DNA and for quantitative polymerase chain reaction. Electrophoresis 13:632-636.

Saghai-Maroof, M.A., R.M. Biyashev, G.P. Yang, Q. Zhang and R. Allard. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proc. Nat. Acad. Sci. USA 91:5466-5470.

Scribner, K.T., J.W. Arntzen and T. Burke. 1994. Comparative analysis on Intra- and Inter-population genetic diversity in *Bufo bufo* using allozyme, single locus microsatellite, Minisatellite and multilocus mini-satellite data. Mol. Biol. Evolution 11:737-748.

Strauss, S.H., Y.P. Hong and V.D. Hipkins. 1993. High levels of population differentiation for mitochondrial DNA haplotypes in *Pinus radiata, muricata* and *attenuata*. Theor. Appl. Genet. 86:605-611.

Tragoonrung, S., V. Kanazin, P.M. Hayes and T.K. Blake. 1992. Sequence-tagged-site facilitated PCR for barley genome mapping. Theor. Appl. Genet. 84:1002-1008.

Vosman, B., P. Arens, W. Rus-Kortekaas and M.J.M. Smulders. 1992. Identification of highly polymorphic DNA regions in tomato. Theor. Appl. Genet. 85:239-244.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:6531-6535.

White, M.B., M. Carvalho, D. Derse, S.J. O'Brien and M. Dean. 1992. Detecting single base substitutions as heteroduplex polymorphisms. Genomics 12:301-306.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Raflaski and S.V. Tingey. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic. Acid Res. 18:6531-6535.

Wolff, K., S.H. Rogstadt and B.A. Schaal. 1994. Population and species variation of minisatellite DNA in Plantago. Theor. Appl. Genet. 87:733-740.

Zabeau, M. and P. Vos. 1992. Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application. No: 0 534 858 A1.

Zhang, Q., M.A. Saghai Maroof and R.W. Allard. 1990. Effects on adaptedness of variations in ribosomal DNA copy number in populations of wild barley (*Hordeum vulgare* ssp. *spontaneum*). Proc. Nat. Acad. Sci. USA 87:8741-8745.

Using molecular markers in genebanks: identity, duplication, contamination and regenaration

Stephen Kresovich[1], James R. McFerson[2] and Anne L. Westman[1]
U.S. Department of Agriculture, Agricultural Research Service,
Plant Genetic Resources Units
[1]Griffin, Georgia 30223-1797 and [2]Geneva, New York 14456-0462

Although collections of genetic resources have existed for hundreds of years as zoos, botanic and estate gardens, hunting and nature reserves etc., the development of extensive, formal *ex situ* genetic resources collections of crop plants is a recent phenomenon. Historically, agricultural development has permitted an expanded world population. As noted by Harlan (1975), traditional agricultural systems were based on variable, integrated, adapted populations called landraces. While these systems narrowed genetic diversity within landraces via artificial and natural section, overall diversity was retained in centres of diversity, first termed centres of origin by Vavilov. As agricultural productivity increased, thereby allowing expanded human activities and industrial development, the process of genetic erosion had begun. This process has been greatly accelerated in the past fifty years with the development of scientific plant breeding and the globalization of agricultural technologies. As public and scientific recognition of this genetic erosion increased, conservation biology emerged as a new, multidisciplinary field. Concomitantly, agricultural scientists worldwide undertook to gather and preserve genetic resources of crop plants.

This collection was often opportunistic and haphazard. Strategies and techniques were still under development, primarily by Frankel and colleagues (Frankel and Soule 1981; Frankel and Brown 1984; Frankel 1985, 1986). Funding, trained scientists and adequate storage facilities were all in short supply. Even so, current Food and Agricultural Organization (FAO) estimates indicate that approximately 4 million accessions are held worldwide in *ex situ* collections. As might be expected, nearly half of these accessions are the major grain or food legume crops. In addition, considerable genetic resources exist in present crop improvement programmes and as cultivars in use on farms. Certainly, these collections have contributed to continuing agricultural development (Shands 1990) and represent an irreplaceable treasure of biotic raw material. Some argue that further collection is unnecessary and cite the principal shortcoming of current collections which is their lack of evaluation. However, when establishing measures of collection quality, it is misleading to consider only sheer numbers, whether considered singly or aggregated worldwide. Recent publications raise questions and highlight the importance of considering the quality of *ex situ* germplasm collections (Kresovich and McFerson 1992; Hintum 1994). Though easily stated, this strategy is more difficult to articulate than number-counting. In a sense, it is based on biology rather than bookkeeping. It requires development and understanding of appropriate tools and interpretation of results.

The first step involves clarifying the nature and purpose of an *ex situ* germplasm collection. We have stated elsewhere (Kresovich and McFerson 1992) that all germplasm collections should contain materials that offer both short-term benefits as well as long-term insurance. A collection of a selected taxon should be well-characterized and represent the maximum variation of a taxon with a reasonable number of accessions. It should include: wild and weedy relatives, landraces, cultivars (obsolete and current) and genetic stocks. This approach, built on the genepool concept first articulated by Harlan and deWet (1971), is easily applied to the needs of plant

breeders, who have constituted the primary users of *ex situ* collections. Obviously, crop improvement programs will preferentially utilize germplasm whose genotype most closely approximates current commercially desirable cultivars (Genepool 1). Usually, this will dictate use of available breeding lines. However, the genetic variation allowing introgression of desirable alleles is increasingly available only in exotic materials (Genepools 2 and 3). Although use of such materials is difficult and expensive, it is increasingly the only avenue for discovering and utilizing genetic variation for more complex traits. Since the genetic base of most crops is limited, it is unlikely that Genepool 1 contains the necessary genetic variation for improving crops for future unknown challenges such as novel biotic and abiotic stresses.

In addition, it is important to recognize that the potential user community of a collection should extend beyond plant breeders. While most *ex situ* collections of plant germplasm are crop-oriented, plant genetic resource curators ought to promote their collections for use in basic and applied science, as well across disciplines. Such studies can lead to tangible benefits, such as identification of characters valuable in production agriculture, as well as intangible benefits, such as increased knowledge of a range of biological systems and phenomena.

In all cases, whether the user is exploring issues of plant ecology or gene action and evolution, the quality of a collection will be determined by its representativeness and level of characterization. Since knowledge of representativeness requires characterization, it is on this aspect we focus. Specifically, we summarize the use of molecular markers to characterize germplasm collections. While our experience and most of our examples lie with crop plants, we believe the strategies and techniques involving molecular markers have broad applicability to genetic resources issues across all organisms.

Effective conservation of genetic resources requires a comprehensive understanding of genetics. Since most *ex situ* collections of plant germplasm centre their efforts around a given crop group, most work is conducted at the species or population (accession) level. However, a comprehensive understanding also includes other organizational levels, from ecosystems through cellular and molecular levels. The genetic issues of most interest to the curator can be summarized as follows:

- identity: the entry in a collection is catalogued correctly and is true to type;
- relatedness: the degree of similarity or difference among individual genotypes in an accession or among accessions in a collection;
- structure: the amount of genetic variation present and its partitioning among accessions and individual genotypes; and
- location: the location of a desired gene/gene complex in an accession and also the physical site of a desired DNA sequence on a particular chromosome in an individual plant.

The curator needs accurate, rapid and cost-effective tools to assess the collection for these parameters. In the past, these tools were limited and predominantly measured characters of agricultural interest. At best, these characters were found to be inherited in a Mendelian fashion. Usually, the characters were not simply inherited and were greatly affected by the test environment. Very often data sets for a collection are incomplete and/or of questionable validity. Even when data sets have been assembled for an entire collection, their utility is often hampered by the limited number of markers available, different test environments and errors associated with G x E variation. While much of this information is still critically important to the curator and the crop-oriented user community, it is insufficient to resolve many of the curators' s questions regarding the genetic parameters outlined previously, even in the best-characterized systems, like barley, maize, pea, tomato or wheat. For many plants — fruits, vegetables,

medicinals, trees and shrubs — almost nothing is known. Given the precarious state of funding for genetic resources programmes, this situation will not change. Many collections already cannot fund activities in the "life-support systems" of maintenance and rejuvenation, let alone fund those for expensive, long-term evaluations. Fortunately, advances in technology and genetic knowledge over the past decade promise new strategies and tools of considerable potential to collection curators. Methods using molecular markers, especially those based on DNA polymorphisms, are evolving at a breathtaking rate. Molecular markers are now a routine tool in many scientific fields, used to study animal, bacterial, fungal, plant and viral genomes at population and individual levels. Technical advances in instrumentation, protocol and data handling capability show no signs of abating and contribute to an ever more favourable cost/unit biological information ratio.

As highlighted by Avise (1994), molecular markers are used most intelligently when they address controversial areas or when they are employed to analyze problems in natural history and evolution (and we include agriculture) that have proven intractable using traditional nonmolecular observation. We reduce this statement to "ask the right questions, use the right tool." Avise (1994) further acknowledges that molecular genetic markers are valuable because: 1) molecular data are genetic; 2) molecular methods open the entire biological world for genetic scrutiny; 3) molecular methods access a nearly unlimited pool of genetic variability; 4) molecular data can distinguish homology from analogy; 5) molecular data provide a common measure for assessing divergence; 6) molecular approaches facilitate mechanistic appraisals of evolution; and finally, 7) molecular approaches are challenging and exciting. These same arguments hold true for the application of molecular genetic markers for more effective conservation of agricultural genetic resources.

Several recent reviews describe in detail a range of molecular markers useful for assessing plant genetic diversity (Strauss *et al.* 1992; Kresovich *et al.* 1993; Bachman 1994; O'Brien 1994) As indicated previously, instrumentation and techniques are being developed rapidly so specific details are best obtained directly from the current literature. However, as one looks to the future, continuing themes become evident. Future molecular markers and assay systems will be low cost, automated, high throughput, user-friendly and easily integrated in parallel processes.

The ultimate goal in genetic analysis is to determine, analyze and store DNA sequence information for a wide variety of applications in agriculture and conservation biology. Through example, the Advanced Technology Program of the U.S. National Institute for Standards and Technology has as a goal to enable industry to deliver DNA diagnostics to a variety of industrial sectors at 0.1 to 0.01 the current price. In addition, the effort strives to enable the user community to access DNA sequencing apparatus for one third of the current cost while reducing actual DNA sequencing costs similarly to the reductions associated with DNA diagnostics. In particular, three approaches to DNA analysis which hold promise in agriculture and conservation include serial sequencing, hybridization analysis and amplification-based analysis (Advanced Technology Program 1994). Independently of approach, improvements must be made in sample preparation, assay technology, detection systems, and data management and analysis.

As described by Rao and Riley (1994), the use of new biotechnologies such as DNA molecular markers can assist the curator in acquisition, maintenance, characterization and evaluation activities. In fact, applying molecular markers and assay systems will allow the curator to systematically and comprehensively characterize an entire collection. Characterization then will act as a catalyst and support actions designed to improve the quality of the collection. Without commenting on the specific strengths

and weaknesses of particular investigations, a bibliography of relevant research associated with the organization and management of plant genetic resource collections has been prepared and is attached to complete the references cited in this paper. Because of frequent additions to the literature in these areas of investigation, this bibliography should not be considered complete and inclusive.

Bibliography

Akkaya, M.A., A.A. Bhagwat and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Alberte, R.S., G.S. Suba, G. Procaccini, R.C. Zimmerman and R.S. Fain. 1994. Assessment of genetic diversity of seagrass populations using DNA fingerprinting: implications for population stability and management. Proc. Nat. Acad. Sci. USA 91:1049-1053.

Aldrich, Amos, B. and A.R. Hoelzel. 1992. Applications of molecular genetic techniques to the conservation of small populations. Biol. Conserv. 61:133-144.

Aldrich, P.R. and J. Doebley. 1992. Restriction fragment variation in the nuclear and chloroplast genomes of cultivated and wild *Sorghum bicolor*. Theor. Appl. Genet. 85:293-302.

Andersen, W.R. and D.J. Fairbanks. 1990. Molecular markers: Important tools for plant genetic resources characterization. Diversity 6:51-53.

Antonius, K. and H. Nybom. 1994. DNA fingerprinting reveals significant amounts of genetic variation in a wild raspberry *Rubus idaeus* population. Mol. Ecol. 3:177-180.

Astley, D. 1992. Preservation of genetic diversity and accession integrity. Field Crops Res. 29:205-224.

Avise, J.C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. Trends Res. Ecol. and Evol. 4:279-281.

Avise, J.C. 1992. Molecular population structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. Oikos 63:62-76.

Avise, J.C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York.

Bachman, K. 1994. Molecular markers in plant ecology. New Phytol. 126:403-418.

Baker, R.J. 1994. Some thoughts on conservation, biodiversity, museums, molecular characters, systematics and basic research. J. Mammol. 75:277-287.

Bark, O.H. and M.J. Havey. 1995. Similarities and relationships among populations of the bulb onion as estimated by nuclear RFLPs. Theor. Appl. Genet. 90:407-414.

Becerra-Velasquez and P. Gepts. 1994. RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. Genome 37:256-263.

Beckman, J.S. and M. Soller. 1990. Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Bio/Technology 8:930-932.

Beeching, J.R., P. Marmey, M.C. Gavalda, M. Moirot, H.R. Haysom, M.A. Hughes and A. Charrier. 1993. An assessment of genetic diversity within a collection of cassava (*Manihot esculenta* Crantz) germplasm using molecular markers. Ann. Bot. 72:515-520.

Beese, K. 1992. Development and application of molecular genetic tools for the assessment and evaluation of genetic diversity. Pp.10-16. *In* United States - Commission of the European Communities Workshop, Biotechnology and Genetic Resources 21-22 Oct. 1992, Airlie, VA, USA.

Bell, C.J. and J.R. Ecker. 1994. Assignment of thirty microsatellite loci to the linkage map of Arabidopsis. Genomics 19:137-144.

Bernatsky, R. and S.D. Tanksley. 1989. Restriction fragments as molecular markers for germplasm analysis and utilization. Pp. 353-362 *in* The Use of Plant Genetic Resources (A.D.H. Brown, D.R. Marshall, O.H. Frankel and J.T. Williams, eds.). Cambridge University Press, Cambridge, United Kingdom.

Besse, P., M. Seguin, P. LeBrun and C. Lanaud. 1993. Ribosomal DNA variations in wild and cultivated rubber tree (*Hevea brasiliensis*). Genome 36:1049-1057.

Beyermann, B., P. Nurnberg, A. Weihe, M. Meixner, J.T. Epplen and T. Borner. 1992. Fingerprinting plant genomes with oligonucleotide probes specific for simple repetitive DNA sequences. Theor. Appl. Genet. 83:691-694.

Bogani, P., D. Cavalieri, R. Petruccelli, L. Polsinelli and G. Roselli. 1994. Identification of olive tree cultivars by using random amplified polymorphic DNA. Acta Hortic. 356:98-101.

Bonner, F.T. 1990, Storage of seeds: potential and limitations for germplasm conservation. For. Ecol. Manage. 35:35-43.

Briscoe, D.A., J.M. Malipica, A. Robertson, G.J. Smith, R. Frankham, R.G. Banks and J.S.F. Barker. 1992. Rapid loss of genetic variation in large captive populations of *Drosophila* flies: Implications for the genetic management of captive populations. Conserv. Biol. 6:416-425.

Brubaker, C.L. and J.R. Wendel. 1993. On the specific status of *Gossypium lanceolatum* Todaro. Genet. Res. Crop Evol. 40:165-170.

Burke, T., G. Dolf, A.J. Jeffreys and R. Wolff (eds.). 1991. DNA fingerprinting: Approaches and applications. Berkhauser Verlag, Berlin, Germany.

Byrne, M. and G.F. Moran. 1994. Population divergence in the chloroplast genome of *Eucalyptus nitens*. Heredity 73:18-28.

Caetano-Anolles, G., B.J. Bassam and P.M. Gresshoff. 1991a. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/Technology 9:553-557.

Caetano-Anolles, G., B.J. Bassam and P.M. Gresshoff. 1991b. DNA amplification fingerprinting: a strategy for genome analysis. Plant Mol. Biol. Reptr. 9:294-307.

Chalmers, K.J., R. Waugh, J.I. Sprent, A.J. Simins and W. Powell. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity 69:465-472.

Chalmers, K.J., R. Waugh, J. Watters, B.P. Forster, A. Nevo, R.J. Abbott and W. Powell. 1992. Grain isozyme and ribosomal DNA variability in *Hordeum spontaneum* populations from Israel. Theor. Appl. Genet. 84:313-322.

Chen, H.B., J.M. Martin, M. Lavin and L.E. Talbert. 1994. Genetic diversity in hard red spring wheat based on sequence-tagged-site PCR markers. Crop Sci. 34:1628-1631.

Clegg, M.T. 1990. Molecular diversity in plant populations. Pp. 99-116 *in* Plant Population Genetics, Breeding, and Genetic Resources (A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir, eds.). Sinauer Assoc., Sunderland, MA, USA.

Collins, G.G. and R.H Symons. 1993. Polymorphisms in grapevine DNA detected by the RAPD PCR technique. Plant Mol. Biol. Reptr. 11:105-112.

Condit, R. and S.P. Hubbell. 1991. Abundance and sequence of 2-base repeat regions in tropical tree genomes. Genome 34:66-71.

Cregan, P.C. 1992. Simple sequence repeat DNA length polymorphisms. Probe 2:18-22.

Cross, R.J., A.G. Fautrier and D.L. McNeil. 1992. IBPGR morphological descriptors - their relevance in determining patterns within a diverse spring barley germplasm collection. Theor. Appl. Genet. 85:489-495.

Crossa, J., S. Taba, S.A. Eberhart, P. Bretting and R. Vencovsky. 1994. Practical considerations for maintaining germplasm in maize. Theor. Appl. Genet. 89:89-95.

Crozier, R.H. 1992. Genetic diversity and the agony of choice. Biol. Conserv. 61:11-15.

Dahlberg, K.A. 1992. The conservation of biological diversity and U.S. agriculture: goals, institutions and policies. Agric. Ecosyst. Environ. 42:177-193.

Dallas, J.F. 1988. Detection of DNA fingerprints of cultivated rice by hybridization with a human minisatellite DNA probe. Proc. Nat. Acad. Sci. USA 85:6831-6835.

Dawson, E.P., K. Wang, S.P Jiao, J.R. Harris and J.R. Hudson. 1994. DNA archival storage and retrieval systems. Pp. 93-99 *in* Conservation of Plant Genes II: Utilization of Ancient and Modern DNA (R.P. Adams, J.S. Miller, E.M. Golenberg and J.E. Adams, eds.). Missouri Botanical Garden, St. Louis, MO.

Dawson, I.K., K.J. Chalmers, R. Waugh and W. Powell. 1993. Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. Mol. Ecol. 2:151-159.

Dellaporta, S.L., J. Wood and J.B. Hicks. 1983. A plant DNA minipreparation: Version II. Plant Mol. Biol. Reptr. 1:19.

Demeke, T., R.P. Adams and R. Chibbar. 1992. Potential taxonomic use of random amplified polymorphic DNA (RAPD): a case study in *Brassica*. Theor. Appl. Genet. 84:990-994.

Demeke, T. and R.P. Adams. 1994. The use of RAPDs to determine germplasm collection strategies in the African species *Phytolacca dodecandra* (Phytolaccaceae). Pp. 131-139 *in* Conservation of Plant Genes II: Utilization of Ancient and Modern DNA (R.P. Adams, J.S. Miller, E.M. Golenberg and J.E. Adams, eds.). Missouri Botanical Garden, St. Louis, MO.

Deu, M., D. Gonzalez de Leon, J.C. Glaszmann, I. Degremont, J. Chanereau, J. Lanaud and P. Hamon. 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor. Appl. Genet. 88:838-844.

Devlin, B. and N.C. Ellstrand. 1990. Male and female fertility in wild radish, a hermaphrodite. Am. Naturalist 136:87-107.

Diwan, N., G.R. Bauchan and M.S. McIntosh. 1994. A core collection for the United States annual *Medicago* germplasm collection. Crop Sci. 34:279-285.

Dodds, J.H. and K. Watanabe. 1990. Biotechnical tools for plant genetic resources management. Diversity 6:26-28.

Doebley, J.F. 1992. Molecular systematics and crop evolution. Pp. 202-222 *in* Molecular Systematics of Plants (P.S. Soltis, D.E. Soltis and J.J. Doyle, eds.). Chapman & Hall, New York, New York.

Doyle, J.J. 1992, Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17:144-163.

Dweikat, I., S. MacKenzie, M. Levy and H. Ohm. 1993. Pedigree assessment using RAPD-DGGE in cereal crop species. Theor. Appl. Genet. 83:497-505.

Ellstrand, N.C. and M.L. Roose. 1987. Patterns of genotypic diversity in clonal plant species. Am. J. Bot. 74:123-131.

Engels, J.M.M. 1993. How can biotechnology be exploited in the conservation and use of biological diversity? *In* Plant Biotechnology in Technical Cooperation. Proc. Workshop GTZ, Legaspi, Phillipines, 6-11 Oct. 1992.

Epplen, J.T., H. Ammer, C. Epplen, C. Kammerbauer, R. Mitreiter, L. Roewer, W. Schwaiger, V. Steimle, H. Zischler, E. Albert, A. Andreas, B. Beyermann, W. Meyer, J. Buitkamp, I. Nanda, M. Schmid, P. Nurnberg, S.D.J. Pena, H. Poche, W. Sprecher, M. Schartl, K. Weising and A. Yassouridis. 1991. Oligonucleotide fingerprinting using simple repeat motifs: a convenient, ubiquitously applicable method to detect hypervariability for multiple purposes. Pp. 50-69 *in* DNA Fingerprinting; Approaches and Applications (T. Burke *et al.*, eds.). Birkhauser Verlag, Berlin, Germany.

Eguiarte, L.E., N. Perez-Nasser and D. Pinero. 1992. Genetic structure, outcrossing rate and heterosis in *Astrocarynum mexicanum* (tropical palm): implication for evolution and conservation. Heredity 69:217-228.

Ellstrand, N.C. 1992. Gene flow by pollen: implications for plant conservation genetics. Oikos 63:77-86.

Ellstrand, N.C. and D.R. Elam. 1993. Population genetic consequences of small populational size: implications for plant conservation. Ann. Rev. Ecol. Syst. 24:217-242.

Erlich, H.A. and N. Arnheim. 1992. Genetic analysis using the polymerase chain reaction. Ann. Rev. Genet. 26:479-506.

Fabbri, A., J.I. Hormaza and V.S. Polito. Random amplified polymorphic DNA analysis of olive (*Olea europaea* L.) cultivars. J. Am. Soc. Hort. Sci. 120:538-542.

FAO. 1993. Data from FAO world information and early warning system on plant genetic resources. (cited in Hintum).

Fairbanks, D.J., A. Waldrigues, C.F. Ruas, P.M. Ruas, P.J. Maughan, L.R. Robison, W.R. Andersen, C.R. Riede and C.S. Pauley. 1993. Efficient characterization of biological diversity using field DNA extraction and random amplified polymorphic DNA markers. Rev. Bras. Genet. 16:11-22.

Falk, D.A. 1990. Integrated strategies for conserving plant genetic diversity. Ann. Missouri Bot. Gard. 77:38-47.

Falk, D.A. and K.E. Holsinger (eds.) 1991. Genetics and Conservation of Rare Plants. Oxford University Press, New York.

Fielder, P.L. and S.K. Jain. (eds.) 1992. Conservation Biology: The Theory and Practice of Nature Conservation, Preservation and Management. Chapman and Hall, New York.

Figuera, A.J., J. Janick and P. Goldsborough. 1992. Genome size and DNA polymnorphism in *Theobroma cacao*. J. Am. Soc. Hort. Sci. 117:673-677.

Flavell, R. 1980. The molecular characterization and organization of plant chromosomal DNA sequences. Annu. Rev. Plant Physiol. 31:569-596.

Frankel, O.H. 1985. Genetic resources: The founding years. I. Early beginnings, 1961-1966. Diversity 7:26-29.

Frankel, O.H. 1986. Genetic resources: The founding years. II. The movement's constituent assembly. Diversity 8:30-32.

Frankel, O.H. and A.H.D. Brown. 1984. Plant genetic resources today: A critical appraisal. Pp. 249-257 *in* Crop Genetic Resources: Conservation and Evaluation (J.W.H. Holden and J.T. Williams, eds.) George Allen and Unwin, London, United Kingdom.

Frankel, O.H. and M.E. Soulé. 1981. Conservation and evolution. Cambridge University Press, Cambridge, United Kingdom.

Fregeau, C.J., R.M. Fourney. 1993. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. BioTechniques 15:304-309.

Jeffreys, A.J., A. MacLeod, K. Tamaki, D.L. Neil and D.G. Monckton. 1991. Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204-209.

Gall, G.A.E and M. Staton. 1992. Integrating conservation biology and agricultural production: conclusions. Agric. Ecosyst. Environ. 42:217-230.

Gall, G.A.E. and G.H. Orians. 1992. Agriculture and biological conservation. Agric. Ecosyst. Environ. 42:1-8.

Garvin, D.F. and N.F. Weeden. 1994. Isozyme evidence supporting a single geographic origin for domesticated tepary bean. Crop Sci. 34:1390-1395.

Gawel, N.J., R.L. Jarret and A. Whittemore. 1992. Restriction fragment length polymorphism (RFLP)- based phylogenetic analysis of *Musa*. Theor. Appl. Genet. 84:286-290.

Gepts, P. 1990. Biochemical evidence bearing on the domestication of Phaseolus (Fabaceae) beans. Econ. Bot. 44:28-38.

Gepts, P., T. Stockton and G. Sonnante. 1992. Use of hypervariable markers in genetic diversity studies. Pp. 41-45 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America/American Society for Horticultural Science/American Genetic Association. CSSA, Madison, WI, USA.

Gerdes, J.T. and W.F. Tracy. 1994. Diversity of historically important sweet corn inbreds as estimated by RFLPs, morphology, isozymes and pedigree. Crop Sci. 34:26-33.

Gizlice, Z., T.E. Carter, Jr. and J.W. Burton. 1993. Genetic diversity in North American soybean. I. Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614-620.

Goffreda, J., W.B. Burnquist, S.C. Beer, S.D. Tanksley and M.E. Sorrells. 1992. Application of molecular markers to assess genetic relationships among accessions of wild oat, *Avena sterilis*. Theor. Appl. Genet. 85:146-151.

Grattapaglia, D, J. Chaparro, P. Wilcox, S. McCord, D. Werner, H. Amerson, S. McKeand, F. Bridgewater, R. Whetten, D. O'Malley and R. Sederoff. 1992. Mapping in woody plants with RAPD markers: Application to breeding in forestry and horticulture. Pp. 37-40 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America/American Society for Horticultural Science/American Genetic Association. CSSA, Madison, WI, USA.

Gupta, V., G. Lakshmisita, M.S. Shaila, V. Jagannathan and M.S. Lakshmikumaran. 1992. Characterization of species-specific repeated DNA sequences from *B. nigra*. Theor. Appl. Genet. 84:397-402.

Hadrys, H., M. Balick and B. Schierwater. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Mol. Ecol. 1:55-63.

Haila, Y. and J. Kouki. 1994. The phenomenon of biodiversity in conservation biology. Ann. Zoo. Fennici. 31:5-18.

Hall, S.J.G. and J. Ruane. 1993. Livestock breeds and their conservation: A global overview. 1993. Conserv. Biol. 7:815-825.

Halward, T.M., H.T. Stalker, E.A. LaRue and G. Kochert. 1991. Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and related wild species. Genome 34:1013-1020.

Hamilton, M.B. 1994. *Ex situ* conservation of wild plant species: Time to reassess the genetic assumptions and implications of seed banks. Conserv. Biol. 8:39-49.

Hamrick, J.L., M.J.W. Godt, D.A. Murawski and M.D. Loveless. 1991. Correlations between species: Traits and allozyme diversity implications for conservation biology. Pp. 75-86 *in* Genetics and Conservation of Rare Plants (D.A. Falk and K.E. Holsinger, eds.). Oxford University Press, New York.

Hanelt, P. 1988. Taxonomy as a tool for studying plant genetic resources. Kulturpflanze 36:169-187.

Harada, T. K. Matsukawa, T. Sato, R. Ishikawa, M. Nizeki and K. Saito. 1993. DNA-RAPDs detect genetic variation and paternity in *Malus*. Euphytica 65:87-92.

Harlan, J.R. 1975. Our vanishing genetic resources. Science 188:618-621.

Harlan, J.R. and J.M.J. de Wet. 1971. Toward a rational classification of cultivated plants. Taxon 20: 509-517.

Havey, M.J. and F.J. Muehlbauer. 1989. Variability for restriction lengths and phylogenies in lentil. Theor. Appl. Genet. 77:839-843.

He, S. H. Ohm and S. MacKenzie. 1993 Detection of DNA sequence polymorphisms among the wheat varieties. Theor. Appl. Genet. 84:573-578.

Hedrick, P.W. and P.S. Miller. 1992. Conservation genetics: techniques and fundamentals. Ecol. Applic. 2:30-46.

Helentjaris, T. G. King, M. Slocum, C. Siedenstrong and S. Wegman. 1985. Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol. Biol. 5:109-118.

Hickey, R.J., M.A. Vincent and S.I Guttman. 1991. Genetic variation in running buffalo clover [*Trifolium stoloniferum* (Fabaceae)]. Conserv. Biol. 309-316.

Hintum, T.J.L. van and D. Haalman. 1994. Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare* s. lat.) Theor. Appl. Genet. 88:70-74.

Hintum, T.J.L. van. 1994. Drowning in the genepool: managing genetic diversity in genebank collections. Thesis, Swedish Univ. Agric. Sciences. Dept. Plant Breeding Research, S-269 31 Sval`f, Sweden.

Hodges, J. 1990. Conservation of animal genetic resources in developing countries. Pp. 128-145 *in* Genetic Conservation of Domestic Livestock (L. Alderson, ed.). CAB Intl., Oxon, UK.

Hodgkin, T. and D.G. DeBouck. 1992. Some possible applications of molecular genetics in the conservation of wild species for crop improvement. Pp. 153-182 *in* Conservation of Plant Genes: DNA Banking and *in vitro* Biotechnology (R.P. Adams and J.E. Adams, eds.). Academic Press, San Diego, CA, USA.

Hong, Y.P., V.D. Hipkins and S.H. Strauss. 1993. Chloroplast DNA diversity among trees, populations and species in the California closed-cone pines (*Pinus radiata, Pinus muricata,* and *Pinus attenuata*). Genetics 135:1187-1196.

Hormoza, J.I., L. Dollo and V.S. Polito. 1994. Determination of relatedness and geographical movements of *Pistacia vera* (Pistachio: *Anacardiaceae*) germplasm by RAPD analysis. Econ. Bot. 48:349-358.

Hu, J. and C.F. Quiros. 1991. Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep. 10:505-511.

Hubby, J.L. and R.C. Lewontin. 1966. A molecular approach to study of genetic heterozygosity in natural populations. Genetics 59:577-594.

Hughes, C.A. and D.C. Queller. 1993. Detection of highly polymorphic micosatellite loci in a species with little allozyme polymorphism. Mol. Ecol. 2:131-137.

Jarret, R.L. and D.F. Austin. 1994. Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas*) and related species as revealed by RAPD analysis. Genet. Resour. and Crop Evol. 41:165-173.

Jarret, R.L. and R.E. Litz. 1986. Isozymes as genetic markers in bananas and plantains. Euphytica 35:539-549.

Jarret, R.L. and N. Bowen. 1994. Simple sequence repeats (SSRs) for sweet potato germplasm characterization. Plant Genet. Resour. Newsl. 100:9-11.

Jung, C., K. Pillen, L. Freese, S. Fahr and A.E. Melchinger. 1993. Phylogenetic relationships between cultivated and wild species of the genus *Beta* revealed by DNA "fingerprinting." Theor. Appl. Genet. 86:449-457.

Kaemmer, D., R. Afza, K. Weising, G. Kahl and F.J. Novak. 1992. Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp). Bio/Tech. 10:1030-1035.

Karron, J.D. 1989. Breeding systems and levels of inbreeding depression in geographically restricted and widespread species of *Astragalus* (Fabaceae). Am. J. Bot. 76:331-1340.

Keim, P., W. Beavis, J. Schupp and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germpl plasm. Theor. Appl. Genet. 85:205-212.

Kesseli, R.V., I. Paran and R.W. Michelmore. 1992. Efficient mapping of specifically targeted genomic regions and the tagging of these regions with reliable PCR-based genetic markers. Pp. 31-40 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America/American Society for Horticultural Science/American Genetic Association. CSSA, Madison, WI, USA.

Kesseli, R., O. Ochoa and R. Michelmore. 1991. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce. Genome 34:430-437.

Kidwell, K.K., D.F. Austin and T.C. Osborn. 1994. RFLP evaluations of nine *Medicago* accessions representing the original germplasm sources for North American alfalfa cultivars. Crop Sci. 34:230-236.

Kijas, J.M.H., J.C.S. Fowler and M.R. Thomas. 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. Genome 38:349-355.

Klinger, T., P.E. Arriola and N.C. Ellstrand. 1992. Crop-weed hybridization in radish (*Raphanus sativus*): Effects of distance and population size. Am. J. Bot. 79:1431-1435.

Kochert, G, T. Halward, W.D. Branch and C.E. Simpson. 1991. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor. Appl. Genet. 81:565-570.

Kohn, L.M. 1992. Developing new characters for fungal systematics: an experimental approach for determining rank of resolution. Mycologia 84:139-154.

Koller, B., A. Lehmann, J.M. McDermott and C. Gessler. 1993. Identification of apple cultivars using RAPD markers. Theor. Appl. Genet. 85: 901-904.

Kresovich, S. 1992. Plant genetic resources conservation and use: an evolving paradigm. Field Crops Res. 29:183-184.

Kresovich, S. and J.R. McFerson. 1992. Assessment and management of plant genetic diversity: conservation of intra- and interspecific variation. Field Crops Res. 29:185-204.

Kresovich, S., J.G.K. Williams, J.R. McFerson, E.J. Routman and B.A. Schaal. 1992. Characterization of genetic identities and relationships of *Brassica oleracea* L. via a random amplified polymorphic DNA assay. Theor. Appl. Genet. 85:190-196.

Kresovich, S., W.F. Lamboy, R.G. Li, J.P. Ren, A.K. Szwewc-McFadden and S.M. Bliek. 1994. Application of molecular methods and statistical analyses for discrimination of accessions and clones of Vetiver grass. Crop Sci. 34:805-809.

Kresovich, S., W.F. Lamboy, A.K. Szewc-McFadden, J.R. McFerson and P.L. Forsline. 1993. Molecular diagnostics and plant genetic resources conservation. AgBiotech News Info. 5:255-258.

Kresovich, S., A.K. Szewc-McFadden, S.M. Bliek and J.R. McFerson. 1995. Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor. Appl. Genet. 91:206-211.

Lagercrantz, U., H. Ellegren and L. Andersson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res. 21:1111-1115.

Lakshmikumaran, M., S. Bhatia, S.S. Banga and S. Prakash. 1994. Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (*Brassica juncea*) and its relationship to heterosis. Theor. Appl. Genet. 88:116-122.

Lamboy, W.F. 1994a. Computing genetic similarity coefficients from RAPD data: the effects of PCR artifacts. PCR Meth. Applic. 4:31-37.

Lamboy, W.F. 1994b. Computing genetic similarity coefficients from RAPD data: correcting for the effects of PCR artifacts caused by variation in experimental conditions. PCR Meth. Applic. 4:38-43.

Lamboy, W.F., J.R. McFerson, A.L. Westman and S. Kresovich. 1994. Application of isozyme data to the management of the United States national *Brassica oleracea* L. genetic resources collection. Genet. Resour. Crop Evol. 41:99-108.

Lande, R. 1988. Genetics and demography in biological conservation. Science 41:1455-1460.

Landry, B.S., R.V.Kesseli, B. Farrara and R.W. Michelmore. 1987. A genetic linkage map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337.

Lashermes, P., P. Cros, P. Marmey and A. Charrier. 1993. Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. Genet. Resour. Crop Evol. 40:91-99.

Laurent, V., A.M. Risterucci and C. Lanaud. 1994. Genetic diversity in cocoa revealed by cDNA probes. Theor. Appl. Genet. 88:193-198.

Lavi, U., J. Hillel, A. Vainstein, E. Lahav and D. Sharon. 1991. Application of DNA fingerprints for identification and genetic analysis of avocado. J. Am. Soc. Hort. Sci. 116:1078-1081.

Lebot, V., K.M. Aradhya, R. Manshardt and B. Meilleur. 1993. Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. Euphytica 67:163-175.

Ledig, F.T. 1988. The conservation of diversity in forest trees. BioScience 38:471-479.

Lefort-Buson, M., Y. Hebert and C. Damerval. 1988. Tools for assessment of genetic and phenotypic diversity. Agronomie (Paris) 8:173-178.

Lessa, E.P. and G. Applebaum. 1993. Screening techniques for detecting allelic variation in DNA sequences. Mol. Ecol. 2:119-129.

Li, P., J. MacKay and J. Bousquet. 1992. Genetic diversity in Canadian hardwoods: implications for conservation. For. Chron. 68:709-719.

Liu, Z.W. and G.R. Furnier. 1993. Comparisons of allozyme, RFLP and markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theor. Appl. Genet. 87:97-105.

Liu, Z.W., R.L. Jarret, S. Kresovich and R.R. Duncan. 1994. Genetic relationships and variation among ecotypes of seashore paspalum (*Paspalum vaginatum*) determined by random amplified polymorphic DNA markers. Genome 37:1001-1007.

Liu, Z.W., R.L. Jarret, S. Kresovich and R.R. Duncan. 1995. Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum*). Theor. Appl. Genet. 91:47-52.

Loukas, M. and C.B. Krimbas. 1983. History of olive cultivars based on their genetic distances. J. Hort. Sci. 58:121-127.

Lu, J. and B. Pickersgill. 1993. Isozyme variation and species relationships in peanut and its wild relatives (*Arachis* L. - Leguminosae). Theor. Appl. Genet. 85:550-560.

Lynch, M. 1991. Analysis of population genetic structure by DNA fingerprinting. Pp. 113-126 *in* DNA Fingerprinting: Approaches and Applications (T. Burke *et al.*, eds.). Birkhauser Verlag, Berlin, Germany.

Lynch, M. and B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Mol. Ecol. 3:91-99.

Mace, G.M. and R. Lande. Assessing extinction threats: toward a reevaluation of IUCN threatened species categories. Conserv. Biol. 5:148-157.

Mackill, D.J. 1995. Classifying japonica rice cultivars with RAPD markers. Crop Sci. 35:889-874.

Mailer, R.J., R. Scarth and B. Fristensky. 1994. Discrimination among cultivars of rapeseed (*Brassica napus* L.) using DNA polymorphisms amplified from arbitrary primers. Theor. Appl. Genet. 87:697-704.

Margale, E., Y. Herve, J. Hu and C.F. Quiros. 1993. Characterization by RAPD markers of a local cole crop cultivar collection from France: optimal sample size and number of markers. Euphytica (In press).

Matlick, J.S., E.M. Ablett and D.L. Edmonson. 1992. The gene library - preservation and analysis of genetic diversity in Australasia. Pp. 153-182 *in* Conservation of Plant Genes: DNA Banking and *in vitro* Biotechnology (R.P. Adams and J.E. Adams, ed.). Academic Press, San Diego, CA, USA.

May, R.M. 1994. Conceptual aspects of the quantification of biological diversity. Phil. Trans. R. Soc. London 345:13-20.

Mayer, M.S. and P.S. Soltis. 1994. Chloroplast DNA phylogeny of *Lens* (Leguminoseae): Origin and diversity of cultivated lentil. Theor. Appl. Genet. 87:773-781.

McGrath, J.M. and C.F. Quiros. 1992. Genetic diversity at isozyme and RFLP loci in *Brassica campestris* as related to crop type and geographical origin. 1992. Theor. Appl. Genet. 83:783-790.

McNeely, J.A. and R.B. Norgaard. 1992. Developed country policies and biological diversity in developing countries. Agric. Ecosyst. Environ. 42:194-204.

Melchinger, A.E., M.M. Messmer, M. Lee, W.L. Woodam and K.R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. 31:669-678.

Melchinger, A.E., A. Graner, M. Singh and M.M. Messmer. 1994. Relationships among European barley cultivars.I. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci. 34:1191-1199.

Mellersh, C. and J. Sampson. 1993. Simplifying detection of microsatellite length polymorphisms. Biotech 15:582-584.

Menges, E.S. 1990. Population viability analysis for an endangered plants. Conserv. Biol. 4:148-157.

Michelmore, R.W., I. Paron and R.V. Kesseili. 1991. Identification of markers linked to disease-resistant genes by bulk segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Nat. Acad. Sci, USA 88:9828-9832.

Miller, J.C. and S.D. Tanksley. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. 80:437-448.

Milligan, B.G. 1991. Chloroplast DNA diversity within and among populations of *Trifolium pratense*. Curr. Genet. 19:411-416.

Milligan, B.G., J. Leebens-Mack and A.E. Strand. 1994. Conservation genetics: beyond the maintenance of marker diversity. Mol. Ecol. 3:423-435.

Morgante, M. and A.M. Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. 3:175-182.

M'Ribu, H.K. and K.W. Hilu. 1994. Detection of interspecific and intraspecific variation in *Panicum* millets through random amplified polymorphic DNA. Theor. Appl. Genet. 88:412-416.

Murty, B.R. and M.H. Mengesha. 1990. World germplasm collections and their potential in crop productivity. Ind. J. Agric. Sci. 60:787-792.

Mussler, A, K.N. Egger and G.A. Hughes. 1992. Low levels of genetic diversity in red pine confirmed by random amplified polymorphic DNA markers. Can. J. For. Res. 22:1332-1337.

Nienhuis, J., M.K. Slocum, D.A. DeVos and R. Muren. 1993. Genetic similarity among *Brassica oleracea* L. genotypes as measured by restriction fragment length polymorphisms. J. Am. Hort. Sci. 118:298-303.

Novy, R.G., C. Kokak, J. Goffreda and N. Vorsa. 1994. RAPDs identify varietal misclassification and regional divergence in cranberry [*Vaccinium macrocarpon* (Ait.) Pursh.]. Theor. Appl. Genet. 88:1004-1010.

Nybom, H. and B.A. Schaal. 1990. DNA "fingerprints" reveal genotypic distributions in natural populations of blackberries and raspberries (*Rubus*, Rosaceae). Am. J. Bot. 77:883-888.

Nybom, H. and B.A. Schaal. 1990. DNA "fingerprints" applied to paternity analysis in apples (*Malus x domestica*). Theor. Appl. Genet. 79:763-768.

Nybom, H., S.H. Rogstad and B.A. Schaal. 1990. Genetic variation detected by use of the M13 DNA fingerprint probed in *Malus*, *Prunus*, and *Rubus* (*Rosaceae*). Theor. Appl. Genet. 79:153-156.

O'Brien, S. J. 1994. A role for molecular genetics in biological conservation. Proc. Nat. Acad. Sci. USA. 91:5748-5755.

Ocampo, C., C. Hershey, C. Inglesias and M. Iwanaga. 1993. Esterase isozyme fingerprinting of the cassava germplasm collection held at CIAT. Pp. 81-92 *in* First international scientific meeting of the Cassava Biotechnology Network (W.M. Roca and A.M. Thro, eds.). Proceedings 25-28 Aug. 1992, Cartagena, Colombia. CIAT, Cali, Colombia.

Olivier, I. and A.J. Beattie. 1993. A possible method for the rapid assessment of biodiversity. Conserv. Biol. 7:562-568.

Orozco-Castillo, C., K.J. Chalmers, R. Waugh and W. Powell. 1994. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor. Appl. Genet. 87:934-940.

Ouazzani, N., R. Lumaret, P. Villemur and F. DiGiusto. 1993. Leaf allozyme variation in cultivated and wild olive trees. J. Hered. 84:34-42.

Pammi, S., K. Schertz, G. Xu, G. Hart and J.E. Mullet. 1994. Random-amplified-polymorphic DNA markers in sorghum. Theor. Appl. Genet. 89:80-88.

Peacock, W.J. 1989. Molecular biology and genetic resources. Pp. 363-376 *in* The Use of Plant Genetic Resources (A.H.D. Brown, O.H. Frankel, D.R. Marshall and J.T. Williams, eds.). Cambridge University Press, Cambridge, United Kingdom.

Phippen, W.B. 1994. Plant genetic identity and relatedness testing to improve genetic resources conservation. MS Thesis, Cornell University.

Plucknett, D.L. and M.E. Horne. 1992. Conservation of genetic resources. Agric. Ecosyst. Environ. 42:75-92.

Polans, N.O. and R.W. Allard. 1989. An experimental evaluation of the recovery potential of ryegrass populations from genetic stress resulting from restriction of population size. Evolution 43:1320-1324.

Pontikis, C.A., M. Loukas and G. Kousounis. 1980. The use of biochemical markers to distinguish olive cultivars. J. Hortic. Sci. 55:333-343.

Poulsen, G.B., G. Kahl and K. Weising. 1993. Abundance and polymorphism of simple repetitive DNA sequences in *Brassica napus* L. Theor. Appl. Genet. 85:994-1000.

Prescott-Allen, R. and C. Prescott-Allen. 1990. How many plants feed the world? Conserv. Biol. 4:365-374.

Queller, D.C., J.E. Strassmann and C.E. Hughes. 1993. Microsatellites and kinship. Trends Ecol. Evol. 8:285-288.

Rafalski J.A. and S.V. Tingey. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet. 9:275-279.

Rao, V.R. 1991. Problems and methodologies for management and retention of genetic diversity in germplasm collections. *In* ASTAF/IBPGR Workshop on Conservation of Plant Genetic Resources (B. Becker, ed.). Proceedings, Bonn, Germany. ATSAF/IBPGR.

Rao, V.R. and K.W. Riley. 1994. The use of biotechnology for conservation and utilization of plant genetic resources. Plant Genet. Resour. Newsl. 97:3-20.

Rafalski, J.A. and S.V. Tingey. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet. 9:275-279.

Ren, J., W.F. Lamboy, J.R. McFerson and S. Kresovich. 1995. Characterization of genetic identities and relationships among Chinese vegetable brassicas using random amplified Polymorphic DNA markers. J. Am. Soc. Hort. Sci. 120:548-555.

Resurrecion, A.P., C.P Villareal, A. Parco, G. Second and B.O. Juliano. 1994. Classification of cultivated rices into indica and japonica types by the isozyme, RFLP and two milled-rice methods. Theor. Appl. Genet. 89:14-18.

Revilla, P. and W.F. Tracy. 1995. Isozyme variation and phylogenetic relationships among open-pollinated sweet corn cultivars. Crop Sci. 35:219-227.

Rick, C.M. 1988. Molecular markers as aids in germplasm management and use in *Lycopersicon*. Hort. Sci. 23:55-57.

Rieseberg, L.H. and G.J. Seiler. 1990. Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annus*, Asteraceae). Econ. Bot. 44:79-91.

Rieseberg, L.H., S.M. Beckstrom-Sternberg, A. Liston and D.M. Arias. 1991. Phylogenetic and systematic inferences from chlorplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae). Syst. Bot. 16:50-76.

Riggs, L.A. 1990. Conserving genetic resources on-site in forest ecosystems. Forest Ecol. Manage. 35:45-68.

Rogers, S.O. 1994. Phylogenetic and taxonomic information from herbarium and mummified DNA. Pp. 47-67 *in* Conservation of Plant Genes II: Utilization of Ancient and Modern DNA (R.P. Adams, J.S. Miller, E.M. Golenberg and J.E. Adams, eds.). Missouri Botanical Garden, St. Louis, MO.

Rogstad, S.H., J.C. Patton and B.A. Schaal. 1988. M13 repeat probe detects DNA minisatellite-like sequences in gymnosperms and angiosperms. Proc. Nat. Acad. Sci. (USA) 85:9176-9178.

Rongwen, J., M.S. Akkaya. A.A. Bhagwat and P.B. Cregan. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. 90:43-48.

Ronning, C.M. and R.J. Schnell. 1994. Allozyme diversity in a collection of *Theobroma cacao* L. J. Hered. 85:291-295.

Ronning, C.M., R.J. Schnell and D.N. Kuhn. 1995. Inheritance of random amplified polymorphic DNA (RAPD) markers in *Theobroma cacao* L. J. Am. Soc. Hort. Sci. 120:681-685.

Roper, M.M. 1993. Biological diversity of micro-organisms: An Australian perspective. Pacific Conserv. Biol. 1:21-28.

Rumbaugh, M.D., W.L. Graves, J.L. Caddel and R.M. Mohammad. 1988. Variability in a collection of alfalfa germplasm from Morocco. Crop Sci. 28:605-609.

Russell, J.B., F. Hosein, E. Johnson, R. Waugh and W. Powell. 1993. Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol. Ecol. 2:89-97.

Saghai-Maroof, M.A., R.W. Allard and Q. Zhang. 1991. Genetic diversity and ecogeographical differentiation among ribosomal DNA alleles in wild and cultivated barley. Proc. Nat. Acad. Sci. USA 87:8486-8490.

Saghai-Maroof, M.A., R.M. Biyashev, G.P. Yang, Q. Zhang and R.W. Allard. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proc. Nat. Acad. Sci. USA 91:5466-5470.

Salwasser, H. 1990. Conserving biological diversity: A perspective on scope and approaches. For. Ecol. Manage. 35:79-90.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular cloning: a laboratory manual (2nd ed.) Pp. 1.6-1.73. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sandlund, O.T., K. Hindar and A.H.D. Brown. 1992. Conservation of biodiversity for sustainable development. Scandinavian University Press, Oslo, Norway.

Santos, J.B. dos, J. Nienhuis, P. Skroch, J. Tivang and M.K. Slocum. 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. Theor. Appl. Genet. 87:909-915.

Sarkar, R. and S.N. Raina. 1992. Assessment of genome relationships in the genus *Oryza* L. based on seed-protein profile analysis. Theor. Appl. Genet. 85:127-132.

Schaal, B.A., W.J. Leverich and S.H. Rogstad. 1991. A comparison of methods for assessing genetic variation in plant conservation biology. Pp. 123-134 *in* Genetics and Conservation of Rare Plants (D.A. Falk and K.E. Holsinger, eds.). Oxford University Press, New York, New York.

Sederoff, R., D. Grattapalia, P. Wilcox, J. Chaparro, D. O'Malley, S. McCord, R. Whetten, L. McIntyre and B. Weir. 1992. Use of PCR-based RAPD markers for genetic mapping in conifers. *In* Abstracts of papers Am. Chem. Soc., 203rd ACS National Meeting, San Francisco, CA. 5-10 April 1992. BTEC73.

Shands, H.L. 1990. Plant genetic resources conservation: the role of the genebank in delivering useful materials to the research scientists. J. Hered. 81:7-10.

Shands, H.L. 1991. Complementarity of *in situ* and *ex situ* germplasm conservation from the standpoint of the future user. Isr. J. Bot. Basic Appl. Plant Sci. 40:521-528.

Siedler, H., M.M. Messmer, G.M. Schachermayr, H. Winzeler and B. Keller. 1994. Genetic diversity in European wheat and spelt breeding material based on RFLP data. Theor. Appl. Genet. 88:994-1003.

Singh, S.P., J.A. Gutierrez, A. Molina and P. Gepts. 1991. Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci. 31:23-29.

Skroch, P., J. Tivang and J. Nienhuis. 1992. Analysis of genetic relationships using RAPD marker data. Pp. 26-30 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America/American Society for Horticultural Science/American Genetic Association. CSSA, Madison, WI, USA.

Smith, J.S.C. 1994. Recent advances in the use of PCR based technology for DNA fingerprinting in plants. Pp. 101-112 *in* Conservation of Plant Genes II: Utilization of Ancient and Modern DNA (R.P. Adams, J.S. Miller, E.M. Golenberg and J.E. Adams, eds.). Missouri Botanical Garden, St. Louis, MO.

Smith, J.S.C. and E.C.L. Chin. 1992. The utility of random primer-mediated profiles, RFLPs and other technologies to provide useful data for varietal protection.Pp. 45-49 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America/American Society for Horticultural Science/American Genetic Association. CSSA, Madison, WI, USA.

Smith, J.S.C. and O.S. Smith. 1989a. The description and assessment of distances between inbred lines of maize: I. The use of morphological traits as descriptors. Maydica 34:141-150.

Smith, J.S.C. and O.S. Smith. 1989b. The description and assessment of distances between inbred lines of maize: II. The use of utility of morphological, biochemical and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. Maydica 34:151-161.

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg and S.J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, heterosis and RFLPs. Theor. Appl. Genet. 80:833-840.

Sobral, B.W.S. and R.J. Honeycutt. 1993. High output genetic mapping of polyploids using PCR-generated markers. Theor. Appl. Genet. 86:105-112.

Soleri, D. and S.E. Smith. 1995. Morphological and phenological comparisons of two Hopi maize varieties conserved *in situ* and *ex situ*. Econ. Bot. 49:56-77.

Soller, M. and J.S. Beckmann. 1983. Genetic polymorphism in varietal identification and genetic improvement. Theor. Appl. Genet. 67:25-33.

Song, K.M., T.C. Osborn and P.H. Williams. 1988. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 2. Preliminary analysis of subspecies within *B. rapa* (syn. *campestris*) and *B. oleracea*. Theor. Appl. Genet. 76:593-600.

Song, K.M., T.C. Osborn and P.H. Williams. 1990. *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor. Appl. Genet. 79:497-506.

Souza, E., P.N. Fox, D. Byerlee and B. Skovmund. 1994. Spring wheat diversity in irrigated areas of two developing countries. Crop Sci. 34:774-783.

Spooner, D.M., G.J. Anderson and R.K. Jansen. 1993. Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes and pepinos (Solanaceae). Am. J. Bot. 80:676-688.

Stein, D.B. 1993. Isolating and comparing nucleic acids from land plants: Nuclear and other organellar genes. *In* Molecular Evolution: Producing the Biochemical Data (E.A. Zimmer, T.J. White, R.L. Cann and A.C. Wilson, eds.). Meth. Enzymol. 224:153-167.

Steiner, J.J. and C.J. Poklemba. 1994. *Lotus corniculatus* classification by seed globulin polypeptides and relationship to accession pedigrees and geographic origin. Crop Sci. 34:255-264.

Steiner, J.J., C.J. Poklemba, R.G. Fjellstrom and L.F. Elliott. 1995. A rapid, one-tube genomic DNA extraction process for PCR and RAPD analysis. Nucleic Acids Res. 23:2569-2570.

Stiles, J.I., L. Lemme, S. Sondur, M.B. Morshidi and R. Manshardt 1993. Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. Theor. Appl. Genet. 85:697-701.

Stockton, T. and P. Gepts. 1994. Identification of DNA probes that reveal polymorphisms among closely related *Phaseolus vulgaris* lines. Euphytica 76:177-183.

Strauss, S.H., J. Bousquet, V.D. Hiplins and Y.P. Hong. 1992. Biochemical and molecular genetic markers in biosystematic studies of forest trees. New For. 6:1255-1258.

Sytsma, K.J. 1994. DNA extraction from recalcitrant plants: Long, pure and simple? Pp. 69-81 *in* Conservation of plant genes II: Utilization of Ancient and Modern DNA (R.P. Adams, J.S. Miller, E.M. Golenberg and J.E. Adams, eds.). Missouri Botanical Garden, St. Louis, MO.

Tao, Y., J.M. Manner, M.M. Ludlow and R.G. Henzell. 1993. DNA polymorphisms in grain sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. 86:679-688.

Templeton, A.R. 1991. Genetics and conservation biology. Pp. 15-30 *in* Advances in Life Sciences: Species Conservation: A Population-Biological Approach (A. Seitz and V. Loeschcke, eds.). Birkhauser Verlag, Basel, Switzerland, Mainz, Germany.

Templeton, A.R. 1994. Biodiversity at the molecular genetic level: Experiences from disparate macroorganisms. Phil. Trans. R. Soc. London 345:59-64.

Thomas, M.R., S. Matsumoto, P. Cain and N.S. Scott. 1993. Repetitive DNA of grapevine: classes present and sequences suitable for cultivar identification. Theor. Appl. Genet. 86:173-180.

Thomas, M.R. and Scott N.S. 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analyzed as sequence-tagged sites (STSs). Theor. Appl. Genet. 86:985-990.

Thormann, C.E. and T.L. Osborn. 1992. Use of RAPD and RFLP markers for germplasm evaluation. Pp. 9-11 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America/American Society for Horticultural Science/American Genetic Association. CSSA, Madison, WI, USA.

Thormann, C.E., M.E. Ferreira, L.E.A. Camargo, J.G. Tivang and T.C. Osborn. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor. Appl. Genet. 88:973-980.

Tingey, T.V., J.A. Rafalski and J.G.K. Williams. 1993. Genetic analysis with RAPD markers. Pp 3-11 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America / American Society for Horticultural Science / American Genetic Association. CSSA, Madison, WI, USA.

Tinker, N.A., A. Vermunt, R. Weide, T. Liharska and P. Zabel. 1993. Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor. Appl. Gent. 85:976-984.

Transue, D.K., D.J. Fairbanks, L.R. Robison and W.R. Andersen. 1994. Species identification by RAPD analysis of grain amaranth genetic resources. Crop Sci. 34:1385-1389.

Trujillo, I., R. Rallo, E.A. Carbonell and M.J. Asins. 1990. Isoenzymatic variability of olive cultivars according to their origin. Acta Hortic. 286:137-140.

Vaillancourt, R.E., N.F. Weeden and J. Barnard. 1993. Isozyme diversity in the cowpea species complex. Crop Sci. 33:606-613.

Vierling, R. and H.T. Nguyen. 1992. Use of RAPD markers to determine the genetic diversity of diploid, wheat genotypes. Theor. Appl. Genet. 84:835-838.

Vierling, R.A., Z. Xiang, C.P. Joshi, M.L. Gilbert and H.T. Nguyen. 1994. Genetic diversity among elite *Sorghum* lines revealed by restriction fragment length polymorphisms and random amplified polymorphic DNAs. Theor. Appl. Gent. 87:816-820.

Virk, P.S., B.V. Ford-Lloyd, M.T. Jackson and H.J. Newbury. 1995. Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74:170-179.

Virk, P.S., H.J. Newbury, M.T. Jackson and B.V. Ford-Lloyd. 1995. The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor. Appl. Genet. 90:1049-1055.

Vogler, A.P. and R. Desalle. 1994. Diagnosing units of conservation management. Conserv. Biol. 8:354-363.

Voysest, O.M.C. Valencia and M.C. Amezquita. 1994. Genetic diversity among Latin American Andean and mesoamerican common bean cultivars. Crop Sci. 34:1100-1110.

Wang, Z., J.L. Weber, G. Zhong and S.D. Tanksley. 1994. Survey of plant short tandem repeats. Theor. Appl. Genet. 88:1-6.

Waugh, R. and W. Powell. 1992. Using RAPD markers for crop improvements. Trends Biotechnol. 10:186-191.

Waycott, W. and S.B. Fort. 1994. Differentiation of nearly identical germplasm accessions by a combination of molecular and morphologic analyses. Genome 37:577-583.

Weatherhead, P.J. and R.D. Montgomerie. 1991. Good news and bad news about DNA fingerprinting. Trends Ecol. Evol. 6:173-174.

Weeden, N.F., G.M. Timmeran, M. Hemmat, B.E. Keen and M.A. Lodhi. 1992. Inheritance and reliability of RAPD markers. Pp. 12-17 *in* Applications of RAPD Technology to Plant Breeding. Proceedings Symposium Crop Science Society of America / American Society for Horticultural Science / American Genetic Association. CSSA, Madison, WI, USA.

Weining, S. and P. Langridge. 1991. Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. Theor. Appl. Genet. 82:209-216.

Weising, K., J. Ramser, D. Kaemmer, G. Kahl and J.T. Epplen. 1991. Oligonucleotide fingerprinting in plants and fungi. Pp. 313-329 *in* DNA Fingerprinting: Approaches and Applications (T. Burke *et al.*, eds.). Birkhauser Verlag, Berlin, Germany.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:7213-7218.

Welsh, J. and M. McClelland. 1991. Genomic fingerprinting using arbitrary primed PCR and a matrix of pairwise combinations of primers. Nucleic Acids Res. 19:5275-5279.

Widen, B. and S. Andersson. 1993. Quantitative genetics of life-history and morphology in a rare plant, *Senecia integrifolius*. Heredity 81:277-289.

Wilde, J., R. Waugh and W. Powell. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. Theor. Appl. Genet. 83:871-877.

Williams, J.T. 1991. Plant genetic resources: Some new directions. Pp. 61-92 *in* Advances in Agronomy (N.C. Brady, ed.). Academic Press, San Diego, CA, USA.

Williams, C.E. and D.A. St. 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., A.R. Kubelik, K.J. Livak, J.A. Rafalski and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Wilson, H.D., J.F. Doebley and M. Duvall. 1992. Chloroplast DNA diversity among wild and cultivated members of *Cucurbita* (Cucurbitaceae). Theor. Appl. Genet. 84:859-865.

Woodruff, D.S. and G.A.E. Gall. 1992. Genetics and conservation. Agric. Ecosyst. Environ. 42:53-73.

Yang, G.P., M.A. Sagahai-Maroof, C.G. Xu, O Zhang and R. Biyashev. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars in rice. Mol. General Genet. 245:187-194.

Yu, K. and K.P. Pauls. 1993. Rapid estimation of genetic relatedness among heterogeneous populations of alfalfa by random amplification of bulked genomic DNA samples. Theor. Appl. Genet. 86:788-794.

Zentgraf, U., K. King and V. Hemleben. 1992. Repetitive sequences are valuable as molecular markers in studies of phylogenetic relationships within the genus *Cucumis*. Acta Botanica Neerlandica 41:397-406.

Zhang, Q., M.A. Sagahi Maroof, T.Y. Lu and B.Z. Shen. 1992. Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. Theor. Appl. Genet. 83:495-499.

Zhang, Q., M.A. Sagahi Maroof and A. Kleinhofs. 1993. Comparative diversity ananysis of RFLPs and isozymes within and among populations of *Hordeum vulgare* ssp. *spontaneum*. Acta Bto. Neerlandica 41:397-406.

Zhang, Q., G.P. Yang, X.K. Dai and J.Z. Sun. 1994. A comparative analysis of genetic polymorphisms in wild and cultivated barley from Tibet using isozyme and ribosomal DNA markers. Genome 37:631-638.

Zhao, X., G. Kochert. 1993. Phylogenetic distribution and genetic mapping of a $(GGC)_n$ 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 and S.R. Diehl. 1992. Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026-10.

The use of molecular markers in the study of genetic diversity in rattan: preliminary results

Suchitra Changtragoon[1], *Alfred E. Szmidt*[2] *and Xiao-Ru Wang*[2]
[1]DNA and Isoenzyme Laboratory, Silvicultural Research Division, Royal Forest Department, Chatuchak, Bangkok 10900, Thailand
[2]Molecular Population Genetics Laboratory, Department of Forest Genetics and Plant Physiology, The Swedish University of Agricultural Sciences, S-901 83 Umea, Sweden

Introduction

Diminishing forest cover and uncontrolled exploitation have seriously depleted wild populations of rattans throughout Southeast Asia to the point where the rattan trade in some countries, including Thailand, is now in danger of collapse and this traditional and important source of income for rural people is likely to disappear (Dransfield 1987). Furthermore, the reproductive biology and ecology of species have been disturbed and extinction of species cannot be ruled out in the near future (Pimentel *et al.* 1986; WRI 1990; Pimentel *et al.* 1992). There is, therefore, a need to conserve the remaining gene pool. However, sampling guidelines for *in situ* and *ex situ* conservation cannot be effectively formulated due to the general lack of information on the genetic structure and diversity of populations, and the mating system of most tropical forest species.

Genetic marker screening is based on the survey of genetic diversity as revealed by variation at specific gene loci and provides information about the amount and distribution of genetic diversity within and among populations. Furthermore, analysis of gene marker data enables estimation of the mating system and monitoring of genetic changes caused by factors affecting the reproductive biology of a species. Information gained from genetic marker screening is invaluable for identifying populations desirable for forestry practices which inadvertently alter the natural gene pools of domesticated species (Changtragoon and Szmidt 1993).

Since rattans are commercially important in Southeast Asia, the genetic resources of these species need to be conserved. There are about 50 species of rattan in Thailand occurring in swamp, evergreen, dry evergreen and mixed deciduous forests (Dransfield 1985). Since the genetic resources of rattan are seriously depleted in Thailand, it is necessary to find an efficient way of conserving these invaluable genetic resources. To achieve this, we should try to evaluate these genetic resources as quickly as possible using molecular markers and integrating genetic assessment into efficient management and conservation plans.

Materials and methods

Materials

The material for this study came from ten species of rattan (Table 1). The leaves of five seedlings of each species and 30 mature plants of *Calamus palustris* were collected and used for DNA extraction.

Table 1: List of rattan species investigated

No.	Species	Common name
1	*Calamus palustris*	Kring
2	*C. longisetus*	Kampuan
3	*C. rudentum*	Keesean
4	*C. manan*	Kordam
5	*C. peregrinus*	Nguay
6	*C. caesius*	Takathong
7	*C. oxleyanus*	Dam
8	*C. latifolius*	Phong
9	*Daemonorops angustifolia*	Nam
10	*D. didymophylla*	Kepaet

Methods

DNA extraction

Total DNA was extracted as described by Doyle and Doyle (1990) and suspended in TE buffer (0.1 mM EDTA).

DNA Amplification and RAPD (Random amplification of polymorphic DNA)

The optimal reaction for RAPD analysis was set up under the following conditions: 1 x reaction buffer, 0.5 U Taq DNA polymerase (Pharmacia), 0.3 uM of the 10-mer random primer, 150 uM dNTPs and 25-50 ng template DNA for total volume of 25 ul (Lu *et al.* 1995). Amplification conditions were set up using a programmable thermocycler PTC 100 (MJ Research). The arbitrary primer kits A were purchased from Operon Technologies (Alameda, CA, USA). A total of 20 primers (OPA01 to 20) were screened in this study.

The amplification products were separated on 1.5 percent agarose gels in 0.5xTBE buffer. The banding patterns were visualized under UV light and photographed using a Polaroid camera. One kilobase ladder (BRL) was used as DNA standard.

Results

Primer screening

To identify primers that detect polymorphism, 20 primers from the OPA kit were screened on the total DNA obtained from the leaf tissues of mature plants of *C. palustris*. Of these 20 primers, ten failed to yield amplification products. The remaining ten (OPA03, OPA04, OPA09, OPA11, OPA12, OPA13, OPA15, OPA16, OPA17 and OPA20) yielded reproducible fragments and at least 61 loci were scorable. The size of the fragments ranged from 300 to 3500 bp. Figures 1, 2 and 3 show the fragments in *C. palustris* amplified DNA obtained with OPA09, OPA16 and OPA17 primers respectively. The ten primers will be used to further study genetic diversity in *C. palustris*.

Species specific RAPD markers

Based on preliminary results of primer screening, four primers(OPA 09, OPA11, OPA13 and OPA17) which gave good amplification products in *C. palutris* were selected to amplify the total DNA of the nine other rattan species. Figure 4 shows the specific RAPD patterns obtained in the nine species using the OPA13 primer.

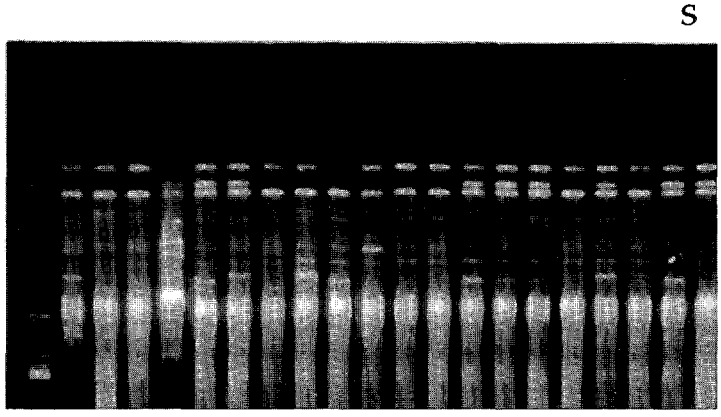

Fig. 1. Polymorphic RAPD fragments amplified by OPA09 primer in *Calamus palustris*. S: 1 kb ladder.

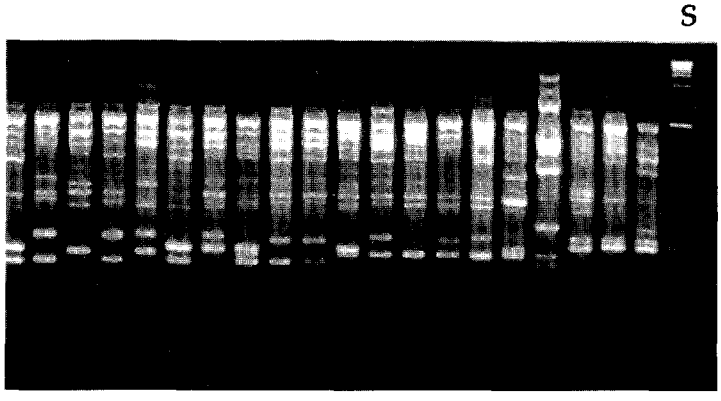

Fig. 2. Polymorphic RAPD fragments amplified by OPA16 primer in *Calamus palustris*. S: 1 kb ladder.

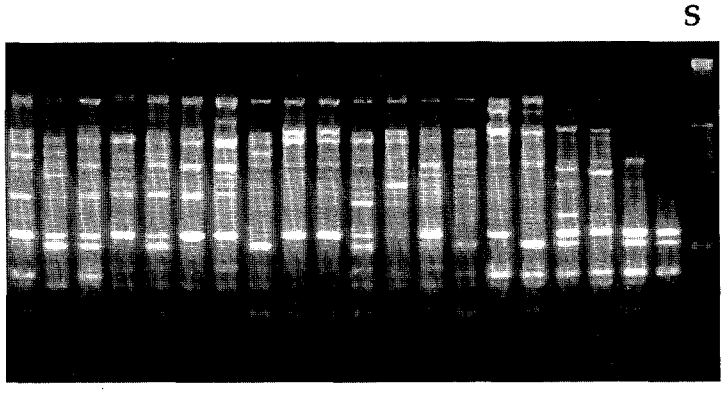

Fig. 3. Polymorphic RAPD fragments amplified by OPA17 primer in *Calamus palustris*. S: 1 kb ladder.

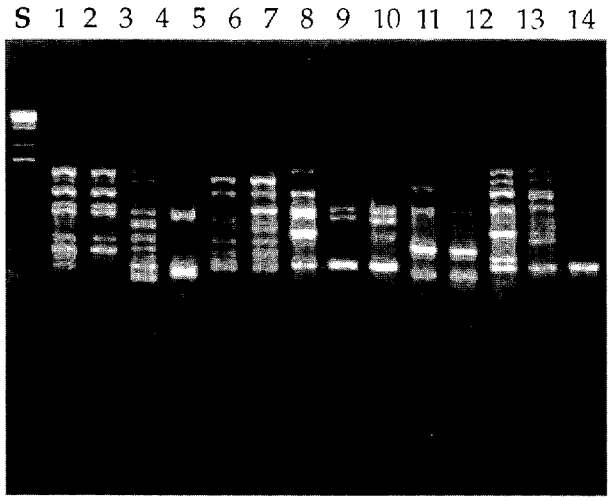

Fig. 4. RAPD fragments amplified by OPA13 primer in ten species of rattan, showing species-specific polymorphism: Lane 1&2: *Calamus manam*, 3: *C. longisetus*, 4: *C. peregrinus*, 5&6: *Daemophylla didymophylla*, 7: *C caesius*, 8&9: *C. rudentum*, 10&11: *C. oxleyanus*, 12&13: *D. angustifolia*, 14: *C. latifolia*, S: 1 kb ladder.

Discussion and plans for future research

Since the present project on the genetic diversity of *C. palustris* was started recently, we can only show some preliminary results of using RAPD markers in rattan. However, these results can be used to further study the genetic diversity of *C. palustris* and other species of rattan and also to identify some rattan species which are difficult to identify using morphological traits.

We plan to survey an additional five to eight populations of *C. palustris* and collect more materials both from mature plants and seeds from plants in different parts of the south of Thailand. Two of these populations have already been analyzed and the data is currently being evaluated by Dr A. Szmidt.

We shall use RAPD and SAPs (Specific amplicon polymorphisms) as well as allozyme markers for evaluating the genetic diversity of this species. The mating system in *C. palustris* will also be investigated using allozyme marker analysis of single plant seedlings. The genetic resources of *C. palustris* obtained to date will also be evaluated. Gene conservation strategies of this species will be designed using, in part, the information obtained from this study.

Acknowledgements

We thank Mr Meng-Zhu Lu, Department of Forest Genetics and Plant Physiology for fruitful discussions. We also thank Mr Chanatip Kuldelok and Mr Chanya Jareanrattawong for helping in collecting materials. This study was supported financially by the International Plant Genetic Resources Institute (IPGRI).

References

Changtragoon, S. and A.E. Szmidt. 1993. An Integrated population genetic approach to conserve forest tree diversity in Thailand. Pp. 217-224 *in* Proceedings of the ASEAN Seminar on Management and Conservation of Biodiversity, Kuala Lampur, Malaysia, 29 November-December 1993 (H.E. Hassan, D. Miedenger, A. Mirza and L.K. Leng, eds.).

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Dransfield, J. 1985. Prospects for lesser known canes. Pp. 107-114 *in* Proceedings of the Rattan Seminar, Kuala Lumpur, Malaysia. The Rattan Information Centre, Kepong, Malaysia.

Dransfield, J. 1987. The conservation status of rattan in 1987: a cause for great concern. Pp. 6-10 *in* Recent Research on Rattans (A.N. Rao, I. Vongkaluang, J. Dransfield, N. Monokaran, B.C. Sastry and G. Dhanarjan, eds.). Proceedings of the International Rattan Seminar, November 12-14, 1987, Chiangmai, Thailand.

Lu, M-Z., A.E. Szmidt and X-R. Wang. 1994. Inheritance of RAPD frangments in haploid and diploid tissues of *Pinus sylvestris* (L). Heredity 74:582-589.

Pimentel, D., U. Stacow, D.A. Takacs, H.W. Brubaker, A.R. Dumas. J.J. Meaney, J.A.S. Oniel, D.E. Onsi and D.B. Corzitlius. 1992. Conservation biological diversity in agricultural forestry system: most biological diversity exists in human-managed ecosytems. Biosciences 42:354-362.

Pimentel, D., W. Dazhong, S. Eigenbrode, H. Lang, D. Emerson and M. Karasik. 1986. Deforestation: interdependency of fuelwood and agriculture. Oikos 46:404-412.

WRI. 1990. World Resources Institute. World Resources: A Guide to Global Environment. Oxford University Press, New York

Molecular analysis of variation in *Lactuca* as a case study for the potential usages of molecular methods in the management of plant genetic resources

Ben Vosman
Centre for Plant Breeding and Reproduction Research (CPRO-DLO), 6700 AA Wageningen, The Netherlands

Introduction

Lactuca L. is a widely distributed genus that consists of about 100 species, which can be classified into four sections (Thompson *et al.* 1941). Seventeen species are found in Europe, nine of which are members of the section *Lactuca* (Feráková 1977). Ten species are found in northern America, 33 in east Africa and approximately 40 in Asia (Boukema *et al.* 1990). Lettuce (*L. sativa* (n = 9)), which is a member of the section *Lactuca* subsection *Lactuca*, has received most of the attention. Other relatively well studied species of the same subsection are *L. serriola*, *L. virosa* and *L. saligna*. Although cultivars are 99% selfing (Thompson *et al.* 1958), some cross-pollination occurs and inter-specific hybrids between different members of the subsection *Lactuca* can be made.

Several techniques have been used to characterize material from the genus *Lactuca*. There are a number of reports on the use of morphological markers (De Vries and Raamsdonk 1994), isozymes (Kesseli and Michelmore 1986) and molecular markers. In the following presentation, the use of RFLP, RAPD, microsatellite fingerprinting and DNA sequencing for the assessment of genetic diversity in *Lactuca* will be discussed.

RFLP

Kesseli *et al.* (1991) used RFLP to access variation among *Lactuca* species. Sixty five accessions of the subsection *Sativa* including 19 commercial and four breeding lines of the crisphead type, 20 commercial lines of the butterhead type, two looseleaf and one looseleaf by crisphead breeding line, three cos types, and one latin type was studied. From the rest of the subsection, eight populations of *L. serriola*, four of *L. saligna* and three of *L. virosa* were included, as well as two species outside the subsection *Lactuca*: *L. perennis* and *L. indica*. Twenty plants were sampled from each accession and DNA was extracted as a bulk. RFLP probes were either genomic or cDNA clones from *L. sativa* with known map positions, that were polymorphic in the intraspecific cross between the cvs Calmar and Kordaat (Landry *et al.* 1987). Fifty five probes were chosen that identified loci flanking regions of known resistance genes, covered the end of a linkage group, or that evenly spanned internal regions of a linkage group. Approximately 88% of the genome is within 20 cM of these selected markers. Three restriction enzymes were used and each probe-enzyme combination was treated as a separate locus. In total 143 loci could be used which identified 834 alleles. Most *Lactuca* species are self pollinated and many of the accessions used are inbred lines. As was to be expected, most of the diversity was found among populations and little within. Only two pairs of accessions, which were sister lines from an F_6 population and sister lines from a BC_5 respectively, could not be distinguished. The dendrogram based on the Nei index, matched the morphologically defined taxa well. Intra- and interspecific mean genetic distances showed the distinctness of the species, except for *L sativa* and *L. serriola*. The

interspecific mean genetic distance for these two was 0.71, which is not different from the intraspecific mean genetic distance in *L. serriola* (0.73).

RAPD

The RAPD technique was evaluated for differentiating individuals within closely related populations (Waycott and Fort 1994). Ten populations of butterhead and one of crisphead lettuce were analyzed with 13 primers that were selected from a set of 400. Nine out of the ten butterhead lines were visually very similar and homogeneous, the tenth line was highly heterogeneous. On average approximately 200 bands were detected in each butterhead line and approximately 350 in the crisphead line. The percentage of polymorphic bands within each line ranged from 22.6 % for the visually highly heterogeneous line to a perfect 0% for one of the other lines. Six lines averaged 3.0% or lower. In total, 93 bands were polymorphic (of a total of 270 sites) between lines. Most of the lines could easily be identified using only eight to ten primers. The relationship of within-line band segregation and standard deviations from morphological studies was positively correlated, with r^2 values ranging from 0.53 to 0.8. These figures suggest that the two parameters are moderately related and that the degree of detectable variation in genotype can imply a certain level of phenotypic variability, and vice versa (Waycott and Fort, 1994). This indicates that phenotypic assessment of material can still make an important contribution to germplasm evaluation by curators.

In another paper, Kesseli *et al.* (1994) compared the levels of polymorphism detected between two cultivars of lettuce for two types of molecular markers; RFLP and RAPDs. From the 1008 probes (1107 putative loci) derived from cDNA, 10% was polymorphic and 9% could be mapped; similar results were obtained with 180 probes (95 putative loci) derived from genomic DNA (11% polymorphic, which all could be mapped). The 50 RAPD primers tested detected 426 loci of which 17% were polymorphic and only 7% (30 loci) could be mapped due to scoring difficulties. Errors in scoring RAPD and RFLP bands were similar for both techniques. RFLP and RAPD markers showed similar distributions throughout the genome, significant clustering was detected in five out of the eight major linkage groups. Both identified similar levels of polymorphism. RAPD loci however, were much quicker to be identified, since one in every two primers detected a polymorphism. Approximately 70% of the RFLP variants were caused by deletion/insertion events.

Microsatellites

Microsatellite sequences are especially suited for distinguishing between closely related genotypes and are therefore favoured in population studies (Smith and Devey 1994) and for identification of closely related cultivars (Vosman *et al.* 1992). Microsatellite polymorphisms can be detected by Southern hybridization or PCR. Van de Wiel *et al.* (1996) have used oligonucleotides complementary to mini- and microsatellite sequences as a probe for detecting polymophisms between cultivars of lettuce as well as accessions of *L. serriola, L. virosa,* and *L. saligna*. The same material has been characterized morphologically (Frietema de Vries *et al.* 1994). Fourteen microsatellite and three minisatellite motifs were tested for fingerprinting in Lettuce. The microsatellite array TCT was the most promising one for fingerprinting in lettuce. Southern blots of a series of 75 *Taq*I-digested cultivars and accessions of *Lactuca sativa, L. serriola* and *L. virosa* were hybridized to $(TCT)_{10}$. In *L. sativa* and *L. serriola* a pattern of two to three highly polymorphic bands was visible in the high-molecular weight

range. This sufficed to distinguish all accessions and cultivars tested. In *L. virosa* more bands were visible, and here too, all accessions tested could be distinguished. The TCT fingerprinting was not suitable for determining relationships among the accessions. The level of polymorphism detected with this probe was too high. Similar observations were made by Rus-Kortekaas *et al.* (1994) who compared RAPDs and GACA microsatellite fingerprints in tomato. The Sequence tagged microsatellite approach is currently being tested on this material.

DNA sequencing

A system used frequently by taxonomists is sequencing of the internal transcribed spacer (ITS) of ribosomal DNA. This sequence has been shown to provide information that is discriminative at least down to the species level. Universal primers for the PCR amplification of the ITS are available (White *et al.* 1990) and have been tested on lettuce (Van de Wiel *et al.* 1996). ITS1 was sequenced and no significant polymorphisms were found between accessions of both *L. sativa* and *L. serriola*, indicating that ITS is not useful for studying variation within lettuce. However, there were several base pair changes between these two species and *L. virosa*. Also, within *L. virosa* there appears to be some variability among accessions. The study on ITS sequences is currently being extended by including related species and more accessions.

Degree of polymorphism

Depending on the information required, one technique is more appropriate than an other. The different molecular techniques are informative at different taxonomic levels. DNA sequencing of ITS appears to be useful for phylogeny of species but is of no use in population structure studies. Microsatellites probably are the fastest evolving pieces of DNA. They can be used for population studies but are of no use to phylogeny. RFLPs, RAPDs and probably also AFLP seem to fall midway between these two extremes and may be used for both purposes.

Reproducibility and repeatability

From the point of reproducibility and repeatability, the RAPD technique does not seem to meet the standards, especially not when results obtained in different laboratories are to be compared (Penner *et al.* 1993). It is unclear how the AFLP technique scores with respect to these points. Results obtained with RFLP, microsatellites and DNA sequencing appear to be reproducible and repeatable.

Data storage and handling

Data obtained by sequencing and by the sequence tagged microsatellite approach are relatively easily stored and handled in the form of base pairs and fragment lengths. This is in contrast to data obtained by RFLP and RAPD, which consist of fragments from which the length cannot be measured exactly or of more or less complex banding patterns. Experience with storage and handling of AFLP data is still largely lacking.

Conservation of germplasm

Molecular tools can be used in two ways for the conservation of genetic diversity:

(A) For screening material collected in genebanks: what is present, how many duplications there are, and whether new material to be added really does contain new information.

(B) For evaluate procedures used by the genebanks with respect to collecting and propagation of materials. Here the following questions may be addressed: how interesting are marginal populations compared to main populations; do they

contain many unique alleles; is putting a lot of effort in identifying and collecting them justified; how can efficient sampling strategies be worked out; how can we locate areas where we should conserve materials; what is the best way to propagate (cross-pollinating) species; and can criteria for the construction of core collections be formulated?

The technique(s) to be used for application A has to be cheap, fast and reliable since large numbers of accessions will have to be screened. Also, data storage and retrieval should be easy. None of the molecular techniques mentioned can satisfy the combination of these demands at this moment, especially the costs associated with the large scale use of molecular markers (between US$ 0.3 and 3 per data point) are a problem. For application B, there are less restrictions since only a limited number of accessions have to be analyzed and costs related to that do not play such a large role. Most of the techniques can, therefore, be put into practice to solve the questions indicated.

Research and development needs

Questions that should be addressed before starting to use molecular markers are:

- how important is an equal sampling of each part of the genome; how should the markers be distributed; is sequencing of one fragment just as informative as an equal number of datapoints, obtained by, for instance, RFLP?
- how many markers are needed to give an accurate description of a genotype; is the information obtained with a limited number of highly polymorphic markers (microsatellites) a good reflection of the total amount of variation present?

References

Boukema, I.W., Th Hazekamp and Th.J.L. van Hintum. 1990. CGN collection reviews: The CGN lettuce collection. December 1990. Centre for genetic resources, The Netherlands (CGN), Wageningen, the Netherlands.

Feráková, V. 1977. The genus *Lactuca* L. in Europe. Universita Komenského, Bratislava.

Frietema de Vries, F.T., R. van der Meijden and W.A. Brandenburg. 1994. Botanical files on lettuce (*Lactuca sativa*). Gorteria Supplement 2:1-44.

Kesseli, R.V. and R.W. Michelmore. 1986. Genetic variation and phylogenies detected from isozyme markers in species of *Lactuca*. J. Heredity 77:324-331.

Kesseli, R., O. Ochoa and R. Michelmore. 1991. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. Sativa*). Genome 34:430-436.

Kesseli, R.V., I. Paran and R.W. Michelmore. 1994. Analysis of a detailed genetic linkage map of *Lactuca sativa* (Lettuce) constructed from RFLP and RAPD markers. Genetics 136:1435-1446.

Landry, B.S., R.V. Kesseli, B. Farrara and R.W. Michelmore. 1987. A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331-337.

Penner, G.A., A. Bush, R. Wise, W. Kim, L. Domier, K. Kasha, A. Laroche, G. Scoles, S.J. Molnar and G. Fedak. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. PCR Methods Applications 2:341-345.

Rus-Kortekaas, W., M.J.M. Smulders, P. Arens and B. Vosman. 1994. Direct comparison of levels of genetic variation in tomato detected by a GACA-containing microsatellite probe and by random amplified polymorphic DNA. Genome 37:375-381.

Smith, D.N. and M.E. Devey. 1994. Occurence and inheritance of microsatellites in *Pinus radiata*. Genome 37:977-983.

Thompson, R.C., T.W. Whitaker and W.F. Kosar. 1941. Interspecific genetic relationships in *Lactuca*. J. Agric. Res. 63:91-107.

Thompson, R.C., T.W. Whitaker, G.W. Bohn and C.W. van Horn. 1958. Natural cross-pollination in lettuce. Proc. Am Soc. Hort. Sci 72:403-409.

Van de Wiel, C., P. Arens and B. Vosman. 1996. Molecular analysis of variation in lettuce by oligonucleotide fingerprinting and ITS sequencing. (In preparation).

Vosman, B., P. Arens, W. Rus-Kortekaas and M.J.M. Smulders. 1992. Identification of highly polymorphic DNA regions in tomato. Theor. Appl. Genet. 85:239-244.

Vries, I.M. de and L.W.D. van Raamsdonk. 1994. Numerical morphological analysis of lettuce cultivars and species (*Lactuca* sect. *Lactuca, Asteraceae*). Pl. Syst. Evol. 193:125-141.

Waycott, W. and S.B. Fort. 1994. Differentiation of nearly identical germplasm accessions by a combination of molecular and morphological analyses. Genome 37:577-583.

White, T.J., T. Bruns, S. Lee and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 *in* PCR protocols. A guide to methods and applications (M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White, eds.). Academic Press, San Diego.

Report of the Working Group on molecular techniques for the analysis, characterization and conservation of plant genetic resources

Identification of techniques suitable for different objectives

The resolution power, genetic information gathered, reliability of the techniques and cost per assay vary dramatically for the different techniques. Some techniques, such as RAPDs, may generally only be useful in specific situations due to low repeatability of results between laboratories and lack of genetic information on the loci assayed. The techniques of microsatellite loci analysis require high set-up costs. The evolution of variants at these loci are not correlated, as present understanding shows, to the fitness of the organism. Dr Karp and Dr Kresovich's presentations brought out very clearly the differences in the techniques. For studying variation at infra-specific level with various levels of polymorphism, techniques such as RFLP, ST-SSR, PCR-sequencing and even RAPD and AFLP may be followed. For resolving inter-specific differences and phylogenetic relationships, PCR-sequencing or PCR-RFLP are ideal. It was suggested that a working group be established to decide the guidelines for selecting molecular techniques suitable for each type of analysis (see Introduction, p.1).

Measures of diversity and correlation between the measures

Dr Marshall and Dr Gonzalez pointed out that the various measures of diversity used are not always the ideal ones for plant genetic diversity analysis. Genetic diversity analysis should be at population level and the data collected should relate to specific gene loci so as to make them meaningful and comparable between laboratories. Proper sampling of the natural populations has to be followed to get a realistic view of the variation existing in nature so that the germplasm collections have complete representations of total genetic diversity occurring in nature.

Use of combination of techniques to solve a problem

No single technique can be recommended as ideal for all situations. Patterns of variation at different regions of the genome differ greatly and the extent of genetic variation/polymorphism observed in different taxa differ widely. A technique suitable in one taxa may not be capable of resolving genotypic differences in another. The number of samples to be analyzed varies drastically. Therefore a combination of techniques should be employed to study genetic diversity taking into account the cost, amount of work involved and type of information to be generated.

What population to use for analyses?

An appropriate sampling strategy is important to avoid duplications and limit the number of samples to be analyzed. Selections based on geographic distributions and morphological divergence are advisable. The quality of data generated is affected severely by the sampling strategy. The breeding system found in behaviour of the taxa to be tested should also have a substantial impact on sampling strategies.

Collaboration and sharing of information

The cost involved in genetic diversity analysis can be reduced drastically by sharing of information and resources. Availability of sequence information, primarily sequences, probes, etc. will reduce set- up costs in the case of microsatellite and RFLP analyses.

New emerging techniques

Dr Smith pointed out that techniques which do not require gel running are essential for reducing the time required for analysis and for increasing the number of samples to be analyzed. The silicon chip band techniques, similar to ELISA techniques, are emerging as possible alternatives.

Study of cytoplasmic diversity

Elucidation of phylogenetic relationships, diversity in cytoplasmic organelles, namely chloroplast and mitochondrial DNA, are also good indicators of genetic diversity and need to be studied more widely. The information is also useful for plant breeders.

Conclusions

The case study of Dr Bonierbale on Cassava and beans (see pp.98-102) needs to be highlighted through a more detailed publication. This would help in generating interest and attracting more funds for this type of study.

IPGRI should participate in the ECU platform **"The Biotechnology for Biodiversity platform (BBP)"** as this would help in developing linkages and keeping in touch with technological developments.

IPGRI should establish global connections between various national programmes in order to facilitate the networking of various programmes.

IPGRI should develop particularly close interactions with national programmes already involved in plant molecular genetic studies, such as those in India, China and Malaysia.

Projects studying specific issues, particularly appropriate sampling strategies, are required and should be developed by IPGRI.

IPGRI should have a population geneticist to assist in the formulation of conservation strategies and molecular diversity analyses.

The services of molecular geneticists, for example, participants in the Workshop, are already available to IPGRI on an *ad hoc* basis and can be utilized.

Analysis, management and exchange of molecular data

Some aspects of the analysis, management and exchange of molecular data and their relationship to more traditional types of genetic resources data

Mark Perry and George Ayad
International Plant Genetic Resources Institute, 00145 Rome, Italy

Background

The types of data that have traditionally been recorded to describe, select and use plant genetic resources are focused towards the agronomic and morphological characteristics of a plant accession. In addition, there are several parameters normally observed that describe the site of collection, ethnobotanical characteristics and obvious but important susceptibilities or resistances to biotic and abiotic factors. It has generally been assumed that knowing the characteristics of a germplasm sample in these categories provides the genetic resources collection curator and the users of the collection with the data needed to access the diversity and utilize the germplasm collection.

These data types have been found to be some of the most important as far as the use and management of the genetic material is concerned. The majority of the more traditional plant breeders need to have data that can be related to the characteristics of what they finally want to produce. At this time, these are generally the morphological and agronomic characteristics and any desirable physiological traits as well as disease and pest resistances. The closer the material is in character to the desired final result, the quicker the breeding process usually is. Genetic resources curators need to invest in the collection and management of data to assist them in knowing the extent of diversity the collection holds and how to assist the user in its use.

Generally, the assumption has been used in genetic resources work that if certain general traits are known, the extent of diversity of the collection is known. These general traits have been morphological, agronomic and susceptibility traits. This notion is untenable since it is impossible at a given time, to know the characteristics of a crop that will be important in the future and the extent to which morphological and agronomic characters actually differentiate between accessions. Also, the differentiation between individuals within an accession is not generally accurate since most individuals are not used for observation, there is always some degree of outcrossing within a species, and the characteristics used may be subject to environmental variation.

The use of molecular techniques for demonstrating and measuring variation within genetic resources collections are promising as tools for finding and measuring diversity. However, there are several issues that should be discussed in relation to the traditional methods of assessing diversity and using genetic resources collections. The purpose of this short paper is to contrast the usefulness of molecular genetic techniques as means of measuring diversity and using genetic resources collections compared to the now widespread use of agronomic, morphological and resistance characters.

Traditional types of genetic resources data

The following groups of data are traditionally used for describing plant genetic resources. These data basically name each accession, describe its history relative to where it was developed or collected and describe its morphological and agronomic characteristics.

A descriptor is generally defined as an identifiable and measurable trait or character of a plant accession. It includes a descriptor name, a descriptor definition and possibly a list of descriptor states and codes used for the descriptor states. The selection of descriptors made during the establishment of the database will determine not only the size of the resulting database, but the usefulness of the database to the users. The choice of descriptors and descriptor states will also help to establish a standard type of language for the exchange and use of data. Also the use of descriptor lists that have been developed by an international group of crop-specific experts will generally assist in compatibility of data during exchange. IPGRI has been responsible for publishing over 60 crop-specific descriptor lists, each of which follows a standard format for descriptor categories.

The use of standard descriptor lists assists with the exchange of data between institutes that maintain or are interested in the same crop. Data incompatibility exists if descriptors with nonequivalent definitions or non-convertible descriptor states or coding schemes are used. For many plant morphological or agronomic descriptors, the conversion between two descriptors lists is generally not possible which results in a loss of reliable information.

Most descriptors that are used in plant genetic resources databases can be placed in one of the following categories:

Passport

These descriptors are used to describe characters of the accession observed at the time of original field collection or those which are important when an accession is bred. They include the original identity of the accession (scientific name and collector's number) as well as the identifier assigned to the accession after it enters into a collection. In the IPGRI descriptor lists, they include both the accession and collection data sections.

Characterization

These descriptors define characters that are highly heritable and that can be seen easily, and are equally expressed in all environments. For many highly inbred species, data for these descriptors can be collected during multiplication trials, but more commonly they are collected during characterization trials.

Preliminary evaluation

These descriptors define characters that have been identified by experts as important to the description/identification of the accession. It may be possible to assess visually or measure these easily, but they are generally environmentally influenced.

Evaluation

These descriptors define characters that are linked to breeding programmes. They require replicated trials for their accurate assessment. Because crop use varies from country to country, these descriptors may or may not be used depending on their importance within a particular country or region.

Management

These descriptors are used to assist in stock control of collections. They help to determine and assist in the management of multiplication and regeneration trials and when these should take place and in the recording of the distribution of material into and out of the genebank.

General information on collections

Besides descriptor type information, data on the collections that hold material of a particular crop are very useful for locating alternative sources of germplasm and for providing a start in locating other people working on similar crops. IPGRI has published this type of information in its "Directories of germplasm collections". The information collected includes the complete name, address, contact numbers, a summary of the holdings (genus and species names, geographical distribution of the material, documentation status and quarantine restrictions that might apply).

Since the development and widespread use of protein and molecular techniques, there has not been a clear categorization of them into one of the above categories. Some working with genetic resources consider them outside the scope of traditional characterization data but closer to evaluation data. However, given the definitions used for them, these techniques should produce data that will not vary if the plant is moved into a different environment. However, in some cases the results from these techniques may vary depending on the particular procedures and laboratory conditions that are employed.

Management of traditional genetic resources data

The data types described in the previous section have traditionally been managed in fixed field database management systems. Many of them can be analyzed statistically to help in locating the most important accessions for a particular use. Even though these databases are widespread, the data held within will vary depending on the location of the genebank and the particular genebank objectives. Access to these data takes the format of electronic copies of the data, printed listings and on-line access which allows users to find the data they are interested in.

Molecular genetics data and genetic resources

The use of molecular genetic techniques to assess the diversity in genetic resources has been occurring for some time. However, to date, they have not been used for genetic resources on any scale. Their use in genebank management and use programmes raises a number of issues and questions, particularly:

1. How closely do these data relate to the characters plant breeders are using in their breeding programmes? The linkage between a banding pattern found using RAPD techniques and useful traits, for example, is not straight forward and provides very little useful data to the breeder unless it is known to be linked to a useful allele.
2. How practical is it to characterize a genetic resources collection for a particular molecular genetics character as far as cost is concerned? With many crops, it is very inexpensive, particularly in developing countries, to perform a characterization trial for several thousand accessions at one time and that includes observation for many morphological and agronomic characters that

either interest breeders directly or have been found to be correlated with characters of interest.

3. How reproducible are the results for many of the molecular genetic techniques? There may be difficulties reproducing results in different laboratories. This will make it difficult to use the data that are generated throughout the user community.
4. In relation to point 3, is there a common method for storing and/or managing data generated from molecular genetic techniques? The fact that a quantitative result can be obtained from the analysis of banding patterns is positive, but can these data be translated into information that can be used by a genetic resources user?
5. The management of data that have been derived quantitatively from banding patterns or sequences may be sufficient for the storage of the data. Is it important for these data to be portrayed in a object-oriented manner to allow for the quick use by those that can draw conclusions from them?
6. How many "differences" have to be determined to provide the indication that two accessions are usefully different? A cost/benefit analysis will generally be needed to assess this aspect.
7. How can the analysis of large amounts of these types of data be accomplished, either in an object-oriented mode or of those derived quantitatively, in order to make the data referring to even a part of the genetic resources collection meaningful? The computational aspects of this task may be more than most genetic resources users are able to access, hence the data are of limited usefulness until they have the needed access.
8. Standardization of the techniques for deriving these data and the methods of analysis needs clarification. Will the data derived from one laboratory be usable in another in an understandable manner? If not, is there some degree or way of creating a standard which will eventually be recognizable to those working in the molecular genetics field and those working in the area of plant genetic resources?
9. What are the data exchange ramifications resulting from these data types? Is it possible to provide these data to users who are not genuinely familiar with them and assume that they will use them? What are the consequences as far as data access is concerned if they are best used in an object-oriented manner?

Analyzing molecular data for studies of genetic diversity

Fernando González-Candelas and Carmen Palacios
Departament de Genètica, Universitat de València, Valencia 46100, Spain

Introduction

One of the main goals in conservation is the preservation of genetic diversity. Traditionally, the study of genetic diversity has fallen within population genetics, which has focussed on measuring its extent in natural populations, in comparing levels of genetic diversity within and among populations and in making inferences on the nature and intensity of evolutionary processes from the observed patterns of genetic diversity. Hence, there is a long tradition as well as a wealth of conceptual tools in population genetics for analyzing, measuring and partitioning genetic diversity.

This review on methods for analyzing variation using molecular markers will start with a brief outline of the main population genetic concepts involved. These ideas were developed for simple situations, such as the one-locus two-alleles case, and were refined and generalized later. However, the main features are best understood by taking the simplest case which, in terms of a molecular marker, can be understood as an allozyme locus with only two alleles. In this situation, we are dealing with codominant markers, for which all possible genotypes (both homozygotes and the heterozygote) can be easily ascertained.

We will then move to the case of the richest possible markers in terms of the amount and quality of the information provided, DNA sequences. For these markers, it is possible to establish measurements of their evolutionary distance, which can further be used to refine the measurements of genetic diversity. Once the direct analysis of nucleotide sequences has been developed, we will consider other markers which provide indirect estimates of nucleotide divergence between alternative alleles, such as RFLPs and restriction site data. After that, we shall consider the complicating effects of using dominant markers, such as RAPDs and multilocus DNA-fingerprinting, for which the "presence" allele is dominant over the "absence" allele. Furthermore, use of these markers usually results in an unknown number of loci being analyzed simultaneously which introduces further complications. Finally, microsatellite markers will be considered, as these can be interpreted with or without reference to the evolutionary relatedness among alleles. Some computer programmes available for use in population genetics and analysis of molecular variation are listed in the Appendix to this paper.

The population genetics description of diversity

Genetic diversity can be measured, as can any other measurement of diversity, in different ways. One of the most commonly used ways defines diversity in a single locus as

$$D_l = 1 - \sum_i p_{li}^2 \qquad \text{Eq.1}$$

where p_{li} represents the frequency of the i-th allele at locus l. An average diversity for several (L) loci is given by

$$D = 1 - \frac{1}{L}\sum_{l} D_l = 1 - \frac{1}{L}\sum_{l}\sum_{i} p_{li}^2 \qquad \text{Eq.2}$$

This definition of diversity is closely related to the expected heterozygosity in a single locus for diploid organisms when populations are in Hardy-Weinberg equilibrium. Under these circumstances, the expected heterozygosity in a locus is given by:

$$H_l = 2\sum_{i \neq j} p_{li} p_{lj} = 1 - \sum_{i} p_{li}^2 = D_l \qquad \text{Eq.3}$$

This relationship provides an interesting solution to the problem of comparing levels of diversity between haploid organisms, where genetic diversity is readily defined but heterozygosity is not, and diploid organisms.

Following Weir (1990), it is possible to obtain a partition of the distribution of genetic diversity estimated directly from heterozygosity values in terms of an analysis of variance (ANOVA), taking into account the several levels at which heterozygosity can be defined. So the effects of subpopulations, of individuals within subpopulations, of the different loci and their interactions on the heterozygosity observed in a population can be easily obtained and tested for the significance of their relative contributions to the observed variability.

Deviations from Hardy-Weinberg equilibrium reflect on differences between observed and expected values of heterozygosity. These deviations, which can be due to many different causes, can be formulated in terms of inbreeding coefficients. So the genotypic frequencies in a two-allele locus, P and Q for homozygotes and H for heterozygotes, can be expressed (Wright 1931) as:

$$P = p^2 + fpq \qquad \text{Eq.4}$$

$$H = 2pq - 2fpq \qquad \text{Eq.5}$$

$$Q = q^2 + fpq \qquad \text{Eq.6}$$

where f is the inbreeding coefficient. Its sign and value reflect deviations from Hardy-Weinberg proportions. When f takes a positive value, there will be an excess of homozygotes and a lack of heterozygotes, as when endogamic reproduction occurs. Conversely, negative values of f are an indication of exogamy. An indirect estimate of the amount of inbreeding for a given locus is obtained from the observed proportion of heterozygotes, H_0, as

$$\hat{f} = 1 - \frac{H_0}{2\hat{p}\hat{q}} \qquad \text{Eq.7}$$

where $\hat{p}$ and $\hat{q}$ are the estimated gene frequencies of each allele. The corresponding sampling variance for a sample of size n is given by (Rasmussen 1964)

$$Var(\hat{f}) = \frac{1}{2n\hat{p}\hat{q}}\left\{(1-\hat{f})\left[2\hat{p}\hat{q}(1+\hat{f}) + \hat{f}(2-\hat{f})(1-2\hat{p})^2\right]\right\} \qquad \text{Eq.8}$$

It immediately follows that the amount of genetic diversity in a given locus is related to the level of inbreeding in that population.

The two major, non-selective causes that move natural populations of diploid organisms away from Hardy-Weinberg equilibrium are drift and inbreeding. Both factors act on reducing the amount of heterozygotes and increasing homozygosity, and this effect is further enhanced whenever populations become structured i.e. mating and dispersal take place only in the close neighbourhood of each individual. Wright (1951) introduced a method of describing genetic population structures of diploid organisms in terms of three *F*-statistics or allelic correlations.

In every subdivided population, there are at least three levels of complexity: individual organisms (*I*), subpopulations (*S*) and the whole population (*P*). These three levels are associated with three different measurements of heterozygosity:

H_I, which can be interpreted as the average heterozygosity of all the genes in a single individual or the probability of heterozygosity in any gene. H_I is the observed heterozygosity averaged over populations. If H_{il} represents the heterozygosity in a single locus in subpopulation *i* (Eq. 9) and *k* subpopulations and *L* loci are considered, then

$$H_{Il} = \frac{1}{k}\sum_{i=1}^{k} H_{il} \qquad H_I = \frac{1}{L}\sum_{l=1}^{L} H_{Il} \qquad \text{Eq.9}$$

H_S represents the heterozygosity level expected in a panmictic subpopulation. Hence, for a diallelic locus with gene frequencies p_i and q_i in subpopulation *i*, H_S always equals $2p_iq_i$. For *k* subpopulations, the average value of the H_S for each subpopulation is represented by $\overline{H}_S$.

H_T represents the expected heterozygosity of all subpopulations when pooled and mating is random in the pooled population. In this case, H_T is given by $2p_0q_0$, being p_0 the average gene frequency across subpopulations.

Wright developed these ideas for the diallelic case, grouping all but the most frequent allele into a single class. An explicit multiallelic form was presented by Nei (1987), although the notation employed was slightly different. Nei uses G_{ST} to refer to themultiallelic form of F_{ST} developed by Wright which, accordingly should be reserved only for the diallelic case (Nei 1987). So, if p_{iX} is the frequency of the *i*-th allele in subpopulation *X*, then:

$$H_S = 1 - \sum_{i=1}^{k} p_{iX}^2 , \qquad \text{Eq.10}$$

and if $\overline{p}_i$ is the average frequency of the *i*-th allele over subpopulations, then

$$H_T = 1 - \sum_{i=1}^{k} \overline{p}_i^2 \qquad \text{Eq.11}$$

The inbreeding coefficient measures the reduction in individual heterozygosity due to deviations from random mating in the local populations. This inbreeding coefficient is represented by F_{IS} and is given by:

$$F_{IS} = \frac{\overline{H}_S - H_I}{H_S} \qquad \text{Eq.12}$$

The effects of population subdivision can be quantified by means of the fixation index F_{ST}, which is the reduction in the heterozygosity in a subpopulation due to nonrandom mating with respect to the total population. F_{ST} is given by:

$$F_{ST} = \frac{H_T - \overline{H}_S}{H_T} \qquad \text{Eq.13}$$

An alternative interpretation of F_{ST} in its diallelic version is as the ratio between the expected and observed variances of gene frequency considered among all subpopulations. So:

$$F_{ST} = \frac{\sigma_p^2}{p_0 q_0} \qquad \text{Eq.14}$$

The following relationship holds for F-statistics:

$$(1 - F_{IS})(1 - F_{ST}) = (1 - F_{IT}) \qquad \text{Eq.15}$$

The estimation of F-statistics by mere substitution in the previous equations of the relevant parameters by their observed values does not necessarily lead to better estimates, especially with small sample sizes. Ideally, the estimates should be corrected for the effects of sampling a limited number of individuals in a limited number of subpopulations. Several corrections have been proposed (Wright 1968; Curie-Cohen 1982; Nei and Chesser 1986; Weir and Cockerham 1984; Nei 1986) although further difficulties arise with their application.

Although several statistical tests have been proposed for testing the significance of each of these F-statistics (Li and Horvitz 1953; Brown 1970; Workman 1970), Cockerham (1969, 1973) pioneered the development of a system for analyzing F-statistics in an adequate context for hypothesis testing. Cockerham worked on the relationship between F-statistics and components of variance under an ANOVA framework. This work was further developed by Weir and Cockerham (1984) and Long (1986).

Excoffier *et al.* (1992) have introduced an alternative method for partitioning genetic variance at different levels which is based on an extension of the works of Cockerham (1973), Long (1986) and Long *et al.* (1987) on the allelic correlation among demes. This method uses the usual setup of the analysis of variance on a transformed measure of distance between different genetic variants (usually haplotypes, see below) obtaining an evolutionary metric distance. The method is implemented in the program WINAMOVA (Excoffier *et al.*).

Summarizing, population geneticists usually evaluate genetic diversity by means of observed and estimated amounts of heterozygosity, and they compare and partition this variation among different hierarchical population levels by means of Wright's F-statistics and related quantities.

There are, however, alternate ways of analyzing the same basic information. Another way to study the level of genetic differentiation among populations is by asking how similar they are. It is generally considered that genetic distance increases with time of divergence from a common population. But this requires a genetic model that specifies those genetic processes, such as migration and drift, that make

populations diverge. Many genetic distance measures have been proposed (Reynolds 1981; Nei 1987) since Cavalli-Sforza and Edward's (1967) attempt to relate their distance measure to the evolutionary changes of gene frequencies among populations. Some of the distance measures used in genetic studies are geometric distances that do not consider special features of evolutionary processes. This is the case of Mahalanobis' (1936), Bhattacharyya's (1946) and Rogers' (1972) distances, to name a few. More appropriate measures when dealing with genetic data have been developed by Nei (1972, 1973). Two of them are especially relevant for this review, the minimum genetic distance (D_m) and the standard genetic distance (D).

Let p_{iX} and p_{jY} represent the frequency of allele i in a given locus in population X and the frequency of allele j in the same locus in population Y. Suppose we take random allele from each population and compare them. The probability that both alleles are identical is given by:

$$j_{XY} = \sum_i p_{iX} p_{iY} \qquad \text{Eq.16}$$

and they will be different with probability $1\text{-}j_{XY}$. When the alleles are different, there is at least one codon or nucleotide, depending on what marker is being used, difference between them. Therefore $d'_{XY}=1\text{-}j_{XY}$ gives the minimum number of differences between both populations. However, when there is a polymorphism, two alleles randomly sampled from one population will not always be identical. Hence, we need to correct for these intrapopulational differences. For each population, the intrapopulation measure of differentiation is given by:

$$d_X = 1 - \sum_i p_{iX}^2 \qquad \text{Eq.17}$$

which equals the expected heterozygosity (and the gene diversity) for that locus in that population. Therefore the net minimum number of differences between two populations is given by:

$$d_{XY} = d'_{XY} - \frac{d_X + d_Y}{2} = \sum_i \frac{(p_{iX} - p_{iY})^2}{2} \qquad \text{Eq.18}$$

In practice, d varies from one locus to another, and in order to estimate the difference between two populations, the average of d over all loci must be taken. This average is known as *minimum genetic distance* and is given by:

$$D_m = D_{XY(m)} - \frac{D_{X(m)} + D_{Y(m)}}{2} \qquad \text{Eq.19}$$

where $D_{XY(m)} = 1 - J_{XY}$, $D_{X(m)} = 1 - J_X$, and $D_{Y(m)} = 1 - J_Y$, and J_{XY}, J_X and J_Y are the averages of j_{XY}, j_X and j_Y over all loci. The main disadvantage of using the minimum genetic distance is that it can seriously underestimate the difference between pairs of populations. However, it can be used for the study of the maintenance of polymorphism among populations (Chakraborty 1974).

When changes occur independently at every position in the genome, the mean number of net substitutions is given by:

$$D = -\log I, \qquad \text{Eq.20}$$

where

$$I = \frac{J_{XY}}{\sqrt{J_X J_Y}} \qquad \text{Eq.21}$$

This is known as the *standard genetic distance*. I takes value 1 when the two populations have identical gene frequencies in all loci and 0 when they share no alleles. Because of this property, I itself has been used as a measure of the genetic similarity between populations, and it is known as the *genetic identity*. Nei (1987) discusses the estimation procedure and sampling properties of several estimators of genetic distances.

Generally, more than two populations of any species are being analyzed, and all possible pairwise genetic distances (or identities) have been estimated. In these cases, the information provided in the corresponding distance matrix can be used to simultaneously analyze the relationships among all the populations. This is usually accomplished by means of different multivariate techniques (Manly 1986), of which clustering methods are most popular. The two most commonly used clustering methods are UPGMA (Sokal and Michener 1958; Sneath 1973) and neighbour-joining (Saitou and Nei 1987). UPGMA is an ultrametric method and should provide accurate clusters when distances are linearly proportional to the amount of divergence, for instance under constancy of evolutionary rates, whereas neighbour-joining is a very robust method which is gaining widespread use because of its relative independence of assumptions (Swofford and Olsen 1990; Nei 1991).

Using nucleotide sequence data

There are two different quantities for measuring the amount of genetic variation at the DNA level: the average number of pairwise nucleotide differences and the number of segregating (polymorphic) sites among a sample of sequences.

The number of segregating sites, S, is the number of sites which are occupied by at least two different nucleotides. The average number of pairwise nucleotide differences among DNA sequences is defined as:

$$d = \frac{2\sum_{i<j} d_{i,j}}{n(n-1)} \qquad \text{Eq.22}$$

where $d_{i,j}$ is the number of nucleotide differences between sequences i and j, and n is the number of DNA sequences sampled from a population. When the number of DNA sequences studied is large, it is advisable to use heterozygosity instead of d. At a site i, heterozygosity is defined as:

$$H_i = 1 - \sum_{j=1}^{4} p_j^2 \qquad \text{Eq.23}$$

where p_j is the relative frequency of nucleotide j (j = 1, 2, 3, 4 corresponding to nucleotides A, T, C and G) and an unbiased estimate is given by (Tajima 1993):

$$\hat{H}_i = \frac{n}{n-1}\left(1-\sum_{j=1}^{4} p_{ij}^2\right) \qquad \text{Eq.24}$$

where p_{ij} is the observed frequency of nucleotide j at site i and n is the number of nucleotide sequences. The following relationship can be shown between heterozygosity and the average number of nucleotide differences among the n sequences studied

$$d = \sum_{i=1}^{m} H_i \qquad \text{Eq.25}$$

where m is the number of nucleotide sites in the DNA sequence.

Both S and d depend on the length of the nucleotide sequence (m) and the amount of DNA polymorphism per site can be used instead, simply by dividing any of those measurements by m. The new measures correspond to the average number of nucleotide differences per site (S/m) and to the average heterozygosity per site (H/m). Tajima (1963) provides an excellent review of both measures under different evolutionary scenarios.

In order to compare the amount of variation at several levels using sequence data, it is necessary to define the average number of nucleotide differences per site between two sequences, an amount also known as *nucleotide diversity*. It can be defined as:

$$\pi = \sum_{i,j} x_i x_j \pi_{ij} \qquad \text{Eq.26}$$

It can immediately be shown that π can be estimated by:

$$\hat{\pi} = \frac{n}{n-1}\sum_{i<j} \hat{x}_i \hat{x}_j \pi_{ij} = \frac{2\sum_{i<j}\pi_{ij}}{n(n-1)} = \frac{S}{m} \qquad \text{Eq.27}$$

An alternative estimate of π was proposed by Nei and Miller (1990):

$$\hat{\pi}_p = \frac{\sum_{i=1}^{m} h_i}{m} \qquad \text{Eq.28}$$

which is equivalent to the average heterozygosity per site.

The average proportion of nucleotide differences between n_X sequences from population X and n_Y sequences from population Y can be calculated as:

$$\bar{p}_{XY} = \frac{\sum_{i=1}^{m} h_{XY_i}}{m} \qquad \text{Eq.29}$$

where h_{XYi}, the proportion of nucleotide differences at site i, is given by:

$$h_{XY_i} = 1 - \sum_{j=1}^{4} p_{Xij} p_{Yij} \qquad \text{Eq.30}$$

An approximate value of d'_{XY} can de obtained as:

$$\hat{d}'_{XY} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\bar{p}_{XY}\right) \qquad \text{Eq.31}$$

and the interpopulation component of the nucleotide differentiation between populations X and Y is then given by:

$$\hat{d}_{XY} = \hat{d}'_{XY} - \frac{\hat{d}_X + \hat{d}_Y}{2} \qquad \text{Eq.32}$$

whose expected value for a pair of populations that diverged t years ago is $2\lambda t$, being λ the rate of nucleotide substitution per site per year.

When several populations are analyzed, Lynch and Crease (1990) devised a method for partitioning nucleotide diversity into intra- and interpopulation components which is analogous to F_{ST} at the DNA level. Their method is based on the average number of nucleotide substitutions per site between pairs of sequences sampled both from each population and from all possible pairs of populations. The intrapopulation component is estimated as:

$$\hat{v}_X = \frac{2}{n_X(n_X - 1)} \sum_{i,j} n_{iX} n_{jX} \hat{d}_{ij} \qquad \text{Eq.33}$$

where n_X is the total number of individuals sampled in population X, n_{iX} and n_{jX} are the number of those individuals with haplotypes i and j, respectively, and d_{ij} is the estimated nucleotide distance between those haplotypes. The combined estimate of intrapopulation differentiation is given by:

$$\hat{v}_w = \frac{\sum_{i=1}^{n_p} \hat{v}_i}{n_p} \qquad \text{Eq.34}$$

being n_p the number of populations studied. The combined estimate of interpopulation differentiation is obtained as:

$$\hat{v}_b = \frac{2\sum_{X<Y} \hat{v}_{XY}}{n_p(n_p - 1)} \qquad \text{Eq.35}$$

The analogue to the indices of population structure (Wright's F_{ST}) proposed by Lynch and Crease (1990) at the nucleotide level is:

$$N_{ST} = \frac{\hat{v}_b}{\hat{v}_b + \hat{v}_w} \qquad \text{Eq.36}$$

which is the ratio between the average genetic distance between genes from different populations and the average global genetic distance. The extreme values of N_{ST}, 0 and 1, are indication of null and complete population subdivision, respectively.

The approximate sampling variance of N_{ST} is given by:

$$Var(\hat{N}_{ST}) = \left(\frac{\hat{N}_{ST}}{\hat{v}_w + \hat{v}_b}\right)^2 \left[\left(\frac{\hat{v}_w}{\hat{v}_b}\right)^2 Var(\hat{v}_b) - 2\left(\frac{\hat{v}_w}{\hat{v}_b}\right) Cov(\hat{v}_w, \hat{v}_b) + Var(\hat{v}_w)\right] \qquad \text{Eq.37}$$

If we assume, as a first approximation, that N_{ST} is normally distributed then the statistic:

$$D = \frac{N_{ST}^2}{Var(N_{ST})} \qquad \text{Eq.38}$$

will be chi-square distributed with 1 degree of freedom under the null hypothesis of no population subdivision.

The analysis of restriction data

In order to analyze genetic diversity using restriction data, it is necessary to introduce a few previous ideas. In the first place, it is important to distinguish between restriction fragment data, for which only the size of the generated fragments is available, and restriction site data, for which the precise location of a recognition sequence for a restriction enzyme is known. These types of data are not adequate for comparing sequences which have diverged considerably, but both are usually acceptable for studying intraspecific variation. In order to compare levels of genetic diversity within and among populations from restriction data, it is necessary to previously estimate the number of nucleotide substitutions between any pair of sequences. Several methods, both for restriction fragment and restriction site data are reviewed in Nei (1987).

For restriction site data, let S denote the probability that two sequences, X and Y, share the same recognition sequence at a given site. This value can be described by:

$$S = (1-p)^r \qquad \text{Eq.39}$$

where p is the probability that the sequences do not share a nucleotide in a given position and r represents the length of the recognition sequence. The probability p is related to the expected number of substitutions per site, d, according to:

$$p = \frac{3}{4}\left[1 - e^{\left(-\frac{4}{3}d\right)}\right] \qquad \text{Eq.40}$$

If the rate of nucleotide substitution per site and per year is λ, then d is also given by $d = 2\lambda t$. Consequently, it is possible to estimate d if the value of S is known. Whenever nucleotide divergence is relatively small ($d < 0.25$), as is certainly the case for sequences from the same and very closely related species, S is usually approximated by (Nei and Li 1979; Kaplan and Risko 1981; Li 1981):

$$S = e^{-2r\lambda t} \qquad \text{Eq.41}$$

The maximum likelihood estimator of S (Nei and Tajima 1983) is given by:

$$\hat{S} = \frac{2m_{XY}}{m_X + m_Y} \qquad \text{Eq.42}$$

with variance given by:

$$Var(\hat{S}) = \frac{\hat{S}(1-\hat{S})(2-\hat{S})}{m_X + m_Y} \qquad \text{Eq.43}$$

where m_X and m_Y are the number of restriction sites in sequence X and Y, respectively, and m_{XY} is the number of restriction sites shared by both sequences.

Once an estimate of S is available, it is possible to estimate the proportion of nucleotide differences, p, by:

$$\hat{p} = 1 - \sqrt[r]{\hat{S}} \qquad \text{Eq.44}$$

and the estimate of d follows immediately:

$$\hat{d} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\hat{p}\right) \qquad \text{Eq.45}$$

When $d < 0.25$, it is possible to use the above approximation (Nei and Li 1979) which leads to the estimate:

$$\hat{d}_1 = -\frac{\ln \hat{S}}{r} \qquad \text{Eq.46}$$

In the preceding derivation, it has been assumed that a single enzyme has been used. If several enzymes with the same length of the corresponding recognition sequences are used, it is possible to use the above expression simply by taking summations over all the enzymes. However, if several enzymes with different lengths in their recognition sequences are employed, then it is convenient to follow the method proposed by Nei and Miller (1990) in order to weigh the data obtained with each enzyme class. When values of d have been estimated for each class of restriction enzyme according to Eqs. 45 or 46 above, then a combined estimate of d for all the enzymes is given by:

$$\hat{d} = \frac{\sum_k \hat{m}_k r_k \hat{d}_k}{\sum_k \hat{m}_k r_k} \qquad \text{Eq.47}$$

where $\hat{m}_k$ is the average number of bands for the k-th class of enzymes, r_k is the length of the sequence recognized by the class of enzymes and $\hat{d}_k$ is the corresponding estimated nucleotide divergence.

This estimate of nucleotide divergence can be extended to be an estimate of interpopulation nucleotide divergence for all possible pairs of populations X and Y simply by considering the nucleotide divergence estimated for all possible pairs of sequences one from each subpopulation. This leads to:

$$\hat{S}_{XY} = \frac{2\sum_{i,j} m_{X_i Y_j}}{\sum_{i,j} m_{X_i(Y_j)} + \sum_{i,j} m_{(X_i)Y_j}} \qquad \text{Eq.48}$$

where m_{XiYj} represents the number of restriction sites shared by the i-th sequence from population X and the j-th sequence from population Y, whereas ; $m_{Xi(Yj)}$ and $m_{(Xi)Yj}$ are the number of restriction sites in sequences i and j form subpopulations X and Y, respectively. This estimate can be obtained for each different class of restriction enzymes, and these estimates can be combined, similarly as above, into the estimate:

$$\hat{d}'_{XY} = \frac{\sum_k \hat{m}_k r_k \hat{d}_{XY_k}}{\sum_k \hat{m}_k r_k} \qquad \text{Eq.49}$$

where:

$$\hat{m}_k = \frac{\hat{m}_{X_k} + \hat{m}_{Y_k}}{2} \qquad \text{Eq.50}$$

This estimate of interpopulation divergence includes both an interpopulation component, due to differences in the frequency of the different sequences in both subpopulations, as well as an intrapopulation component, due to variation within each subpopulation. If we are interested merely in interpopulation divergence, then the interpopulation component of the above estimate of nucleotide divergence between both subpopulations can be estimated according to:

$$\hat{d}_{XY} = \hat{d}'_{XY} - \frac{\hat{d}_X + \hat{d}_Y}{2} \qquad \text{Eq.51}$$

Unfortunately, the method developed by Nei and Li (1979) to estimate the sampling variance of d_{XY} cannot be applied in this situation (Nei and Miller 1990) and resampling estimates, by jackknifing or bootstrapping, must be obtained.

When only restriction fragment length data are available, the estimate of nucleotide divergence, as the average number of nucleotide substitutions per site, between a pair of sequences can be obtained as:

$$\hat{d} = -\frac{2}{r} \ln \hat{G} \qquad \text{Eq.52}$$

where $G = e^{-r\lambda t}$ is the probability that no nucleotide substitution has occurred at a restriction site and is related to the proportion of shared fragments (F) by means of:

$$F = \frac{G^4}{3-2G} \qquad \text{Eq.53}$$

and F can be estimated from:

$$\hat{F} = \frac{2m_{XY}}{m_X + m_Y} \qquad \text{Eq.54}$$

Nei (1987) has proposed an iteration procedure to estimate G, once an estimate for F has been obtained from Eq. 54.

Similar expressions for the estimates of intra- and interpopulation nucleotide divergence can be obtained. González-Candelas *et al.* (1995) have recently derived an approximate expression for the sampling variance of d in this situation.

There are two main ways of measuring the amount of polymorphism in a given population. The haplotype diversity, h, was defined by Nei and Tajima (1981) as:

$$H = 1 - \sum_{i=1} p_i^2 \qquad \text{Eq.55}$$

where p_i is the frequency of the i-th haplotype. An estimate of H can be obtained as:

$$\hat{H} = \frac{2n}{2n-1}\left(1 - \sum_i \hat{p}_i^2\right) \qquad \text{Eq.56}$$

This definition is equivalent to that of heterozygosity or genic diversity described above.

A second way to measure the amount of polymorphism is by the average number of differences in the restriction sites between pairs of randomly chosen haplotypes. This number is given by:

$$v = \sum_{i,j} p_i p_j v_{ij} \qquad \text{Eq.57}$$

and an unbiased estimate is obtained as:

$$\hat{v} = \frac{n}{n-1} \sum_{i,j} \hat{p}_i \hat{p}_j v_{ij} \qquad \text{Eq.58}$$

However, these two measures of polymorphism depend on the length of the DNA fragment studied by restriction analysis. This can be avoided by using a measure of variation at the nucleotide level. As most polymorphisms in restriction sites are due

to nucleotide substitutions, this can be used to estimate nucleotide diversity, defined as:

$$d = \sum_{i,j} p_i p_j d_{ij} \qquad \text{Eq.59}$$

where d_{ij} is the proportion of nucleotide differences between haplotypes i and j and p_i and p_j are their respective population frequencies. We have already seen how d can be estimated from restriction site or restriction fragment data. An unbiased estimate of nucleotide diversity is then given by:

$$\hat{d} = \frac{n}{n-1} \sum_{i,j} \hat{p}_i \hat{p}_j d_{ij} \qquad \text{Eq.60}$$

Once an estimate of d for any given population is available, it is possible to extend the procedure in order to analyze the amount of divergence among populations. Let d_X represent the average number of nucleotide substitutions between a pair of haplotypes randomly chosen in population X. This number can be estimated as:

$$\hat{d}_X = \frac{n_X}{n_X - 1} \sum_{i,j} \hat{p}_i \hat{p}_j d_{ij} \qquad \text{Eq.61}$$

where n_X represents the sample size in population X. The average number of substitutions between pairs of haplotypes randomly chosen one from each population X and Y can be estimated from:

$$\hat{d}'_{XY} = \sum_{i,j} \hat{p}_i \hat{p}_j d_{ij} \qquad \text{Eq.62}$$

when the i-th and j-th haplotypes have been sampled from populations X and Y, respectively. As before, in this measure both an intra- and an interpopulation component of variability are included. The net amount of nucleotide substitutions between these two populations is then estimated by:

$$\hat{d}_{XY} = \hat{d}'_{XY} - \frac{\hat{d}_X + \hat{d}_Y}{2} \qquad \text{Eq.63}$$

If the rate of nucleotide substitution per site and per year between two populations that diverged t years ago is given by λ, then the expected value of d_{XY} is:

$$d_{XY} = 2\lambda t \qquad \text{Eq.64}$$

Hence, in order to compute the time since divergence between two populations (t), it is necessary to subtract from d_{XY} the average nucleotide difference between polymorphic alleles at the instant of separation. The sampling variance of d_A is given by:

$$Var(\hat{d}_{XY}) = Var(\hat{d}'_{XY}) + \frac{1}{4}\left[Var(\hat{d}_X) + Var(\hat{d}_Y)\right] - \left[Cov(\hat{d}'_{XY}.\hat{d}_X) + Cov(\hat{d}'_{XY}.\hat{d}_y)\right] \qquad \text{Eq.65}$$

Nei (1987) gives expressions for the corresponding variances and covariances in the above expression.

In the above derivations, it has been assumed that changes in recognition sites occur only once, and hence no provision has been made for superposition of substitutions. This is a good approximation as long as divergence between sequences is rather small, but when the number of changes increases, the approximation no longer holds. Then it becomes necessary to correct for the probability of several substitutions on the same recognition site. The corresponding expression was derived by Nei and Tajima (1983):

$$\delta_{XY} = -\frac{3}{4}\ln\left(1-\frac{4}{3}d'_{XY}\right) \qquad \text{Eq.66}$$

When several populations are analyzed, the method proposed by Lynch and Crease (1990) for partitioning nucleotide diversity the into intra- and interpopulation components described in the preceding section can be readily applied to both restriction site or restriction fragment data.

The analysis of RAPD data

The analysis of RAPD data has been hampered by the failure of usual methods to correct for the inability of detecting genotypes with dominant markers. This can result in a serious underestimation of the actual level of genetic diversity (Clark and Lanigan 1993). Recently, two methods for overcoming this difficulty and thus enabling the use of RAPD data have been proposed (Clark and Lanigan 1993; Lynch and Milligan 1994).

The method proposed by Clark and Lanigan (1993) uses the frequency of the absence of a fragment in a population sample as an estimate of the population frequency of recessive heterozygotes (q^2) and then uses this value to correct for the relative detectability of individuals who have one versus two copies of a fragment. Once this correction has been taken into account, data are treated in a very similar way to that already described for restriction fragment data. This is feasible if the following assumptions are met:

1. Amplification of a fragment is dependent on the primer hybridizing to the flanking sequences. In other words, if one single substitution is present in the sequence complementary to the primer this will not hybridize and the corresponding fragment will not be amplified.
2. Individuals being analyzed should diverge in less than 0.10, because the model does not allow for multiple hits. This assumption is usually met in population studies.
3. Sizes of the bands can be determined accurately and all different bands can be told apart from each other.
4. Different bands represent independent loci in linkage equilibrium. Incidentally, this is one of the less realistic assumptions in this model.
5. Populations are panmictic and samples are taken randomly.
6. Hardy-Weinberg equilibrium can be assumed for genotypic proportions.

As above, let P be the probability that no mutation has occurred at a primer site since the divergence from the common ancestor of two sequences. If F is the expected proportion of fragments that remain unchanged then, as for restriction fragments, Nei and Li (1979) showed the approximate relationship:

$$F = \frac{P^4}{3-2P} \qquad \text{Eq.67}$$

From the number of bands shared by two individuals (m_{XY}), and those present in individuals X (m_X) and Y (m_Y), the following estimate of F can be obtained:

$$\hat{F} = \frac{2m_{XY}}{m_X + m_Y} \qquad \text{Eq.68}$$

The expected nucleotide divergence between two sequences is $d = 2\lambda t$ if λ is the rate of nucleotide substitution per site and per year, and since $P = \exp(-r\lambda t)$, it is possible to estimate d from the relation:

$$\hat{d} = -\frac{2}{r}\ln \hat{P} \qquad \text{Eq.69}$$

The preceding estimate of d is the nucleotide divergence for a pair of haploid individuals. If several individuals from each of two populations are examined, it is possible to estimate the interpopulational nucleotide divergence following Nei and Miller (1990):

$$\hat{F}_{XY} = \frac{2\sum_{i,j} m_{X_iY_j}}{n_Y \sum_i m_{X_i} + n_X \sum_j m_{Y_j}} \qquad \text{Eq.70}$$

where m_{XiYj} is the number of bands shared by individual i from population X and individual j from population Y, m_{Xi} is the number of bands scored in individual i from population X and n_X and n_Y are the number of individuals sampled in the corresponding populations. Sample sizes are used to weigh the number of bands scored in each population in order to make the number of within and between-population comparisons the same.

The correction for dominance is based on the assumption of Hardy-Weinberg equilibrium. The expected heterozygosity under Hardy-Weinberg equilibrium is $2pq$. The conditional heterozygosity of band i in an individual from population X, given that band is observed, is defined as:

$$H_{X(i)} = \frac{2pq}{p^2 + 2pq} \qquad \text{Eq.71}$$

The numbers m_X, m_Y and m_{XY} can be tallied by summing over the $i = 1 - k$ bands for a pair of individuals from within and between populations X and Y:

$$m_X = \sum_i \left[4\left(1 - H_{X(i)}\right)^2 + 4H_{X(i)}\left(1 - H_{X(i)}\right) + H^2_{X(i)}\right], \qquad \text{Eq.72}$$

$$m_Y = \sum_i \left[4\left(1 - H_{Y(i)}\right)^2 + 4H_{Y(i)}\left(1 - H_{Y(i)}\right) + H^2_{Y(i)}\right],$$

$$m_{XY} = \sum_i \left[4\left(1 - H_{X(i)}\right)\left(1 - H_{Y(i)}\right) + 2\left(1 - H_{X(i)}\right)H_{Y(i)} + 2H_{X(i)}\left(1 - H_{Y(i)}\right) + H_{X(i)}H_{Y(i)}\right]$$

After the weighted values of m_{XY}, m_X and m_Y are tallied for all bands and pairs of individuals, F and d can be calculated from Equations 67 - 70.

Nei and Takezaki (1994) proposed a modified estimate of F values based on the frequencies of each band instead of their direct count, and on taking a geometric rather than an arithmetic average for comparing the shared bands with the bans present in the common ancestor. This leads to:

$$\hat{F} = \frac{\sum_i p_{X_i} p_{Y_i}}{\sqrt{\sum_i p_{X_i}^2 \sum_i p_{Y_i}^2}} \qquad \text{Eq.73}$$

where p_{Xi} represents the frequency of the i-th DNA fragment in population X. This frequency, for diploid organisms and in populations in Hardy-Weinberg equilibrium, can be estimated from:

$$p_{X_i} = 1 - \sqrt{Q_{X_i}} \qquad \text{Eq.74}$$

where Q_{Xi} is the frequency of individuals lacking the i-th fragment in population X.

From the estimate of nucleotide divergence, d, obtained using Eq. 69 it is possible to evaluate the interpopulational component of the total diversity as:

$$d_{XY} = d'_{XY} - \frac{d_X + d_Y}{2} \qquad \text{Eq.75}$$

When several different primers are used and sample sizes differ from one primer to another and/or primers have different lengths, it is possible to combine the different data into one single estimate by using an approach similar to Nei and Miller's (1990):

$$\hat{d}_{XY} = \frac{\sum_k \bar{n} r_k \hat{d}_{XY_k}}{\sum_k \bar{n} r_k} \qquad \text{Eq.76}$$

where $\bar{n}$ is the average number of individuals assayed in each population X and Y, r_k is the length of the k-th primer and $\hat{d}_{XYk}$ is the corresponding estimate of interpopulation divergence for that primer.

Lynch and Milligan (1994) have adopted a different approach for analyzing population structure using RAPDs. They simply assume that alleles from different loci do not comigrate to the same position in the gel, that the researcher is capable of matching bands from different lanes within and among gels, and that each locus can be treated as a two-allele system, with a presence and an absence allele. They adopt the following estimate for the gene frequency, q, of the null allele at one locus:

$$\hat{q} = \frac{\sqrt{\hat{x}}}{1 - \frac{Var(\hat{x})}{8\hat{x}^2}} \qquad \text{Eq.77}$$

where x is the frequency of null homozygotes. This is an asymptotically unbiased estimator of q and has lower bias than the correction proposed by Clark and Lanigan (1993).

Once gene frequencies have been estimated, it is possible to estimate gene diversity within a population. The usual measure of gene diversity:

$$H_{X_i} = 2p_{Xi}q_{Xi} = 1 - \sum_i p_{Xi}^2 \qquad \text{Eq.78}$$

which is the probability that two genes randomly chosen from population X differ at the i-th locus, is equivalent to the expected herterozygosity under Hardy-Weinberg equilibrium. An estimator of this quantity is given by

$$\hat{H}_{X_i} = 2\hat{q}_{Xi}\hat{p}_{Xi} + 2Var(\hat{q}_{Xi}) \qquad \text{Eq.79}$$

whose sample variance is approximately:

$$Var(\hat{H}_{X_i}) = 4(1 - 2\hat{q}_{Xi})^2 Var(\hat{q}_{Xi}) \qquad \text{Eq.80}$$

If L loci have been sampled in population X, the average gene diversity in this population is:

$$\hat{H}_X = \frac{1}{L}\sum_{i=1}^{L}\hat{H}_{X_i} \qquad \text{Eq.81}$$

and if n populations have been sampled, the average within-population gene diversity can be estimated by:

$$\hat{H}_W = \frac{1}{n}\sum_{X=1}^{n}\hat{H}_X \qquad \text{Eq.82}$$

Expressions for the corresponding sample variances can be found in Lynch and Milligan (1994).

The heterozygosity between populations X and Y at the i-th locus can be estimated by:

$$\hat{H}'_{XY_i} = \hat{q}_{Xi} + \hat{q}_{Yi} - 2\hat{q}_{Xi}\hat{q}_{Yi} \qquad \text{Eq.83}$$

If there is no population subdivision, the gene frequencies in all populations are the same, so $\hat{H}'_{XY_i} = \hat{H}_{Xi} = \hat{H}_{Yi}$, and the interpopulational component of diversity can be estimated as usual by:

$$\hat{H}_{XY_i} = \hat{H}'_{XY_i} - \frac{\hat{H}_{X_i} + \hat{H}_{Y_i}}{2} \qquad \text{Eq.84}$$

Averaging over all loci, the estimated mean gene diversity between populations X and Y is:

$$\hat{H}_{XY} = \frac{1}{L}\sum_{i=1}^{L}\hat{H}_{XY_i} \qquad \text{Eq.85}$$

and the mean between population gene diversity can be obtained by averaging over all possible pairs of populations:

$$\hat{H}_B = \frac{2\sum_{X<Y}\hat{H}_{XY}}{n(n-1)} \qquad \text{Eq.86}$$

Lynch and Milligan (1994) propose an asymptotically unbiased estimate of F_{ST} by using:

$$\hat{F}_{ST} = \frac{\hat{H}_B}{\hat{H}_B + \hat{H}_W}\left[1 + \frac{\hat{H}_B Var(\hat{H}_W) - \hat{H}_W Var(\hat{H}_B) + (\hat{H}_B - \hat{H}_W)Cov(\hat{H}_B, \hat{H}_W)}{\hat{H}_B(\hat{H}_B + \hat{H}_W)^2}\right]^{1} \qquad \text{Eq.87}$$

The expression for the variances and covariances in the above equation can be found in Lynch and Milligan (1994). There are two packages of freeware programmes for the analysis of RAPD data, RAPDISTANCE (Armstrong *et al.* 1995) and RAPDIS (Dopazo 1995).

An alternative approach for estimating F-statistics from RAPDs data has been used by Huff *et al.* (1993) and Peakall *et al.* (1995). They have used the already described analysis of molecular variance (Excoffier *et al.* 1992) implemented in the programme WINAMOVA in order to estimate population differentiation statistics. This procedure is especially useful when more than two population levels have to be considered.

One of the potential uses of RAPDs is their capability for providing estimates of the relatedness among individuals in the population. This measure is based on the expectation that related individuals will have more similar genotypes than nonrelatives, and hence the fraction of loci for which two individuals are identical should increase with the degree of relatedness. An estimate for the relatedness, r, between individuals a and b using data at the i-th locus is given by (Lynch and Milligan 1994):

$$\hat{r}_{ab_i} = \frac{S_{ab_i} - \hat{\theta}_i}{1 - \hat{\theta}_i} + \frac{1 - S_{ab_i}}{[1 - \hat{\theta}_i]^3} Var(\hat{\theta}_i) \qquad \text{Eq.88}$$

where $S_{abi} = 1$ or 0 denotes whether individuals a and b have the band at the i-th locus or not, θ_i is the probability that two nonrelatives have matching phenotypes at the locus, and:

$$\hat{\theta}_i = 1 - 2Q_i(1 - Q_i)\left(1 - \frac{1}{N}\right) \qquad \text{Eq.89}$$

$$Var(\hat{\theta}_i) = \frac{4Q_i(1 - Q_i)(2Q_i - 1)^2}{N} \qquad \text{Eq.90}$$

A more accurate estimate is obtained by averaging over all L loci:

$$\hat{r}_{ab} = \frac{1}{L}\sum_{i=1}^{L}\hat{r}_{ab_i} \qquad \text{Eq.91}$$

Such an analysis is only applied to polymorphic loci. Nevertheless, due to the overlap between the distributions of similarity for RAPDs for different degrees of relatedness, the utility of relatedness estimation using RAPDs is rather limited (Lynch and Milligan 1994).

Analyzing DNA fingerprinting data

Hypervariable minisatellite DNA, when analyzed by Southern hybridization following restriction digestion, produce DNA fingerprints. These fingerprints usually correspond to several loci which share a core sequence, hence showing all at once in a single gel. These loci usually exhibit large allelic diversity, and hence only on a few occasions will individuals from exogamous populations sampled at random show exactly the same DNA fingerprint pattern.

In order to develop similarity estimates, potential indicators of the relative level of population homozygosity, a few assumptions on the technical ability of the researcher have to be made (Lynch 1990; Lynch and Crease 1990). First, it is assumed that the DNAs of individuals to be compared are run in close lanes and/or with adequate controls so that errors on the assignment of identity to pairs of fragments are minimized. Second, it is assumed that all individuals are sampled at random from the population. Third, it is assumed that all comigration of non-allelic markers can be resolved either by differences in band intensity or by some other means. Fourth, marker loci are assumed unlinked and in Hardy-Weinberg equilibrium within and among loci. And last, the same set of homologous loci is tested in all individuals.

Similarity is usually defined as the fraction of shared bands. For two individuals, x and y, it can be defined as the number of common fragments (n_{xy}) divided by an estimate of the number of fragments in any individual:

$$S_{xy} = \frac{2n_{xy}}{n_x + n_y} \qquad \text{Eq.92}$$

It is necessary to relate this index with a population genetic parameter such as the identity by state between pairs of individuals and population homozygosity. The identity in state for two individuals can be defined as 100% for pairs of individuals AA-AA or Aa-Aa and as 50% for pairs such as AA-Aa or Aa-Aa'. The expected genotypic identity in state for a panmictic population is:

$$E(I) = \frac{\sum_k \sum_i P_{ki}^2 + P_{ki}^2 (1-P_{ki})^2}{L.} \qquad \text{Eq.93}$$

where p_{ki} is the frequency of the i-th allele at the k-th locus and L is the number of loci. Alternatively, the identity in state can be defined from the standpoint of gametes taken at random from two individuals. Under panmictic mating, the expected gametic identity in state is equivalent to population homozygosity:

$$E(H)=\frac{\sum_k\sum_i p_{ki}^2}{L} \qquad \text{Eq.94}$$

Jeffreys *et al.* (1985) and Lynch (1988) showed that:

$$E(S)=\frac{\sum_k\sum_i p_{ki}^2(2-p_{ki})}{L} \qquad \text{Eq.95}$$

Hence, the similarity index is always a biased estimator by excess both of I and H. Lynch (1988) developed the sampling theory for the similarity index. When large numbers of polymorphic loci are sampled, the sampling variance of the average population similarity can be directly estimated from the observed data as:

$$Var(\bar{S})=\frac{NVar(S_{xy})+2N'Cov(S_{xy},S_{xz})}{N^2} \qquad \text{Eq.96}$$

where N is the total number of similarity measures used and N' is the number of pairs of those measures shared by an individual. When average similarities from different populations are being compared and there is no certainty that the same loci have been sampled in all populations or we are interested in making inferences on properties of the whole genome from the sampled loci, it is convenient to use the following expression that takes into account both the effect of sampling on the studied loci and the error due to sampling a finite number of loci:

$$Var'(S_{xy})=\frac{2\bar{S}(1-\bar{S})(2-\bar{S})}{\bar{n}(4-\bar{S})} \qquad \text{Eq.97}$$

where $\bar{n}$ is the average number of bands present in any individual.

A measure of interpopulation similarity corrected for intrapopulation similarity is given by:

$$\bar{S}_{ij}=1+\bar{S}'_{ij}-\frac{\bar{S}_i+\bar{S}_j}{2} \qquad \text{Eq.98}$$

where $\bar{S}_i$ is the average similarity for individuals belonging to the i-th populations and $\bar{S}'_{ij}$ is the average similarity between pairs of individuals randomly sampled from populations i and j. As the similarity index is not an unbiased estimator of population homozygosity, caution should be taken when using it for estimating the usual measure of population subdivision, Wright's F-statistics. Nevertheless, if the biases

corresponding to $\overline{S}_{ij}$, $\overline{S}_i$ and $\overline{S}_j$ are approximately equal, then they cancel out in the above expression for $\overline{S}_{ij}$. Consequently, $\overline{D}_{ij} = 1 - \overline{S}_{ij}$ is an unbiased estimator of the interpopulation genetic diversity.

Let D_b denote the average of $\overline{D}_{ij}$ for all i, j and let D_w denote the average value of $1 - S_i$. Then:

$$F' = \frac{D_b}{D_b + D_w} \qquad \text{Eq.99}$$

provides a downwards biased estimate of population subdivision.

An unbiased estimate of the average heterozygosity for a system with L loci is given by (Stephens *et al.* 1992):

$$\hat{H} = \frac{\frac{2n}{2n-1}\sum_{i=1}^{L}\left(1-\sum_{j=1}^{A_i} p_{ij}^2\right)}{L} = \frac{2n}{2n-1}\left(1-\frac{\sum_{k=1}^{A} p_k^2}{L}\right) \qquad \text{Eq.100}$$

where A_i is the number of alleles in the i-th locus and p_{ij} is the estimated frequency of the j-th allele in the i-th locus. The second equality represents the lack of importance for the estimation of heterozygosity of the specific distribution of alleles throughout loci.

As every band or allele in a fingerprint is effectively dominant, it is necessary to estimate the allele frequency (p_k) from the frequency with which the k-th band appears (S_k). Assuming Hardy-Weinberg equilibrium for the genotypes, then:

$$p_k = 1 - \sqrt{1 - s_k} \qquad \text{Eq.101}$$

The individual p_k estimates can be summed up to provide an estimate of L (Gilbert *et al.* 1990). An improved estimate of heterozygosity is obtained when monomorphic and polymorphic bands are considered separately. Let L_M represent the number of monomorphic loci and A_P that of polymorphic bands, such that $A_P = A - L_M$, being A the total amount of bands. Heterozygosity will be at a maximum when all allele frequencies in polymorphic loci are uniform, i.e. when for each allele frequency $p_{ij} = 1/A_{ij} = L_P / A_P$. Then the estimate of the maximum heterozygosity will be:

$$H_{max} = \frac{2n}{2n-1}\frac{L_P\left(1-\frac{L_P}{A_P}\right)}{L_M + L_P} \qquad \text{Eq.102}$$

and the value of L_P is given by

$$L_P = \sqrt{L_M A} - L_M. \qquad \text{Eq.103}$$

Microsatellites and SSR loci

The main difficulty posed by microsatellite loci for their use in the evaluation of genetic distance is their relatively high mutation rate. This makes it difficult to adopt any of the two main mutation models used in population genetics, the infinite alleles or the stepwise mutation model. There is still uncertainty as to whether allele sizes are unconstrained or whether there are certain limits to the number of repeats present (Estoup *et al.* 1995; Garza *et al.* 1995; Meyer *et al.* 1995). Assuming a stepwise mutation model, Slatkin (1995) and Goldstein *et al.* (1995) have recently proposed a distance measure for microsatellite alleles. The distance between two alleles is a simple transformation of the number of repeat units. The within population measure of distance is obtained as the average sum of squares of the differences in number of repeats between alleles:

$$S_{Wj} = \frac{2}{2n(2n-1)} \sum_{i<i'} \left(a_{ij} - a_{i'j}\right)^2 \qquad \text{Eq.104}$$

where a_{ij} is the allele size of the i-th copy ($i = 1,\ldots,2n$) in the j-th population ($j = 1,\ldots,d_s$). The average within population distance S_w from Slatkin is equivalent to D_0 from Goldstein *et al.* (1995):

$$S_W = \frac{1}{d_s} \sum_{j=1}^{d_s} S_{Wj} \qquad \text{Eq.105}$$

In order to estimate the average distance between all possible pairs of alleles, it is necessary to define the between population component, S_b as:

$$S_B = \frac{2}{(2n)^2 d_s(d_s-1)} \sum_{j<j'} \sum_{i<i'} \left(a_{ij} - a_{i'j'}\right)^2 \qquad \text{Eq.106}$$

which is equivalent to D_1 of Goldstein *et al.* (1995). The global distance is obtained by a weighted average of the intra- and interpopulation components:

$$\bar{S} = \frac{2n-1}{2nd_s-1} S_W + \frac{2n(d_s-1)}{2nd_s-1} S_B \qquad \text{Eq.107}$$

where the coefficients represent the probability of choosing two different copies of the locus from the same and from different populations, respectively. In practice, it is easier to compute S_w and $\bar{S}$. directly from the variances of allele sizes, as S_w is twice the average of the variances of allele size within each population and $\bar{S}$ is twice the estimated variance of allele size in the collection of populations together. MICROSAT is a programme that can be used for computing these distances.

Given that S_w and $\bar{S}$ are proportional to the within-population and total variances, the fraction:

$$R_{ST} = \frac{\bar{S} - S_W}{\bar{S}} \qquad \text{Eq.108}$$

has the same properties for microsatellite loci that follow the stepwise mutation model as F_{ST} has for allozyme loci. R_{ST} is simply the fraction of the total variance in allele size that is due to interpopulation differences.

An extension of this method, incorporating the analysis of microsatellite data into an ANOVA framework, has been recently proposed by Michalakis and Excoffier (1995). In this method, the partition of genetic variance at different levels is achieved by means of an analysis of molecular variance, as described above, by using the programme WINAMOVA.

An alternative distance measure, the shared allele distances D_{AS} (Chakraborty and Jin 1993) has been advocated by Estoup *et al.* (1995) for use with microsatellite data. This distance is computed by averaging the values over all loci at pairs of individuals. For each locus, the distance is 1 if both individuals have the same genotype, 0 if they have no allele in common and 0.5 if they share only one allele. With the use of this distance, it is possible to group individuals by any of the different methods of clustering (see above). The programme MICROSAT can also be used to compute D_{AS} distances.

Shriver *et al.* (1995) have proposed the use of a stepwise weighted genetic distance measure (D_{SW}), which is an extension of Nei's minimum genetic distance. This measure has several advantages over minimum and standard genetic distances when applied to loci evolving via a stepwise mutation mechanism. Let p_{Xi} represent the allele frequency of the i-th allele in population X. The proposed distance weighs the probability that two alleles are different when randomly sampled from one or two populations by the absolute value of the difference in steps (number of repeats for tandem repeat loci) between the two alleles. That is:

$$d_{XW} = \sum_{i \neq j} p_{Xi} p_{Xj} \delta_{ij} \qquad \text{Eq.109}$$

$$d_{YW} = \sum_{i \neq j} p_{Yi} p_{Yj} \delta_{ij} \qquad \text{Eq.110}$$

$$d_{XYW} = \sum_{i \neq j} p_{Xi} p_{Yj} \delta_{ij} \qquad \text{Eq.111}$$

where:

$$\delta_{ij} = |i - j| \qquad \text{Eq.112}$$

D_{SW} can then be defined as:

$$D_{SW} = d_{XYW} - \frac{d_{XW} + d_{YW}}{2} \qquad \text{Eq.113}$$

As in the case of Nei's distance measures, D_{SW} can be estimated by means of the unbiased estimates of d_{XW}, d_{YW} and d_{XYW}, given by:

$$\hat{d}_{XW} = \frac{n_X}{n_X - 1} \sum_{i \neq j} \hat{p}_{Xi} \hat{p}_{Xj} \delta_{ij} \qquad \text{Eq.114}$$

$$\hat{d}_{YW} = \frac{n_Y}{n_Y - 1} \sum_{i \neq j} \hat{p}_{Yi} \hat{p}_{Yj} \delta_{ij} \qquad \text{Eq.115}$$

$$\hat{d}_{XYW} = \sum_{i \neq j} \hat{p}_{Xi} \hat{p}_{Yj} \delta_{ij} \qquad \text{Eq.116}$$

where n_X and n_Y are the number of chromosomes sampled from populations X and Y, respectively. For the estimation of D_{SW} from multilocus data, the averages over all loci of the above estimators can be used.

Acknowledgements

This work has been supported by grant PB93-0350 from DGICYT (Ministerio de Educación y Ciencia, España).

References

Armstrong, J., A. Gibbs, R. Peakall and G. Weiller. 1995. RAPDistance programs; Version 1.03 for the analysis of patterns of RAPD fragments.

Bhattacharyya, A. 1946. On a measure of divergence between two multinomial populations. Sankhya 7:401-406.

Brown, A.H.D. 1970. The estimation of Wright's fixation index from genotypic frequencies. Genetica 41:399-406.

Cavalli-Sforza, L.L. and A.W.F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. Am. J. Hum. Genet. 19:233-257.

Chakraborty, R. 1974. A note on Nei's measure of gene diversity in a substructured population. Hummangenetik 21:85-88.

Chakraborty, R. and L. Jin. 1993. A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. Pp. 153-175 *in* DNA Fingerprinting: State of the Science (S.D.J. Peña, R. Chakraborty, T.J. Epplen and A.J. Jeffreys, eds.). Birkhauser Verlag, Basel.

Clark, A.G. and C.M.S. Lanigan. 1993. Prospects for estimating nucleotide divergence with RAPDs. Molecular Biology and Evolution 10:1096-1111.

Cockerham, C.C. 1969. Variance of gene frequencies. Evolution 23:72-84.

Cockerham, C.C. 1973. Analysis of gene frequencies. Genetics 74:679-700.

Curie-Cohen, M. 1982. Estimates of inbreeding in a natural population: A comparison of sampling properties. Genetics 100:339-358.

Dopazo, J. 1995. RAPDIS (In preparation).

Estoup, A., L. Garnery, M. Solignac and J.M. Cornuet. 1995. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics 140:679-695.

Excoffier, L., P.E. Smouse and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

Gilbert, D.A., Y.A. Reid, M.H. Gail, D. Pee, C. White, R.J. Hay and S.J. O'Brien. 1990. Application of DNA fingerprints for cell-line individualization. Am. J. Hum. Genet. 47:499-514.

Goldstein, D.B., A. Ruiz Linares, L.L. Cavalli-Sforza and M.W. Feldman. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-471.

González-Candelas, F., S.F. Elena and A. Moya. 1995. Approximate variance of nucleotide divergence between two sequences estimated from restriction fragment data. Genetics 140:1443-1446.

Huff, D.R., R. Peakall and P.E. Smouse. 1993. RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm.]. Theor. Appl. Genet. 86:927-934.

Jeffreys, A.J., V. Wilson and S.L. Thein. 1985. Hypervariable "minisatellite" regions in human DNA. Nature 314:67-73.

Jin, L. and J.W.H. Ferguson. 1990. Neighbor-joining tree and UPGMA tree software.

Kaplan, N. and K. Risko. 1981. An improved method for estimating sequence divergence of DNA using restriction endonuclease mappings. J. Molec. Evolution 17:156-162.

Li, C.C. and D.G. Horvitz. 1953. Some methods of estimating the inbreeding coefficient. Am. J. Hum. Genet. 95:107-117.

Li, W.-H. 1981. A simulation study if Nei and Li's model for estimating DNA divergence from restriction enzyme maps. J. Molec. Evolution 17:251-255.

Long, J.C. 1986. The allelic correlation structure of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright's F-statistics. Genetics 112:629-647.

Long, J.C., P.E. Smouse and J.W. Wood. 1987. The allelic correlation structure of Gainj. and Kalam-speaking people. II. The genetic distance between population subdivisions. Genetics 117:273-283.

Lynch, M. 1988. Estimation of relatedness by DNA fingerprinting. Molec. Biol. and Evolution 5:584-599.

Lynch, M. 1990. The similarity index and DNA fingerprinting. Molec. Biol. and Evolution 7:478-484.

Lynch, M. and B.G. Milligan. 1994. Analysis of population genetic structure with RAPD markers. Molec. Ecol. 3:91-99.

Lynch, M. and T.J. Crease. 1990. The analysis of population survey data on DNA sequence variation. Molec. Biol. and Evolution 7:377-394.

Mahalanobis, P.C. 1936. On the generalized distance in statistics. Proc. Natl. Inst. Sci. India 2:49-55.

Manly, J.B.F. 1986. Multivariate Statistical Methods. A Primer. Chapman and Hall, London.

Michalakis, Y. and L. Excoffier. 1995. A generic estimation of population subdivision using distances between alleles with specieal interest to microsatellite loci. Genetics (In press).

Miller, J.C. 1990. A program for computing distances between phylogenetic groups based on restriction-site or fragment data.

Nei, M. 1972. Genetic distance between populations. The American Naturalist 106:283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Nat. Acad. Sci., USA 70:3321-3323.

Nei, M. 1973. The theory and estimation of genetic distance. Pp. 45-54 *in* Genetic Structure of Populations (N.E. Morton, ed.). University Press of Hawaii, Honolulu.

Nei, M. 1986. Definition and estimation of fixation indices. Evolution 40:643-645.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei, M. and F. Tajima. 1981. DNA polymorphism detectable by restriction endonucleases. Genetics 97:145-163.

Nei, M. and F. Tajima. 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207-217.

Nei, M. and J.C. Miller. 1990. A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. Genetics 125:873-879.

Nei, M. and N. Takezaki. 1994. Estimation of genetic distances and phylogenetic trees from DNA analysis. Proc. 5th World Cong. Genet. Appl. Livestock Prod. 21:405-412.

Nei, M. and R.K. Chesser. 1983. Estimation of fixation indices and gene diversities. Am. J. Hum. Genet. 47:253-259.

Nei, M. and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Nat. Acad. Sci. USA 76:5269-5273.

Peakall, R., P.E. Smouse and D.R. Huff. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. Molec. Ecol. 4:135-147.

Rasmussen, D.I. 1964. Blood group polymorphism and inbreeding in natural populations of the deer mouse *Peromyscus maniculatus*. Evolution 18:219-229.

Raymond, M. and F. Rousset. 1995. GENEPOP (V. 1.2): A population genetics software for exact tests and ecumenicism. J. Heredity (in press).

Reynolds, J. 1981. Genetic Distance and Coancestry. North Carolina State University, Raleigh, NC, USA.

Rogers, J.S. 1972. Measures in genetic similarity and genetic distance. Studies in Genetics VII. University of Texas Publ. 7213:145-153.

Rohlf, F.J. and D.E. Slice. 1992. NTSYS-pc.

Rozas, J. and R. Rozas. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating Population Genetics parameters from DNA sequence data. Computer Application in Biosciences (in press).

Saitou, N. and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. J. Molec. Evolution 4:406-425.

Shriver, M.D., L. Jin, E. Boerwinkle, R. Deka, R.E. Ferrell and R. Chakraborty. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. Molec. Biol. and Evolution 12:914-920.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.

Sneath, P.H.A. and R.R. Sokal. 1973. Numerical Taxonomy. W.H. Freeman, San Francisco.

Sokal, R.R. and C.D. Michener. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 28:1409-1438.

Stephens, J.C., D.A. Gilbert, N. Yuhki and S.J. O'Brien. 1992. Estimation of heterozygosity for single-probe multilocus DNA fingerprints. Molec. Biol. and Evolution 9:729-743.

Tajima, F. 1993. Measurement of DNA polymorphism. Pp. 37-59 *in* Mechanisms of Molecular Evolution (N. Takahata and A.G. Clark, eds.). Sinauer, Sunderland.

Weir, B.S. 1990. Genetic Data Analysis. Sinauer Associates Inc. Sunderland.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F statistics for the analysis of population structure. Evolution 38:1358-1370.

Workman, P.L. and J.D. Niswander. 1970. Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. Am. J. Hum. Genet. 22:24-49.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

Wright, S. 1951. The genetical structure of populations. Ann. Eugen. 15:323-354.

Wright, S. 1978. Evolution and the Genetics of Populations. IV. Variability within and among Natural Populations. University of Chicago Press, Chicago.

Appendix

Computer programmes for use in population genetics and analysis of molecular variation

BIOSYS-1 (Swofford and Selander 1981, 1989)
Programme for the analysis of allelic variation in population genetics.
DnaSP (Rozas and Rozas 1995)
This is interactive programme for estimating population genetics parameters from DNA sequence data.
http://www.ebi.ac.uk
ftp://ftp.ebi.ac.uk
GENEPOP (Raymond and Rousset 1995)
GENEPOP is a population genetic software package, able to perform two major tasks:
1) It computes exact tests: for Hardy-Weinberg equilibrium, for population differentiation and for genotypic disequilibrium among pairs of loci.
2) It converts the input GENEPOP file to formats used by other programmes, like Biosys (Swofford and Selander 1981), Diploid (Weir 1990), Linkdos (Garnier-Gere and Dillman 1992) and M. Slatkin's (1993) isolation-by-distance programme (the last three programmes are also provided with GENEPOP, with the authorization of their authors).

ftp:// ftp.cefe.cnrs-mop.fr/ pub/msdos/genepop
MICROSAT (Goldstein *et al.* 1995)
Programme for computing distance measures with microsatellite data.
http://lotka.stanford.edu/research/microsat.html
ftp://lotka.stanford.edu/pub/Programs/microsat.c
NEIGHBOR (Jin and Ferguson 1990)
NJTREE, UPGMA and TDRAW is a group of software used for creating neighbour-joining trees or UPGMA trees.
NTSYS (Rolfe and Slyce 1992)
General package for multivariate analysis in population and evolutionary biology.
RAPDIS (Dopazo 1995)
Programme for the analysis of RAPD data.
http://www.tdi.es/
RAPDISTANCE (Armstrong *et al.* 1995)
RAPDistance Programmes; Version 1.03 for the Analysis of Patterns of RAPD Fragments.
ftp://life.anu.edu.au/pub/RAPDistance
http://life.anu.edu.au/molecular/software/rapd.html
RESTSITE (Miller 1990)
Programme for computing distances between phylogenetic groups based on restriction-site or fragment data.
WINAMOVA (Excoffier *et al.* 1992)
Programme for the analysis of molecular variance.
ftp://acasun1.unige.ch/pub/comp/win/amova
http://acasun1.unige.ch/LGB/Software/Windoze/amova

Managing, storing and using molecular data

David E. Matthews[1] *and Olin D. Anderson*[2]
[1]Cornell University, Dept. Plant Breeding and Biometry, Ithaca, NY, USA
[2]USDA Western Regional Research Laboratory, Albany, CA, USA

Introduction

Molecular data are being collected on a large scale in the world's genome mapping and sequencing projects. Data management for these projects is similar in some ways to the management of molecular data for germplasm resources, though there are also many differences. This article describes some aspects of **GrainGenes, the *Triticeae* Genome Database**, particularly aspects relevant to germplasm resources data.

The requirements for managing and using molecular data are of two basic types. Laboratory databases manage primary data for monitoring and directing work flow within a project, whereas archival databases manage more summarized results for use within a project and for sharing with other projects. An example of a large laboratory database is that of the Japanese Rice Genome Research Program (RGP) [RICE GENOME RESEARCH PROGRAM (RGP) HOME PAGE. URL: http://www.staff.or.jp], using 4th Dimension software for Macintosh. For archival purposes, the most popular genome database software is ACEDB, which will be used for the examples in this article.

Genome databases to archive molecular information

ACEDB

Unlike 4th Dimension, ACEDB was never intended to be general purpose database software. Originally developed by and for researchers on *Caenorhabditis elegans* ("**A** *C. Elegans* **DataBase**"), it is primarily designed for genome data and includes special graphical modules for displaying genetic maps (Fig. 1), physical maps and sequences.

Some subsidiary data types important to genome projects are also relevant to the management of molecular data on genetic resources: DNA clones, genetic stocks, bibliographic references and colleagues' addresses. The general display for these and any other curator-defined data class is a multiwindowed text interface, one window per record, with relations between records available as hypertext links. There is also a facility for displaying image files of photographs or drawings, which can be linked to any data record.

ACEDB is free software primarily designed to run under Unix, though a Macintosh version is also available. Like some other Unix database software such as Oracle, it can be gatewayed to the World Wide Web to allow Internet access. ACEDB/Web gateways for most of the USDA-sponsored plant genome databases, as well as a number of genome databases for man and microorganisms are running on the Agricultural Genome Information Server [Agricultural Genome Information Server (AGIS). URL: http://probe.nalusda.gov:8000], maintained by the USDA National Agricultural Library (NAL). This site is also the primary source for information and documentation about the ACEDB software itself [ACEDB Documentation. URL: http://probe.nalusda.gov:8000/ acedocs/index.html].

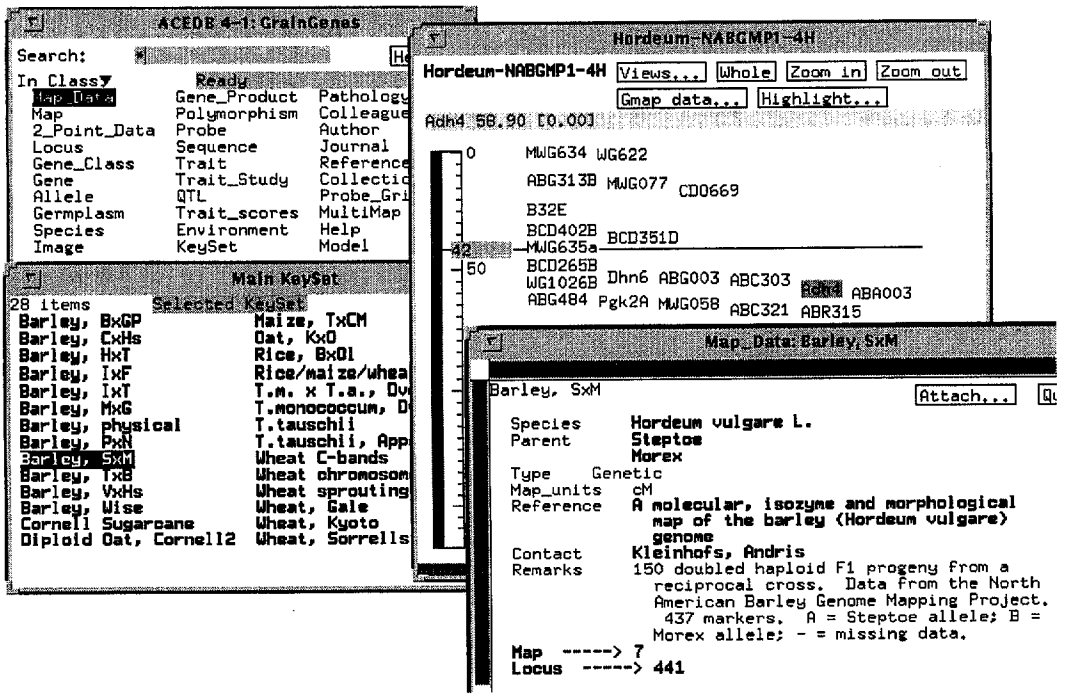

Fig. 1. The ACEDB user interface. The main window (upper left) lists the data classes. Clicking the mouse on a class produces a list of records in the Keyset window. Clicking on the name of a record shows the record in a text window (lower right) or a graphical display such as the genetic map. Boldface words in the text window, and most words in the map display, are links to other records: clicking on them causes the corresponding records to be displayed, each in a separate window.

The value of a database gateway to the Web is not only that it allows individuals to query the database anonymously via free, easy-to-use client software such as Netscape. These gateways also provide a simple mechanism for inter-database connectivity, as described below.

GrainGenes

The USDA plant genome databases — **GrainGenes, MaizeDB, Soybase, TreeGenes, RiceGenes, SolGenes** [Paul, E. *et al.* 1994], and many more — are highly individualistic, according to the interests of their diverse research communities. They share a goal of using genome maps in support of plant breeding, so some elements are common to most of them, especially the characterization of DNA clones and germplasm, and information about molecular markers linked to genes or quantitative trait loci.

Areas of particular emphasis in **GrainGenes** include:

- comparative mapping between species
- catalogues of genes, alleles and their reference stocks
- genotypes of cultivars and breeding material with regard to disease resistance, quality and other agronomically important genes
- taxonomy of the *Triticeae*
- diseases, insect pests and other pathologies

A recent profile of the data classes in **GrainGenes** is as follows.

Map	28	Wheat, oat, diploid relatives, barley, sugarcane
Linkage_Group	260	
2_Point_Data	20	Molecular markers for genes and QTLs
Locus	6400	
Probe	3300	
Polymorphism	1400	Many with images of autoradiograms
Sequence	200	End sequences of probes
Gene	660	
Allele	630	
Gene_Product	60	HMW glutenin subunits
Germplasm	11000	Wheat, rye, triticale
Species	1400	Including plants, pathogens and insects
Trait_Scores	14000	24th International Spring Wheat Yield Nursery
QTL	4	Raw data and statistical analyses from QTL studies
Pathology	450	With images of symptoms
Image	1400	
Colleague	1000	
Reference	1400	

Genome databases for germplasm characterization

In addition to the classical mandate of a genome database to archive information about genetic stocks for mutations and chromosomal derangements, many of the USDA plant genome databases have taken on the more complex challenge of characterizing cultivars, breeding lines and other germplasm resources. A few, including **GrainGenes** and **Soybase**, have even ventured to load data on evaluations of environment-sensitive traits, data largely imported from real germplasm databases which now have Web gateways of their own: GRIN [Search GRIN for accessions. URL: http://www.ars-grin.gov/cgibin/npgs/html/search.pl?] and CIMMYT [AGIS: Database: CIMMYT. URL: http://probe.nalusda.gov:8300/cgi-bin/browse/cimmyt].

Strictly genetic characters such as alleles, gene products and DNA polymorphisms are simpler to represent in a database. Although, databasing trait evaluations well is very difficult. Alleles and gene products such as isozymes are simply connected to the germplasm records by many-to-many relations:

Germplasm : "Abbondanza"

Allele	Glu-A1a (Triticum)
Allele	Glu-B1u (Triticum)
Allele	Glu-D1a (Triticum)
Gene_product	Glu-1 subunit 1
Gene_product	Glu-1 subunit 8
Gene_product	Glu-1 subunit 7*
Gene_product	Glu-1 subunit 12
Gene_product	Glu-1 subunit 2
Development_site Italy	
Data_source	Graybosch, Robert A. 94.04

Allele :	"Glu-A1a (Triticum)"	
Gene	Glu-A1 (Triticum)	
Gene_product	Glu-1 subunit 1	CRC-11-29
Germplasm	Abbazia	JBR-43-17
Germplasm	Abbondanza	JBR-43-17
Germplasm	ABE	PBR-110-48
...		
Gene_Product :	"Glu-1 subunit 1"	
Allele	Glu-A1a (Triticum)	
Germplasm	Abbazia	JBR-43-17
Germplasm	Abbondanza	JBR-43-17
Germplasm	ABE	PBR-110-48
...		

and so on, where the codes CRC-11-29, JBR-43-17 etc. are reference citations and, of course, are hypertext-linked to the corresponding reference records in the database.

Molecular polymorphisms

Molecular polymorphisms generally are not given individual names in the way alleles and gene products are, in part because they are more abundant and complex. They can be represented in several ways depending on the purposes for which the data are to be used. For example, the following format is used in GrainGenes for an RFLP survey of a small number of cultivars against a large number of probes.

Polymorphism : "BCD127 EcoRI"
Probe BCD127
Enzyme EcoRI

TABLE	Filter	Size (Kb)	Intensity	Germplasm
Size	T1639	1.8	Dark	Oryza sativa IR36
		20.3	Dark	Ogle
		14	Dark	Ogle
		6.7	Dark	Ogle
		14.9	Dark	Shin Ebisu 16
		11.9	Faint	Shin Ebisu 16
		5.9	Faint	Shin Ebisu 16
		18.7	Dark	CHINESE SPRING
		9.9	Dark	CHINESE SPRING
		2.3	Faint	CHINESE SPRING
		9.3	Faint	Saccharum spontaneum SES208
		3.5	Faint	Saccharum spontaneum SES208
		3.1	Faint	Saccharum spontaneum SES208
		2.9	Dark	Saccharum spontaneum SES208
		2.7	Faint	Saccharum spontaneum SES208
	A613	23.7	Faint	Kanota Clav2265
				Ogle Clav9401
		15.8	Faint	Kanota Clav2265
		6.4	Medium	Kanota Clav2265
		15.6	Faint	Ogle Clav9401
		6.7	Medium	Ogle Clav9401
Image	BCD127 BROWS autoradiogram			
	BCD127 KO autoradiogram			

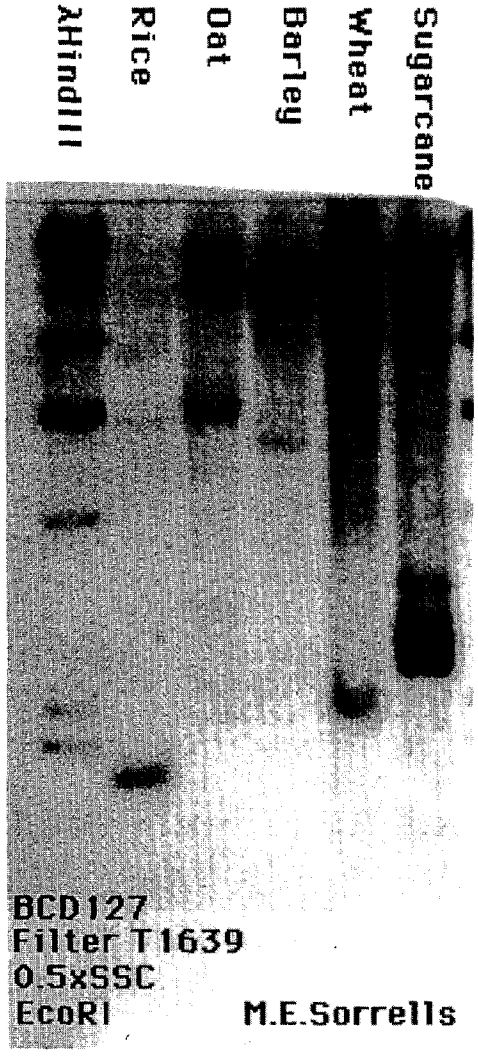

Fig. 2. Image "BCD127 BROWS Autoradiogram". From an RFLP survey of barley, rice, oats, wheat and sugarcane.

Individual polymorphism bands can be designated by their estimated molecular size, as in this example, or by arbitrary numbering. In either case attaching an appropriately labelled image of the autoradiogram is extremely valuable and is not expensive. The image of filter T1639 in Figure 2 occupies about 14 kilobytes as a 19-colour GIF file, and on a video display it faithfully reproduces the original autoradiogram.

Another way of representing polymorphisms is exemplified in the following two records excerpted from a survey of 60 probes against 80 oat cultivars. The results are presented both from the "Probe side" and the "Germplasm side".

Polymorphism : "BCD385 EcoRI"

Probe	BCD385					
Enzyme	EcoRI					
Band_size	17.2	14.9	13.4	12.4	4.6	
Pattern	0	1	1	0	0	Advance Clav3845
						Banner Clav751
						Chief Clav9080
	1	1	0	0	1	Appaloosa Clav9297
						Appler Clav7003
						Astro Clav9160

Germplasm : "Advance Clav3845"

Polymorphism	BCD385 EcoRI	Present	14.9	13.4	
		Absent	17.2	12.4	4.6
	BCD719 EcoRI	Present	6.6	4.4	
		Absent	14.5		
	BCD1150 EcoRI	Present	18.1	13.3	7.2
		Absent	15.0	9.6	5.8
	BCD1230 EcoRI	Present	5.8	3.3	1.8
		Absent			
	...				

A further example comes from **RiceGenes**, in which an RFLP survey is stored in Polymorphism and Germplasm records similar to those above and, in addition, each polymorphic band is represented as an Allele:

Allele : CDO99:HindIII-A

Locus CDO99	
MW_of_band	HindIII 9.9 IRRI/CU Labs
Polymorphism	CDO99/HindIII
Absent_in	Arikarai
Absent_in	Aswina
Absent_in	Babawee
...	
Present_in	ASD1
Present_in	Ai-Chiao-Hong
Present_in	Aichi Asahi
...	

Allele : CDO99:HindIII-B

Locus CDO99	
MW_of_band	HindIII 9.0 IRRI/CU Labs
Polymorphism	CDO99/HindIII
Absent_in	ASD1
Absent_in	Ai-Chiao-Hong
Absent_in	Aichi Asahi
...	
Present_in	Arikarai
Present_in	Aswina
Present_in	Babawee
...	

Yet other structures for representing polymorphisms can be found in Soybase [e.g. Probe "pA007", *In*: AGIS: Database: SoyBase. URL: http://probe.nalusda.gov:8300/cgi-bin/dbrun/soybase?find+probe+pa007) and TreeGenes (e.g.Germplasm "NC-6-13", *In*: AGIS: Database: TreeGenes. URL: http://probe.nalusda.gov:8300/cgi-bin/dbrun/ treegenes?find+Germplasm+nc-6-13]. In short, many structures are possible, with differences not only in their immediate appearance to the user but also in the kinds of queries they make possible. It is feasible to store the same data in more than one structure, but probably not in all possible structures. So thought should be given first to the kinds of queries and reports desired, and then the database should be structured to address those goals.

Names, names, names!

Of all the issues **GrainGenes** - or any database - must deal with, by far the biggest is the unambiguous identification of each data object. There are many examples of a single cultivar that is known by several synonyms, as well as a single name being given to entirely different cultivars (homonymy). To manage this problem, **GrainGenes** is moving toward using accession numbers for all germplasm records, combined with a common name if available: for example, "Advance CIav3845" instead of just "Advance". A cultivar also often has accession numbers in several different collections, and we are trying to cross-reference all these alternate names and accession numbers, as in the following record:

```
Germplasm : "ANZA"
    Other_name          D6413
                        D6923
                        II-8739-4R-1M-1R
                        II8739-4R-1M-1R-0USA
                        BW4820
                        WW15
                        KARAMU
                        Mexicani
                        T4
    Collection_and_ID   USDA/ARS/NSGC        CItr15284
                        CIMMYT               CID/SID: 6733/19
    Cross_number        II8739
    Abbreviation        ANZA
    ...
```

The same problem applies, to a lesser extent, to Probe records. Many of the known-function clones in particular are known under multiple published names, more than one of which is "Adh". Including the Genebank accession number is helpful in sorting out the ambiguities. Note that here again we are relying on other databases to perform the real disambiguation work, rather than assigning **GrainGenes** accession numbers.

A special problem with Probes is the distinction between a DNA clone and a PCR primer pair designed to mimic it, since in fact the hope is often not realized and different polymorphisms are observed. Currently **GrainGenes** is storing PCR primer data in the same record with the clone itself except where there is evidence of a difference in specificity.

Inter-database connections via World Wide Web

As mentioned above, one of the most exciting features of gatewaying a database to the Web is the opportunity to connect it to other Web-accessible databases. These connections are made directly from record to record. For example, when the Germplasm "Anza" record, shown above as it appears in the ACEDB database itself, is viewed through the Web, its Collection_and_ID field is modified to:

Collection_and_ID	USDA/ARS/NSGC	Cltr15284 [GRIN accession]
	CIMMYT	CID/SID: 6733/19 [CIMMYT]

Clicking on "[GRIN accession]" will look up the record for Anza in the *GRIN* database, running on Oracle software (in the same city, as it happens, but it could be anywhere in the world) and display it directly. Likewise clicking on "[CIMMYT]" will look up the corresponding record in the online version of the CIMMYT *IWIS* database. Similar connections are made from Probe records to Genbank, again via the accession number.

The gateway software being developed by the Genome Informatics Group at NAL is very flexible and powerful for creating these links, and the only limitation is the ability to match accession numbers in the foreign databases to the appropriate records of *GrainGenes*. Again, proper synonymy is the key. In the real world such synonymies really mean something between "probably the same" and "might be the same", due to uncertainties ranging from disagreements between data sources to genetic changes during regeneration of samples. Our new capabilities for molecular fingerprinting combined with a high degree of connectivity between databases can only improve our ability to detect and resolve discrepancies.

Acknowledgements

GrainGenes, the other plant genome databases mentioned, and the Genome Informatics Group are projects of the Plant Genome Program, conceived and directed by Jerome P. Miksche, and supported by the U.S. Department of Agriculture, the Agricultural Research Service and the National Genetic Resources Program.

Reference

Paul, E., M. Goto and S.D. Tanksley. 1994. SolGenes - a Solanaceae database. Euphytica 79:181-186.

Report of the Working Group on analysis, management and exchange of molecular data

1. As molecular data become available, they should be included as molecular descriptors in genebank databases.
2. The basic guidelines for the analysis, management and exchange of molecular marker data should be set in the near future. The premises of the current list of descriptors should be used as reference points for establishing the guidelines. A consultant group of specialists in molecular marker technology, analysis and data management should be formed to set up the general guidelines and act as a periodic advisory panel.
3. The acceptance and recommendation of particular molecular markers as descriptors should be determined by a committee that includes, for example, agronomists, molecular geneticists, population geneticists and crop specialists, as currently commissioned by IPGRI for morphological descriptors.
4. Molecular markers will be useful descriptors for genebanks only if they are reproducible and informative. Markers may or may not be genetically informative. It should remain possible to expand character states for each marker. Depending on the current knowledge of certain species, the description of the character should include or refer to genetic and phylogenetic information.
5. The ideas and discussions of the Human Genetic Diversity group should be taken into consideration as these could be very helpful. IPGRI should contact other groups and ongoing initiatives such as the Plant Genome Database and European Initiatives.
6. Data analysis should be used with an understanding of the assumptions and limitations of the techniques employed. More effort should be devoted to the development and dissemination of analytical methods for the analysis molecular marker data.
7. Information on experiences with long term studies and data gathering on isozymes for certain plant species should be compiled (for example, maize experience at NCSU and Costa Rican data on tropical forest trees, etc.) and made available as this might increase the understanding of the usefulness of these approaches.
8. Thorough description of populations and sampling methods should be included for each accession in genetic resources database.
9. IPGRI should provide basic information and cross-references through the Internet and other dissemination methods. IPGRI should also facilitate literature searches for users and provide copies of basic papers on the development and application of molecular markers.
10. It might be useful to establish reference data sets for the comparative testing of new techniques while helping to provide overall systematic information on a taxon.

11. A possible format for IPGRI descriptors for molecular data (see point 2 above) which can provide a basis for further discussion is set out below:

Descriptors for molecular polymorphisms

Accession Accession-1D

Polymorphism-study Polymorphism study-1D Present Text
Polymorphism-study Polymorphism study-1D Absent Text

Polymorphism-study Polymorphism-study-1D
Type Text
    Size-units Text
    Probe Text
    Enzyme Text
    Primers Text
    Buffer-system Text
    Protocol Text
    Band Text Reference_accession Accession 1-D Image Image-1D

Example
    Accession "Manihot #1"
    Polymorphism-study "ABC123 EcoRI CIAT1" Present "11.9 4.3"
    Polymorphism-study "ABC123 EcoRI CIAT1" Absent "9.8 3.7"
    Polymorphism-study "BCD456 HindIII CIAT1" Present "14.8"
    Polymorphism-study "BCD456 HindIII CIAT1" Absent "10.5 9.0 3.2"
    Polymorphism-study "ABC123 EcoRI CIAT1"
    Type "RFLP"
    Size-units "Kb"
    Probe "ABC123"
    Enzyme "EcoRI"
    Protocol "Agarose, 20V/cm, 8hr"
    Band "11.9" Reference-accession "Manihot #1" Image "ABC123 CIAT47"
    Band "4.3" Reference-accession "Manihot #1" Image "ABC123 CIAT47"
    Band "9.8" Reference-accession "Manihot #236" Image "ABC123 CIAT52"
    Band "3.7" Reference-accession "Manihot #17" Image "ABC123 CIAT30"

Increasing the use of plant genetic resources

Molecular techniques for increased use of genetic resources

Claire Lanaud and Vincent Lebot
CIRAD/BIOTROP, 34032 Montpellier, France

Introduction

The study of morphoagronomic variability is the classical way of assessing genetic diversity for plant breeders. For many species, especially minor crops, it is still the only approach used by breeders. However, with molecular markers techniques, powerful tools have been developed so that genetic resources can be accurately assessed and characterized.

Molecular markers are efficient not only because various and numerous points can be accessed along the genome but also because of their variable nature: single or repeated sequences, coding or non coding. It is, therefore, possible to study nuclear, chloroplast or mitochondrial DNA and to conduct different types of evolutionary studies such multiple and complementary information which subsequently enhance the value of genetic resources. Applications are numerous and we will present here general examples to illustrate them. These will be taken from tropical crop studies developed at CIRAD.

Monitoring plant material and assisting germplasm collection management

One major application of molecular markers is for monitoring plant material and assisting in the management of collections. One objective is to determine the breadth of the genetic base of germplasm collections used by breeders using molecular markers analysis. These markers are suitable for assessing how much allelic diversity is present in a crop and they have the potential for providing unique fingerprints for each genetically distinct genotype, a useful means of identifying different cultivars.

For perennials and trees, this identification is difficult and uncertain at a juvenile stage. If there is sufficient intraspecific polymorphism, isozymes might be powerful enough. This is the case, for example, for *Hevea*. In this case, 12 isozyme systems allow the differentiation of 95% of the cultivated clones. For this species, a "portable laboratory" has been developed in order to fingerprint accurately the clones used in industrial plantations (Laconte *et al.* 1994). The identification is possible at the nursery level where high yielding clones are propagated.

For other species, molecular markers corresponding to various DNA sequences will be necessary. Such molecular markers are used on a routine basis, for example in Australia, to verify controlled crosses between *Erianthus arundinaceus* and *Saccharum officinarum* (D'Honta *et al.* 1995). This enables breeders to exploit efficiently the related genus in sugarcane improvement.

Organizing genetic diversity

An assessment of genetic diversity based only on morphoagronomic traits might be biased because distinct morphotypes can result from a few mutations and share a common genetic background. This has been demonstrated for cocoa : distinct criollo morphotypes such as Pentagona or Porcelana are, in fact, genetically similar.

The opposite is also true, especially for sorghum. For this species, some cultivars are classified into races according to the panicle morphology. According to this, cultivars belonging to race guinea in West and South Africa were grouped together in the same morphotype. However, after isozyme studies (Ollitrault *et. al.* 1990), as well as with RFLPs (Deu *et al.* 1994), it appeared that a clear genetic differentiation existed between guinea cultivars originating from these two geographic zones. This has opened new prospects for the genetic improvement of this race in particular and for the use of genetic resources of the species in general.

Rapid characterization of germplasm

Early characterization of germplasm from collecting trips can be conducted using molecular markers. This is particularly interesting for perennials and allows the rapid identification of interesting genotypes in certain areas. This has been demonstrated so far in the laboratory for cocoa (Lanaud 1987) and *Hevea* spp. (Chevalier 1988).

Assisting in the construction of core collections

Thanks to the wide array of molecular tools that are now available, different but complementary approaches can be combined to study genetic diversity. This helps improve our knowledge of the hidden diversity of the species.

The aim is to organize germplasm collections so that the potential of these accessions are fully characterized and useful for breeders. It is also an approach used by curators to organize the diversity and variability existing within these collections. Molecular markers are essential for explaining whether existing genetic variability, which is assessed by measuring morphoagronomic traits, is related to genetic diversity, which is assessed by measuring allelic frequencies using molecular markers. This information can be used to construct core collections.

Assessing genetic distances to guide the use of genetic resources

Genetic relationships and distances between individuals are revealed by diversity studies and provide useful information for breeders; this can then be exploited to obtain the benefit from the potential of each population. For sorghum, crosses between groups Kafir, South African Guinea and Caudatum demonstrate a relationship between genetic distance and grain product (Chantereau *et al.* 1994). However, this relationship does not exist between progenitors from races belonging to Durra or West African Guinea and the first three above mentioned groups. Molecular markers have, therefore, provided a classification of sorghum cultivars which also allows better use to be made of the genetic resources of this species.

Tracing the origin of cultivated varieties

Most of the bananas consumed in the world today are triploid cultivars, sterile, seedless and parthenocarpic plants, which means that the fruits are formed without fertilization of the ovule. Their major characteristic is their thick flesh. In South East Asia, the area of origin of *Musa* spp., complex seeded and seedless diploids are widely distributed. At least four different species, the two major being *M. acuminata* and *M. balbisiana*, contribute to the genomic composition of diploid and triploid cultivars. They have a wide range of uses such as for desserts, in cooking and for making beer.

The improvement of triploids is very difficult and it is necessary to go through the diploid level to broaden the genetic base and to recombine the existing germplasm. Due to spatial and temporal isolation, *M.acuminata* has diverged into distinct subspecies which differ by one or more chromosomes rearrangements (inversions, translocations). Most of the diploid cultivars are sterile because they are, in fact, hybrids between subspecies. In order to improve our understanding of the genetic organization of triploids domesticated several hundred of years ago, the nuclear, chloroplast and mitochondrial DNA have been studied. At least one of these genomes permits the identification of different *Musa* species and the *M. acuminata* subspecies (Careel *et al.* 1994). Moreover, bananas are peculiar in that their chloroplast DNA is maternally inherited while the mitochondrial DNA is paternally inherited (Faure *et al.* 1993). Thanks to this trait, it has been possible to study phylogenies of cultivars and relationships between subspecies and diploid or triploid cultivars (Careel 1994). It appears that subspecies *banksii* and *errans* contribute to the starchy character of the fruits, while *malaccensis* contributes to the sweetness character of some cultivars. Furthermore, it has been shown that the *banksii* or *errans* subspecies are involved in almost all the diploid and triploid cultivars. It is very likely that parthenocarpy, which is the major character for domestication, appeared in both of these subspecies. This implies that the first area of domestication of cultivars might be located in the Papua New Guinea/Philippines region. Consequently, it is now possible to elaborate breeding strategies taking into account the specific characteristics of these subspecies and the role they have played during domestication.

Improving the use of wild species

Molecular markers also contribute to our understanding of the origins of polyploids. Numerous tropical crops are polyploids (bananas, sugarcane, *Citrus* spp.) and it is of great help for plant breeders to elucidate their genomic composition. This can be useful not only for taxonomic purposes (i.e.to differentiate between AAA and AAB banana cultivars) but also for helping to breed polyploids and to confirm ploidy levels. For some interspecific polyploids, different characters are contributed by different species and markers can trace the presence or absence of a genome exhibiting a peculiar trait.

Wild relatives of crops species are often a valuable source of resistance to diseases. However, their use often involves transferring deleterious agronomic characters that are also present in wild species. Thanks to molecular markers and to genomic maps, it is now possible to monitor introgression throughout breeding cycles and to target the areas of the genome possessing the desired characters. This approach is now being developed at CIRAD to assist sugarcane breeding programmes. Sugarcane varieties have a complex genomic structure. They are interspecific hybrids between *S. officinarum* and *S.spontaneum*. These hybrids have about one hundred chromosomes and less than 20 are contributed by *S. spontaneum* which is the wild species that contributes vigour,

resistance to diseases and tolerance of environmental stress. The objectives of breeding are to preserve the useful agronomic traits from *S. spontaneum,* such as vigour and resistances. This is now being done by crossing varieties while limiting the number of chromosomes of *S. spontaneum* transferred to those which contribute useful characters.

Furthermore, hybrids between *S. officinarum* and the related genus *Erianthus* are used in breeding programmes. The complex genome of sugarcane creates difficulties for breeders and studies on the genetics of specific characters are almost impossible. Molecular markers and *in situ* hybridization are valuable tools for understanding and monitoring breeding cycles. A genetic map composed of 428 markers, has recently been developed by working first on *S. spontaneum* chromosomes where specific markers have been identified (Grivet 1995). A search for QTLs is now being conducted in order to identify chromosomes from this species that could contribute the targeted characters.

In situ hybridization is of particular interest here for understanding and controlling the structure of genome variation: the aim is to identify accurately chromosomes from each species or genus that compose hybrids. It is possible to hybridize the DNA from each parental species directly onto interspecific or intergeneric hybrid metaphases chromosome preparations. This DNA will be detected by different fluorochromes. Thanks to a higher specificity of hybridizing on its original species, chromosomes or chromosome parts from each species in the hybrid will be identified and revealed by the fluorochromes.

The use of this method enables *Erianthus* chromosomes in hybrids (D'Honta *et al.* 1995) to be identified easily. Furthermore, it is possible to locate and count the number of *S. spontaneum* chromosomes existing in each variety and to reveal rare recombinations occurring between genomes of the two species. These techniques offer new opportunities for a better understanding of gene transfers that are essential.

Identifying different sources of interesting genes in the genome

For diploid species, the interest in Marker Assisted Selection is growing and genome mapping allows us to understand the genetic control of characters and to identify favourable or unfavourable genetic linkages. This improves our knowledge of possible introgression of a particular character and the stability of its expression as well as enabling us to control recombination. The variability of the genes directly involved is also revealed.

This approach is now being developed to help improve resistance to *Phytophthora* in cocoa breeding. One of the objectives is to produce progenitors that are more resistant using the genetic resources of this species. This is because well identified resistant progenitors are found in several genetic groups. For this species, a complete map has already been produced (Lanaud *et al.* 1996) and studies currently being carried out on several progenies will allow elucidation of the nature of resistance and an understanding of the diversity of genes involved in this resistance. This will allow us to target our choices of progenitors to be used in breeding in order to capture resistant genes in improved genotypes. Moreover, the use of molecular markers at the juvenile stage will hasten breeding cycles for this perennial crop.

Which picture of the genetic diversity is revealed with which marker?

The currently available tools do not work for all species and in all laboratories. Ideal markers that are codominant, numerous, easy to use and non-specific have yet to be

identified. For major crops and well studied species of economic importance, numerous markers will be developed and RFLP in particular will provide tremendous genetic information. However, for minor crops and under-exploited species, or for small research teams that cannot use RFLPs, non-specific markers such as isozymes or RAPDs can be developed. It is useful to clarify which picture of the existing genetic diversity will be revealed by which marker.

Comparisons have already been made between isozymes and RFLPs for sorghum (Ollitraut *et al.* 1990; Deu *et al.* 1994), and the same geographic groupings have been revealed by both types of markers. RFLPs however, revealed different races of sorghum that could not be identified with isozymes. In the case of rice, a similar classification of rice varieties has been obtained with RFLPs (Wang and Tanksley 1989) and isozymes (Glaszmann 1987).

Comparisons have also been made between RFLP and RAPDs for cocoa and there, too, a very similar structure of the genetic diversity has been revealed (N'Goran 1994). However, some differences appear for certain individuals.

Isozymes and RAPDs are easier to use than RFLPs and appear to be efficient tools for revealing the overall genetic diversity structure within a species. When working on an under-exploited species, a team might be interested in combining several techniques. For example, if several hundred accessions have been collected to study the ecogeographical variation existing within a given species, it would seem appropriate for all the accessions to be screened first for isozymes. Isozymes will reveal global groupings and the variability existing within groups can subsequently be revealed with more powerful markers such as RFLPs. This cascade approach is also more cost-efficient and useful when working on the identification of a core collection.

Furthermore, because isozymes are codominant markers, it is possible to conduct population genetic studies and to reveal the extent of genetic diversity existing within and between populations, the levels of heterozygosity, mating systems, rates of outcrossing and levels of inbreeding depression. They also allow one to monitor legitimacy of controlled crosses as has been done in cocoa to control production of hybrid seeds (Lanaud *et al.* 1987).

However, codominant molecular markers can allow more detailed studies of the genetic structure of the existing diversity and provide a great wealth of information. Their use is sometimes limited because of the absence of probes. Heterologous probes, corresponding to well preserved genes common to different species, can provide solutions to this problem. Since related species share conserved parts of their genomes, it was possible to map sorghum and sugarcane genomes using maize probes (D'Hont *et al.* 1994). Consequently, we can compare their genomes and reveal similarities between maps (Grivet *et al.* 1994). This is also the case for species that are less related. For example, recent studies have been conducted on coconut using rice probes to study coconut genetic diversity (Lebrun, pers. com.).

Conclusions

Molecular markers have a wide array of applications from the simple control of germplasm to interesting gene identification. For poorly improved species, or those with complex genetic structure, molecular markers have helped remove constraints faced by traditional breeding while enhancing the use of genetic resources. However, molecular marker techniques need to be simplified in order to allow a greater number of species, including less well known or poorly improved species, to be studied using these tools.

References

Carreel, F. 1994. Etude de la diversité génétique des bananiers (genre *Musa*) à l'aide des marqueurs RFLP. Thèse I.N.A. Paris-Grignon.

Carreel, F., S. Faure, D. Gonzalez-de-Leon, P.J.L. Lagoda, X. Perrier, F. Bakry, A. Tezenas-du-Montcel, C. Lanaud and J.P. Horry. 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp.). Gen. Sel. Evo. 26 (suppl. 1):125-136.

Chantereau, J., M. Deu, J.C. Glaszmann, I. Degremnot, D. Gonzalez de Leon and P. Hamon. 1994. RFLP diversity in sorghum in relation to racial differentiation and heterosis in hybrids. Pp. 38-45 *in* Use of Molecular Markers in Sorghum and Pearl Millet Breeding in Developing Countries (J.R. Witcombe and R.R. Duncan, eds). ODA, London, UK.

Chevellier, M.H. 1988. Genetic variability of *Hevea brasiliensis* germplasm, using isozyme markers. J. Nat. Rubber Res. 3:42-53.

DíHont, A., Y.H. Lu, D. Gonzalez-de-Leon, L. Girvet, P. Feldman, C. Lanaud and J.C. Glaszmann. 1994. A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. Genome 37:222-230.

Deu, M., D. Gonzalez-de-Leon, J.C. Glaszmann, I. Degremont, J. Chantereau, C. Lanaud and P. Hamon. 1994. RFLP diversity in cultivated sorghum in relation to racial differentiation. Theor. Appl. Genet. 88:838-844.

D'Hont, A., P.S. Rao, P. Feldmann, L. Grivet, N. Islam Faridi, P. Taylor and J.C. Glaszmann. 1995. Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* x *Erianthus arundinaceus*, with molecular markers and DNA *in situ* hybridisation. Theor. Appl. Genet. (in press).

Faure, S., J.L. Noyer, F. Carreel, J.P. Horry, F. Bakry and C. Lanaud. 1993. Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Curr. Genet. 25:265-269.

Glaszmann, J.C., 1987. Isozymes and classification of Asian native rice varieties. Theor. Appl. Genet. 74:21-30.

Grivet, L. 1995. Marquage moléculaire chez la canne à sucre (Saccharum spp.) ; décomposition d'une structure génétique complexe et application à l'amélioration variétale. Université de Paris Sud Centre d'Orsay. Thèse du 4 Avril 1995.

Grivet, L., A. DíHont, P. Dufour, P. Hamon, D. Roques and J.C. Glaszmann. 1994. Comparative mapping of sugarcane with other species within the tribe Andropogoneae. Heredity 73:500-508.

Lanaud, C. 1987. Nouvelles données sur la biologie du cacaoyer (*T. cacao* L.): Diversité des populations, système d'incompatibilité, haploïdes spontanés. Leurs conséquences sur l'amélioration génétique de cette espèce. Université de Paris XI, Centre d'Orsay, Doctorat d'Etat.

Lanaud, C., A.M. Risterucci, A.K.J. NíGoran, D. Clement, M.H. Flament, V. Laurent and M. Falque. 1996. A genetic linkage map of *Theobroma cacao* L. Theor. Appl. Genet. (in press).

Lanaud, C., O. Sounigo, Y.K. Amefia, D. Paulin, Ph. Lauchenaud and D. Clement. 1987. Nouvelles données sur le fonctionnement du système d'incompatibilité du cacaoyer et ses conséquences pour la sélection. Café, Cacao, Thé XXXI(4):267-282.

Leconte, A., P. Lebrun, D. Nicolas and M. Seguin. 1994. Electrophorèse: application à l'identification clonale de l'hévéa [Electrophoresis application to Hevea clone identification]. Plantations, Recherche, Développement 1:28-36.

NíGoran, J.A.K., V. Laurent, A.M. Risterucci and C. Lanaud. 1994. Comparative genetic diversity studies of *Theobroma cacao* L. using RLFP and RAPD markers. Heredity 73:589-597.

Ollitrault, P., M. Arnaud and J. Chantereau. 1990. Polymorphisme enzymatique des sorghos. II. Organisation génétique des sorghos cultivés. L'agron. Trop. 44:211-222.

Wang, Z.Y. and S.D. Tanksley. 1989. Restriction fragment length polymorphism in *Oryza sativa* L. Genome 12:1113-1118.

Molecular genetic techniques in relation to sampling strategies and the development of core collections

M. Bonierbale, S. Beebe, J. Tohme and P. Jones
CIAT, Cali, Colombia

Introduction

Most experience in plant genetic resources supports the fact that genetic diversity is not uniformly distributed across geographic, political or taxonomic boundaries. Yet international efforts in germplasm conservation must often use these indicators as operational and organizational features of large collections. For both wild and cultivated germplasm, factors such as breeding system, community composition or cultural system, and habitat disturbance or other selection pressures bear significantly on the distribution of genetic variability. Within its general range of geographic or taxonomic distribution, genetic diversity also has internal structure at the community, species, population and even the genotype level. The conservation and use of genetic diversity depends on understanding this organization and, ideally, the processes that influence it. Organization of the cassava and bean collection at CIAT have taken advantage of such parameters as ecogeographic distribution, and morphological, biochemical and, more recently, molecular indicators of genetic diversity in attempts to assemble representative samples in germplasm banks, which serve the dual purpose of conservation and improved access to genetic resources. Progress in the formation of core collections, development and application of molecular tools and constraints to accomplishing these goals are discussed in this paper.

Selection of core collections

From the commodity or crop improvement perspective, a specific requirement for the conservation and utilization of diverse genetic resources is an assessment of the amount of variation present within various components of the genepool, including traditional varieties, improved lines and wild relatives. The evaluation of carefully assembled germplasm collections facilitate this determination, but the task is formidable. Core collections, or representative samples of larger genebanks, provide a means for application of more expensive characterization activities on representative subsets than would be practical on large collections. This process serves to orient use of the germplasm and may also reveal patterns of distribution of variability which would impact on the formation of the collection, giving both core and reserve collections a dynamic nature. While stability in a core collection permits the comparison of data on a common set of genotypes over sites and years, total stability is not to be expected. At the outset, any core collection will suffer imperfections. Internal duplication may occur that can be eliminated later, or variability that should be included in the core may subsequently be discovered or recognized. The constitution of core collections should probably be reviewed periodically to attend to these adjustments.

With regard to the use of molecular markers in the selection of core collections, these probably will not play a primary role in the initial selection. This is due to the very nature of the initial selection, that of reducing a very large number of accessions to a manageable number for more intensive study. Due to their cost, molecular markers cannot yet be applied to most large collections, but their application to core collections

may be feasible. Therefore molecular markers may be very useful in the exercise of adjusting the core for the elimination of duplicates, or for the inclusion of variability that was absent in the original core.

Bean core

The bean (*Phaseolus vulgaris*) core collection was selected from a collection of approximately 23000 accessions of common bean in CIAT's genebank. A combination of three types of criteria was utilized: 1) knowledge about the historical development of the crop, to emphasize traditional regions of cultivation in primary centres; 2) the agroecological origin of the accessions, with respect to soil characteristics, rainfall, length of growth cycle including the effect of temperature and day length at flowering period; and 3) plant and seed morphology, placing emphasis on non-commercial seed types and indeterminate growth habits. Sampling was performed to represent the two principal genepools of cultivated common bean.

Cassava core

A core collection of cassava (*Manihot esculenta*), consisting of 630 genotypes, was assembled at CIAT in the early 1990s to represent the genetic diversity of the base collection of 5263 to represent the available genetic variability in a more manageable size from the point of access. It was defined through the use of three parameters - geographic origin, morphology and diversity of isozyme patterns. Selection by geographic origin considered the historical development of the crop, weighting more heavily its centre of diversity and areas of traditional cultivation and distinct ecology. Morphology is described by the application of a relatively stable discriminatory set of 21 non botanical descriptors. Isozyme data was included in an effort to achieve a good representation of alleles present in the base collection, on the assumption that biochemical markers are neutral characters with respect to selection and environmental effects. The base collection was stratified by the three parameters of diversity, and random samples were drawn from the resulting groups. The group was then complemented by *a priori* selections of genotypes with particular agronomic importance or ecological adaptation. The collection is expected to be dynamic as a consequence of the elimination of duplicates, acquisition of new materials, and better characterization of the genetic diversity of the genepool, including wild relatives.

Validation and use of core collection

Two exercises were carried out to verify the representativeness of the bean core collection. First, among accession derived from Peru, isoenzymes were assayed to compare those captured in the core collection with those in the reserve. In 100 core accessions, 95.7% of the polymorphic forms were recovered from 382 accessions in the reserve. Secondly, an analysis of DNA utilizing RAPDs was employed to compare two samples of 90 Mexican accessions each, one of the core, and one of the reserve collection. No significant difference could be detected between the two samples by any criteria tested.

RAPDs have been used to study the structure of the cultivated Mesoamerican bean germplasm. Molecular analysis revealed results which complemented previous concepts of the racial structure of *P. vulgaris*. Unique variation was found in germplasm from the highlands of Guatemala and the state of Chiapas in Mexico which was recognized to be unique based on plant morphology. Future revision of the core collection might consider whether such germplasm is adequately represented in the core.

Analysis by AFLPs was applied to the core of wild *P. vulgaris* with great success, providing more details on the genetic structure than was previously possible. Besides

the two genepools widely reported for wild beans, additional groups were recognized in Northern Peru and Colombia. Additionally, the internal structure of wild bean genepools could be discerned that revealed evolutionary tendencies within and among groups. In the Andean zone in particular, groups appear to have been isolated and to have followed independent evolutionary paths. AFLPs may have special potential for genetic studies, because large amounts of data can be generated quickly.

The cassava core collection has been evaluated across different production ecosystems and years using a series of representative testing sites in Colombia. This permits the determination of genetic variability and genotype by environment effects for important traits which are difficult to measure on the whole collection, and suggests particular sections of the collection which offer specific useful genetic variability.

The core collection of cassava is presently being used for a quantitative assessment of relative variability in cultivated versus wild germplasm, by comparing RFLP patterns in subgroups of the collections at CIAT. A third group of material in this study is elite germplasm, included to give a first estimate of the effect of selection of successful genotypes in specific environments on overall genetic diversity.

Elimination of duplicates in germplasm collections

An important consideration in the management of any germplasm collection is the minimization of duplicate accessions, whose maintenance consumes valuable resources with no return. The strategy currently being applied at CIAT to the identification of duplicates in the cassava collection relies on the database of characterization information, and consists of three steps: 1) first, candidate duplicates are identified by screening the data base for accessions with identical characterization data, including morphological descriptors and isozyme patterns; 2) as the morphological data often results from activities conducted on distinct parts of the collection in different years, it is necessary to grow the putative duplicates together side-by-side in the same year, and repeat the characterization process; and 3) finally, when the number of putative duplicates is reduced, the remaining groups are characterized with a molecular probe, M13 minisatellite, which assays many regions of the genome at the same time, presenting a highly discriminatory fingerprint for each genotype.

Accessions which are identical by these three criteria are eliminated from the field collection, keeping only one representative, but maintained at least temporarily in the *in vitro* gene bank. According to the level of duplication so far detected, it is expected that around 600 accessions could be eliminated from the collection, saving considerable resources in germplasm conservation. The current fingerprinting procedure relies on radioactive labelling of a DNA probe, which can be prepared and applied at CIAT. Current research in the development of additional markers for cassava, and non-radioactive methods of marker detection, will facilitate the application of this model to other institutes, and reduce the cost. One marker type under consideration for this purpose is AFLP.

Development of tools for sampling the plant genome

Nuclear RFLPs derived genomic and cDNA libraries of cassava and RAPDs have been used at CIAT to select a polymorphic population for the development of a molecular genetic map of cassava. Evaluation of a set of cassava genotypes showing diversity of morphology and origin were found, in a survey with several libraries and restriction enzymes, to have a rate of polymorphism of approximately 10%. This rate was

subsequently enhanced by the selection of a most informative set of enzymes and genotypes for future study.

RAPD markers have been dev cDNA libraries of cassava and RAPDs have been used at CIAT to select a polymorphic population for the development of a molecular genetic map of cassava. Evaluation of a set of cassava genotypes showing diversity of morphology and origin were found, in a survey with several libraries and restriction enzymes, to have a rate of polymorphism of approximately 10%. This rate was subsequently enhanced by the selection of a most informative set of enzymes and genotypes for future study.

RAPD markers have been developed for cassava by applying 10-nucleotide random primers to three interspecific crosses of cassava, selecting primers which produced polymorphic amplification products that were inherited as single dose markers. This work resulted in the recommendation of primers detecting clear polymorphisms for mapping and genetic diversity analysis.

Microsatellites have been developed recently in cassava, through collaboration with the University of Georgia, and appear to be more highly polymorphic than RFLPs. Those developed to date are mostly simple repeats of base-pair elements and their evaluation under hot and cold PCR is being compared for cost advantages and resolution in germplasm characterization and mapping. These markers will only become practical for cassava if existing methods such as multiplexing and non-radioactive detection of high resolution products are applied.

An F_1 hybrid population of 90 individuals from a cross between two cassava varieties with complementary agronomic properties was identified as the most highly polymorphic among the three intraspecific crosses compared. Linkage analysis has been conducted among approximately 150 RFLP markers segregating from the heterozygous male parent, to produce a framework genetic map. This framework is being used as the basis for combining information from other types of markers which segregate in the same population. For example, it was possible to localize one of the isozyme loci from the ab esterase system which has been used to characterize the international cassava collection at CIAT to a linkage group with 6 RFLP markers. One microsatellite marker was similarly mapped in the progeny after resolution of the cold PCR products on polyacrylamide gels.

The molecular genetic linkage map of cassava provides genomic distribution information on the markers it is comprised of, improving their value for germplasm studies. It is also highly probable that the markers will detect linkage to desirable traits for cassava improvement.

Ecogeographic information applied to germplasm sampling strategies

Great potential exists for applying Geographic Information Systems (GIS) to understanding biodiversity. GIS has been applied at CIAT to determine homology among crop production environments, to search for and deploy natural enemies of crop plants, and to explore the distribution and potential use of genetic diversity in wild relatives. Ecological parameters including latitude, longitude and elevation taken from passport data, and temperature, diurnal variation in temperature, and rainfall interpolated from the CIAT climate database have been determined for the collection sites of over 1000 wild *P. vulgaris* accessions held in the germplasm collection at CIAT. This information was used to develop a prediction model for the probability of distribution of wild species in regions where there has been no previous germplasm collecting. This manner of determining homologous sites is in the process of verification for collecting orientation through

exploration in specific areas such as the eastern cordillera of Colombia and the Venezuelan Andes.

The same model of site prediction may serve in efforts to develop a conservation strategy for germplasm that is not suited to long-term, inactive storage, as is the case for the wild relatives of cassava. Due to less consistent exploration and collecting, only a limited amount of passport data is available for manihot species. Collection site data is being complemented with herbarium data to determine the ecogeographic range of selected species, for application to the same site prediction model. For the species of Mexico, this georeferenced information has been mapped and overlaid on the FAO soils map at 1:5 000 000. It is expected that the output will indicate undocumented regions with high probability of hosting manihot species, such as within the Colombian Amazons which should be explored and sampled for germplasm conservation and evaluation. Furthermore, a complete description of ecology of origin will suggest particular germplasm as possible sources of desirable adaptive traits.

The role of molecular markers

As a complement to morphological, biochemical, ecological and genetic information, molecular markers can contribute greatly to the use of genetic diversity through the descriptive information they provide on the structure of gene pools and genotypes.

A caveat is warranted with respect to the use of molecular markers. While they are unquestionably one of the most powerful tools for evaluating diversity, they are still basically indicators of potentially useful variability. Even several hundred markers scattered over the genome constitute a very small segment of the total genome. The analysis of such markers probably reveals patterns of broad evolutionary trends or lines of evolution. These patterns probably reflect random mutations accumulated over millennia of descent from distant ancestors, and may not necessarily be related to the occurrence of genes which control traits of economic or agronomic utility. The occurrence of useful genes may respond more to external selection pressures than the markers utilized to study genetic structure. Their distribution, therefore, could be more related to the distribution of selection pressures than to the line of descent from an ancestral stock. It remains to be determined to what degree the distribution of these genes is related to those markers that reflect evolutionary trends.

Constraints

The sheer scale of the task of evaluating genetic diversity, or applying molecular markers to practical breeding objectives, is a good incentive for developing low cost characterization methods. Safety and availability of reagents in diverse research institutes is also an issue that motivates the development of methodology and the choice of analytical tools. In the case of understudied species, such as manihot, limited past exploration and scarce availability of germplasm in collections impedes efforts to evaluate genetic diversity. This question of access is one of the main incentives for assembling collections and should be taken into consideration when setting research priorities. Given adequate resources, an integrated approach to exploration, conservation and evaluation with molecular, geographic and other tools could efficiently remedy this situation for neglected species. An additional less technical issue concerns policy and/or setting priorities, in which, for example, native plant species benefit society at large, implying an international role in conservation orientation which also involves sovereignty.

The Japan Rice Genome Project: enhanced use of genetic resources

Takuji Sasaki
National Institute of Agrobiological Resources, Tsukuba, Ibaraki 305, Japan

Introduction

The Rice Genome Research Programme (RGP) started in 1991 as a collaborative research programme between NIAR (National Institute of Agrobiological Resources) and STAFF (Society of Techno-innovation for Agriculture, Forestry and Fisheries). The programme will last for seven years and, during this period, it aims primarily to make a catalogue of expressed rice genes, to complete a fine resolution RFLP (restriction fragment length polymorphism) linkage map and to make a contig of YAC (Yeast artificial chromosome) clones using DNA markers on the linkage map.

Rice is the most important staple food not only in Japan, but also in Asia and Africa. Improvement of rice for its productivity in any environment and against serious diseases has so far been accomplished by breeding based on crossing, and the empirical knowledge of breeders. Recent progress in biotechnology enables new rice varieties to be developed, with new characteristics previously thought to be difficult to obtain. To use this new technology effectively, a segregating population with regard to the targeted trait has to be established and the DNA markers closely linked to it found. A RFLP linkage map with markers of an average distance of ca. 1 cM is desirable for accurate tagging of major genes. Even quantitative traits controlled by several loci, such as culm length and heading time, can be analyzed precisely using a fine resolution RFLP map. Further, when the YAC clone hybridizing with DNA markers which are closely linked to a specific trait is identified, the gene responsible for this trait can be isolated by map-based cloning. The characterization of a gene as DNA is the first step towards producing a new variety of rice using biotechnology.

cDNA cataloguing

The number of genes in rice is estimated to be around 30 000. Some of them are expressed in various tissues depending on the developmental stage or on environmental conditions. To capture all genes, cDNA libraries containing the artificial transcripts of mRNA were made from various tissues: young root, green shoot, etiolated shoot, flowering panicle, ripening panicle and callus cultured in the presence of 2,4-dichlorophenoxyacetic acid, benzyladenine, or gibberellin or cultured by changing the temperature once and so on. Random cloning of cDNAs and partial sequencing of them enables us to capture rapidly characteristic genes expressed in the rice tissues described above. Similarity search, based on translated amino acid sequences against a public database, PIR (Protein Information Resources), gives information about the function of gene product of the analyzed cDNA clone. If the score calculated using the FASTA algorithm to judge the similarity between rice and the targeted amino acid sequence gives a significant value, the function of the cDNA can be putatively identified.

We have sequenced about 25 000 cDNA clones from the cDNA libraries described above and about 25% of them were putatively identified (Saski *et al.* 1994, Havukkala *et al.* 1995). Our strategy did not adopt a sequencing from both ends of the clone and this makes it difficult to estimate the identity of each sequence. However, providing an overlapping of 50 nucleotides with more than 90% matching or that of 30 nucleotides

with perfect matching, about one half of analyzed cDNA clones was judged to be independent. Although this estimation is invalid for non-overlapping sequences, we seem to have captured 40% of all rice genes.

Of the analyzed sequences, 10,990 have already been submitted to the DDBJ/GenBank/ EMBL database. The list of accession numbers is available in our newsletter, Rice Genome, which is also published on the INTERNET via WWW (http://www.staff.or.jp) (Sasaki *et al.* 1994; Havukkala *et al.* 1995).

Genetic mapping

To tag phenotypic traits precisely, a genetic map with many RFLP markers is required. Mainly using cDNAs as probes for Southern hybridization, DNAs (2 mg) from 186 F2 plants produced by crossing a *japonica* cultivar, Nipponbare, and an *indica* cultivar, Kasalath, were digested by eight kinds of restriction enzymes; *Bam* HI, *Bgl* II, *Eco* RV, *Hin* dIII, *Apa* I, *Dra* I, *Eco* RI or *Kpn* I. A non-radioactive detection system, enhanced chemiluminescence (ECL, Amersham), was used for getting hybridizing signals and this system enabled us to use a filter for more than 30 kinds of probes repeatedly. Published in 1994, our map published in 1994 contained 1374 DNA markers at an average interval of 300 kb. The markers comprised 876 cDNAs, 263 genomic DNAs, 147 random amplified polymorphic DNAs (RAPD) and 88 other DNAs (Inoue *et al.* 1994; Kurata *et al.* 1994: Monna *et al.* 1994). The parental Southern hybridization patterns are available on the INTERNET via WWW and these markers are available as DNAs through the DNA bank of the National Institute of Agrobiological Resources.

RAPDs (random amplified polymorphic DNAs) were detected by PCR using one or two kinds of 10-mer of nucleotides of random sequence as primers. This is a convenient and rapid method for detecting polymorphism among many DNA samples. However, one problem is the unstable reappearance of amplified DNAs which show polymorphisms. To overcome this problem, it is desirable to clone polymorphic DNA and to establish PCR primers which can reproducibly give polymorphisms corresponding to the specified RAPD. A site specified by PCR is called STS (sequence tagged site). We have so far established 168 sets of novel PCR primers for detecting STSs (Fukuoka *et al.* 1994, Monna *et al.* 1995 and Nagamura *et al.* 1995).

RGP has by now added more than 500 DNA markers to the published map, though the updated map has not yet been released. To tag a targeted trait by DNA markers, first of all, a segregating population for the trait needs to be established and, secondly, recombination with DNA markers being located near the trait locus has to be checked. To use DNA markers as a tool for rapid screening of trait carriers, the markers close to the trait locus are desirable (Lin *et al.* 1994). If the nucleotide sequence of RFLP marker can produce PCR primers that give a corresponding polymorphic product, they are the most powerful tools for accelerated breeding.

Quantitative traits, which are strongly related to rice productivity, can also be accurately determined for their loci using a high density genetic map. Each locus can be separated by repeated backcrossing and by checking graphical genotypes of selected progenies. Once a locus which mainly contributes to a quantitative trait is isolated and is proved to be inherited in a Mendelian manner, the judgement of presence or absence of such a locus might be facilitated by DNA markers.

Occasionally, the frequency of polymorphism dramatically decreases among usual cultivars, because the preference to rice taste or adaptability of rice to environment in the specified area has made cultivars monomorphic as a result of inbreeding with closely related cultivars. To overcome this problem, DNA markers should be shared

among researchers to increase the possibility to detect polymorphism. RGP is now collaborating with researchers in Cornell University, who have also made another linkage map, to correlate the loci of common markers in both maps for their more effective use (Fukuoka *et al.* 1994; Inoue *et al.* 1994; Kurata *et al.* 1994; Lin *et al.* 1994; Monna *et al.* 1994; Monna *et al.* 1995; and Nagamura *et al.* 1995).

Physical mapping

In order to be able to locate genes exactly and isolate them, a physical map has to be constructed or a contiguous alignment (contig) made using large-sized genomic DNA fragments, such as YAC clones. Our YAC library comprises 7000 clones with an average insert size of 350 kb (Umehara *et al.* 1995). This means that the total length of the insert covers 5.5 times of rice genome. Screening of this library with DNA markers on a genetic map can select YAC clone(s) involving probe sequences. If one DNA marker hybridizes with more than two YACs and they overlap to cover regions that hybridizes to a neighbouring DNA marker, a YAC contig bridging these two markers can be obtained. If no YAC is found to make a contig, other methods, such as fingerprinting or screening of other types of genomic DNA libraries are to be used to fill gaps.

Constructing a physical map is very laborious. However, the result obtained from this work is indispensable for physically isolating and characterizing genes. Phenotypic traits tagged by our DNA markers might be located in our YAC clone. However, our YAC library was constructed from only one *japonica* cultivar, Nipponbare. To isolate a gene responsible for a specific trait which resides in another variety, we should make genomic libraries such as cosmid library and a cDNA library carrying that gene in an expressed form from this cultivar. It takes a fairly long time to isolate a gene responsible for a phenotypic trait this way (Umehara *et al* 1995).

Synteny analysis

Mapping of wheat DNA markers to rice map and vice versa revealed that the ordering of DNA markers was conserved (Ahn and Tanskley 1993). For example, the rice chromosome 6 shared strong colinearity of gene ordering with wheat chromosome 7(H). Also between rice and barley, maize and foxtail millet, genomic regions in common gene ordering are observed. Because the genome size of rice is the smallest among grasses, gene tagging in crops other than rice is facilitated by using rice YAC clones (Kurata *et al.* 1994, Dunford *et al.* 1995). The idea that, if gene ordering is conserved, similar genes responsible for a trait observed in, for example, wheat should also exist in rice, supports the utilization of rice as a model crop among cereals (Ahn and Tanksley 1993; Kurata *et al.* 1994; Dunford *et al.* 1995 and Kilian et al. 1995).

Conclusion

RGP has produced a lot of valuable results which help understand the structure of the rice genome and indicate possible applications for improving all cereal crops. However, our genome research is just beginning and greater advances in basic and applied research are necessary. To thoroughly unravel the rice genome, that is, to complete a physical map, to characterize genes corresponding to phenotypic traits and to sequence a wide range of interesting genomic regions, much more work is required. In order To

apply our linkage map to breeding, careful work is needed to prepare a segregating population and for further technical development in treating many DNA samples simultaneously within a short period of time is also required. Improving rice by introducing a desirable gene obtained by genome research into a cultivar to be transformed is still a dream. We basic researchers should make efforts step by step to overcome many difficulties and problems for improving rice to supply sufficient food to peoples living in a limited environment.

References

Dunford, R.P., N. Kurata, D.A. Laurie, T.A. Money, Y. Minobe and G. Moore. 1995. Conservation of fine-scale DNA marker order in the genomes of rice and the Triticeae. Nucleic Acids Res. 23:2724-2728.

Fukuoka, S., T. Inoue, A. Miyao, L. Monna, H.S. Zhong, T. Sasaki and Y. Minobe. 1994. Mapping of sequence-tagged sites in rice by single strand conformation polymorphism. DNA Research 1:271-277.

Havukkala, I., H. Ichimura, Y. Nagamura and T. Sasaki. 1995. Plant-Wide Names for Partial cDNA Sequences of Rice Using Blast and 4th Dimension. Plant Molec. Biol. Reporter 13:38-55.

Inoue, T., H.S. Zhong, A. Miyao, I. Ashikawa, L. Monna, S. Fukuoka, N. Miyadera, Y. Nagamura, N. Kurata, T. Sasaki and Y. Minobe. 1994. Sequence-tagged sites (STSs) as standard landmarkers in rice genome. Theor. Appl. Genet. 89:728-734.

Kilian, A., D.A. Kudrna, A. Kleinhofs, M. Yano, N. Kurata, B. Steffenson and T. Sasaki. 1995. Rice-barley synteny and its application to saturation mapping of the barley Rpg1 region. Nucleic Acids Res. 23:2729-2733.

Kurata, N., G. Moore, Y. Nagamura, T. Foote, M. Yano, Y. Minobe and M. Gale. 1994. Conservation of genome structure between rice and wheat. Bio/technology 12:276-278.

Kurata, N., Y. Nagamura, K. Yamamoto, Y. Harushima, N. Sue, J. Wu, B.A. Antonio, A. Shomura, T. Shimizu, S-Y. Lin, T. Inoue, A. Fukuda, T. Shimano, Y. Kubok, T. Toyama, Y. Miyamoto, T. Krihara, K. Hayasaka, A. Myao, L. Monna, H.S. Zhong, Y. Tamura, Z-X. Wang, T. Monna, Y. Umehara, M. Yano, T. Sasaki and Y. Minobe. 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genetics 8:365-372.

Lin, S.-Y., Y. Nagamura, N. Kurata, M. Yano, Y. Minobe and T. Sasaki. 1994. DNA markers tightly linked to genes, *Ph*, *alk* and *Rc*. Rice Genet. Newsl. 11:108-109.

Monna, L., A. Miyao, H.S. Zhong, T. Sasaki and Y. Minobe. 1995. Screening of RAPD markers linked to the photoperiod-sensitivity gene in rice chromosome 6 using bulked sergeant analysis. DNA Research 2:101-106.

Monna, L., A. Miyao, T. Inoue, S. Fukuoka, M. Yamazaki, H.S. Zhong, T. Sasaki and Y. Minobe. 1994. Determination of RAPD markers in rice and their conversion into sequence tagged sites (STSs) and STS-specific primers. DNA Research 1:139-148.

Nagamura, Y., T. Inoue, B.A. Antonio, T. Shimano, H. Kajiya, A. Shomura, S.Y. Lin, Y. Kubiki, Y. Harushima, N. Kurata, Y. Minobe, M. Yano and T. Sasaki. 1995. Conservation of duplicated segments between rice chromosome 11 and 12. Breed. Sci. 45:373-376.

Sasaki, T., J. Song, Y. Koga-Ban, E. Matsui, F. Fang, H. Higo, H. Nagasaki, M. Hori, M. Miya, E. Murayama-Kayano, T. Takiguchi, A. Takasuga, T. Niki, K. Ishimaru, H. Ikeda, Y. Yamamoto, Y. Mukai, I. Ohta, N. Miyadera, I. Havukkala and Y. Minobe. 1994. Toward cataloguing all rice genes: large scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J. 6:615-624.

Umehara, Y., A. Inagaki, H. Tanoue, Y. Yasukochi, Y. Nagamura, S. Saji, Y. Otsuki, T. Fujimura, N. Kurata and Y. Minobe. 1995. Construction and characterization of a rice YAC library for physical mapping. Molec. Breed. 1:79-89.

Molecular markers for characterization and identification of genetic resources of perennial crops

K.V. Bhat[1], S. Lakhanpaul[1], K.P.S. Chandel[1] and R.L. Jarret[2]
[1]National Plant Tissue Culture Repository, National Bureau of Plant Genetic Resources, New Delhi-110 012, India
[2]Genetic Resources Conservation Unit, USDA/ARS, Griffin, GA, USA

Introduction

Conservation and increased utilization of genetic resources of crop plants requires detailed characterization and classification of genetic diversity. Evaluation of perennial crops such as banana is more demanding in terms of space and time because of their long vegetative phase. Lack of sufficient morphological markers that can be scored during the vegetative phase of plant growth result in difficulties in distinguishing and identifying germplasm accessions. Highly polymorphic molecular markers like random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) and microsatellites are ideal for characterizing the genetic resources of perennial crops since such markers are independent of the growth stage of plants and their growing conditions. Various molecular techniques such as RAPD (Kaemmer *et al.* 1993; Bhat and Jarret 1995), RFLPs (Gawel *et al.* 1992; Lanaud *et al.* 1993; Bhat *et al.* 1994), chloroplast DNA RFLPs (Gawel and Jarret 1991), variable number tandem repeat (VNTR) loci analysis using synthetic oligonucleotide probes (Kaemmer *et al.* 1993, Bhat *et al.* 1995) and polymerase chain reaction (PCR) analysis of simple sequence repeats (SSRs) using locus specific primers (Akkaya *et al.* 1992) have been used for plant diversity analyses. They have been proved to be ideal for the characterization and classification of plant genetic diversity.

The National Plant Tissue Culture Repository (NFPTCR) established at the National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India is an *in vitro* repository for genetic resources of asexually propagated crop plants. The rich indigenous genetic resources of *Musa, Colocasia, Curcuma, Ipomoea, Dioscorea, Allium* species and medicinal plants are being conserved in the repository. The repository has maintained over 750 accessions of these crop plants for the last nine years. Since the *in vitro* repository has severe constraints on the number of accessions that can be conserved and is very expensive to sustain, duplicates and accessions with low genetic difference from existing material are generally avoided unless they are of special interest.

Molecular techniques are being utilized in NFPTCR to solve the problems related to plant genetic diversity conservation. The major applications are for:

1. Characterizing and classifying germplasm accessions.
2. Identifying specific cultivars and landraces.
3. Estimating the genetic diversity existing in centres of diversity and their representation in the genebanks.
4. Identifying elite types for conservation.
5. Screening of duplicate accessions.
6. Monitoring the genetic stability of conserved germplasm.

These objectives are fulfilled using the following techniques:

1. Isozyme electrophoresis
2. RAPD/DNA amplification fingerprinting
3. RFLP
4. Chloroplast DNA RFLP analysis
5. Simple sequence repeats analysis

Use of above techniques in NFPTCR is detailed here for analyses of the *Musa* collection for which India is one of the important centres of diversity.

Problems associated with *Musa* germplasm conservation

Musa germplasm comprises of few wild species and mainly cultivated forms which are parthenocarpic. The phenomenon of parthenocarpy along with female sterility and structural hybridity (Stover and Simmonds 1989) have made banana breeding a tough task. Therefore banana cultivation has been dependent largely on the naturally occurring variants selected by farmers. The cultivated forms of banana and plantain include *M. acuminata* (A genome) diploids and triploids (AA and AAA), and their hybrid forms with *M. balbisiana* (B genome). The hybrid forms of cultivars belong to AB, AAB, ABB and ABBB genomic groups (Simmonds and Shephards 1955).

Areas of diversity for wild species of *Musa* and the cultivars span from southern India to Japan and Samoa (Simmonds 1966). The cultivars which have been in cultivation for over a century, are an important source of useful genes. However, the practice of conferring names in local dialects has resulted in numerous synonyms and homonyms (Shanmugavelu *et al.* 1992). This has created problems in germplasm conservation as elaborate field testing is required for germplasm characterization and classification using the morphological descriptors (Shanmugavelu *et al.* 1992). The problems are confounded because banana requires one to two years to flower and fruit. In a recent banana germplasm collecting trip, NBPGR collected over 250 accessions from varied agroclimatic regions of India. Classification of these accessions according to the cultivar names revealed that more than 75% of the collections were bearing just 14 cultivar names. Field evaluation of these accessions revealed that as many as 26 out of 35 accessions named Poovan could not be distinguished by morphology even though they had subtle differences (Amalraj *et al.* 1993). Table 1 highlights the problems associated with banana germplasm characterization using morphological characteristics.

Table 1. Classification of *Musa* accessions on cultivar names

Cultivar name	Genomic group	No. of accessions
1. Dwarf Cavendish	AAA	3
2. Karpuravalli	ABB	16
3. Koombillakkai	AAB	3
4. Monthan	ABB	22
5. Morris	AAA	7
6. Naadu	ABB	15
7. Nendran	AAB	10
8. Ney Poovan	AB	16
9. Pachanadan	AAB	24
10. Poongalli	AB	3
11. Poovan	AAB	35
12. Rasthali	AAB	30
13. Red Banana	AAA	3
14. Sakkai	ABB	8
	Total	**195**

The suitability of molecular techniques for solving problems of germplasm conservation was evaluated using a selected set of banana and plantain accessions. The 57 accessions belonging to six different genomic groups are shown below:

Table 2. *Musa* germplasm analyzed for RAPD and RFLP markers

Species/cultivar	Abbr.	Genome[a)]	Source[b)]	Origin[c)]
1. *M. acuminata* ssp. *malaccensis*	Malccn	AA	INIBAP	Introd.
2. *M. acuminata* ssp. *banksii*	Banksi	AA	INIBAP	Introd.
3. Naivedya Kadali	N.kadl	d)	NBPGR	S-W India
4. Gros Michel?	Gr.mic	d)	IIHR	Introd.
5. Amrit Sagar	A.sagr	AAA	NBPGR	N-E India
6. Kanai Bansi	K.bans	d)	NBPGR	N-E India
7. Bar Jahaji	B.jahj	AAA	NBPGR	N-E India
8. Venkel	Venkel	d)	NBPGR	N-E India
9. Red Banana	Redban	AAA	IIHR	S-W India
10. Pacha Bale	Pachbl	d)	NBPGR	S-W India
11. Agneshwar	Agnewh	d)	IIHR	N-E India
12. Ambala Kadali	Ambala	AAA	NBPGR	S-W India
13. Anai Komban	A.komb	AAA	NBPGR	S-E India
14. *M. balbisiana*	Balbis	BB	NBPGR	S-W India
15. *M. balbisiana* cv. Tani	Tani	BB	INIBAP	Introd.
16. Njali Poovan	Nj.poo	AB	IIHR	S-W India
17. Safed Velchi	S.velc	AB	INIBAP	S-E India
18. Ney Poovan	Neypoo	d)	IIHR	S-W India
19. Hu Bale	HUbale	AB	IIHR	S-W India
20. Sakari	Sakari	d)	NBPGR	N-E India
21. Peyan	Peyan	ABB	IIHR	S-W India
22. Sambrani Monthan	S.mont	ABB	IIHR	S-W India
23. Poon Kannan	Poonkn	d)	NBPGR	S. India

Table 2. (contd.) *Musa* germplasm analyzed for RAPD and RFLP markers

Species/cultivar	Abbr.	Genome[a]	Source[b]	Origin[c]
24. Bontha	Bontha	[d]	IIHR	S-W India
25. Thenel Arachi	T.arac	[d]	NBPGR	S-E India
26. Plantain (unnamed)	Plantn	ABB	NBPGR	N-E India
27. Nalla Bontha	N.bont	[d]	IIHR	S-W India
28. Bluggoe	Bluggo	ABB	NBPGR	Introd.
29. Klue Teparot	K.tepa	ABBB	INIBAP	Introd.
30 *M. velutina*	Velutn	[e]	D. Univ.	Introd.
31. Rasthali	Rasthl	AAB	IIHR	S-W India
32. Poovan	Poovan	AAB	IIHR	S-W India
33. Kapur	Kapur	[d]	IIHR	N-E India
34. Agni Rasthali	A.rast	AAB	NBPGR	S-W India
35. Tiruvananthapuram	Tiruvn	AAB	NBPGR	S-W India
36. Hati Dat	Hatidt	[d]	NBPGR	N-E India
37. Vanang	Vanang	AAB	NBPGR	N-E India
38. Walha	Walha	AAB	IIHR	S-E India
39. Kallar Ladan	K.ladn	AAB	IIHR	S-E India
40. Karpura Chakkarakeli	K.chkr	AAB	NBPGR	S. India
41. Mysore Bale	My.ble	AAB	IIHR	S-W India
42. Nendra Bale	Nd.ble	AAB	IIHR	S-W India
43. Ladan	Ladan	AAB	IIHR	S-E India
44. Kaith Khullong	K.klng	AAB	NBPGR	N-E India
45. Nattu Vazhai	Nt.vzh	AAB	NBPGR	S-E India
46. Sambrani Poovan	S.poov	[d]	NBPGR	S-W India
47. Adukka Kunnan	Adk.kn	AAB	NBPGR	S-E India
48. Booditha Bontha Batheesa	BB.bat	[d]	NBPGR	N-E India
49. Myndoli	Myndol	AAB	NBPGR	N-E India
50. Char Padathi	C.padt	AAB	NBPGR	N-E India
51. Nepali Chinia	Npl.ch	[d]	NBPGR	N-E India
52. Atru Singhan	Atr.sg	AAB	NBPGR	S-E India
53. Jira Banana	Jr.ban	AAB	NBPGR	N-E India
54. Local Pakte	L.pakt	[d]	NBPGR	N-E India
55. Garu Maharaj	G.mahj	[d]	NBPGR	N-E India
56. Honda	Honda	[d]	NBPGR	N-E India
57. Sap Kel	Sapkel	[d]	NBPGR	N-E India

a) According to Simmonds and Shepherd, 1955.
b) Abbreviations: IIHR - Indian Institute of Horticultural Research, Bangalore, India; NBPGR - National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi, India; INIBAP - International Network for the Improvement of Banana and Plantain, Germplasm Transit Centre, Leuven, Belgium; D. Univ. - Department of Botany, University of Delhi, New Delhi, India.
c) S-W India - South western states of India; S-E India - Southeastern states of India; N-E India - Northeastern states of India; Introd. - INIBAP, Montpellier, France.
d) Not previously classified.
e) No genome designation.

The accessions included were chosen for their distinctiveness as well as for the close similarities making them difficult to distinguish using morphological markers.

RAPD analyses

PCR amplification of total genomic DNA using 60 random 10-mer primers yielded 605 scorable amplification products (Bhat and Jarret 1995). The amplification products obtained with each of these primers was resolved on 1.4% agarose gels and were scrutinized for the polymorphism and consistency of amplifications and the 12 most discriminatory primers listed in Table 3 were selected as the most useful for *Musa* germplasm characterization and identification. The size of amplification products scored in 1.4% agarose gels ranged between 0.2 and 3.0 kb. The number of accessions distinguishable individually with the selected primers varied from 35 in the case of OPD-08 to 55 with OPC-15. Collectively, these 12 primers were sufficient to distinguish all the cultivars and accessions analyzed in the study. These are probably sufficient for identifying the distinct cultivars of banana and plantains, as the 57 accessions included in this study represent the whole range of variation occurring naturally in AA, BB, AB, AAA, AAB, ABB and ABBB genomic groups.

Table 3. The twelve most discriminatory primer sequences and the characteristics of their amplification products.

Primer[a)]	Sequence	No. bands detected	Size of products (kb)	No. of cultivars distinguishable
OPA-03	5′-AGTCAGCCAC-3′	20	0.3-2.0	45
OPA-04	5′-AATCGGGCTG-3′	23	0.4-3.0	42
OPA-10	5′-GTGATCGCAG-3	20	0.3-2.5	42
OPA-13	5′-CAGCACCCAC-3′	18	0.2-3.0	45
OPC-06	5′-GAACGGACTC-3′	18	0.5-2.5	50
OPC-11	5′-AAAGCTGCGG-3′	10	0.3-1.7	50
OPC-15	5′-GACGGATCAG-3′	24	0.3-3.0	55
OPD-02	5′-GGACCCAACC-3	15	0.3-2.5	37
OPD-03	5′-GTCGCCGTCA-3′	15	0.3-3.0	36
OPD-07	5′-TTGGCACGGG-3′	14	0.2-3.0	45
OPD-08	5′-GTGTGCCCCA-3′	18	0.3-2.5	35
OPD-13	5′-GGGGTGACGA-3′	22	0.2-2.0	45

[a)]Operon Technologies, Alameda, CA, USA.

The usefulness of a technique for germplasm characterization depends on its ability to sample any portion of the genome, study markers on all the linkage groups, detect genetic differences among distinct genotypes, classify the accessions into specific groups which should be comparable to the accepted classifications and screen large number of samples as required in a genebank. RAPD meets all these requirements, although it suffers from other serious drawbacks such as the dominant nature of markers, non-reproducibility of patterns and difficulty in establishing homology of amplification products with similar molecular weights, which reduce the quality of information obtained.

The reliability of RAPD data for the classification of *Musa* germplasm was tested by subjecting the data to unweighted pair group method analysis of arithmetic means (UPGMA) in order to explore the possibility of classifying the cultivars using RAPD analysis. The phenetic analysis primers revealed the presence and extent of genetic

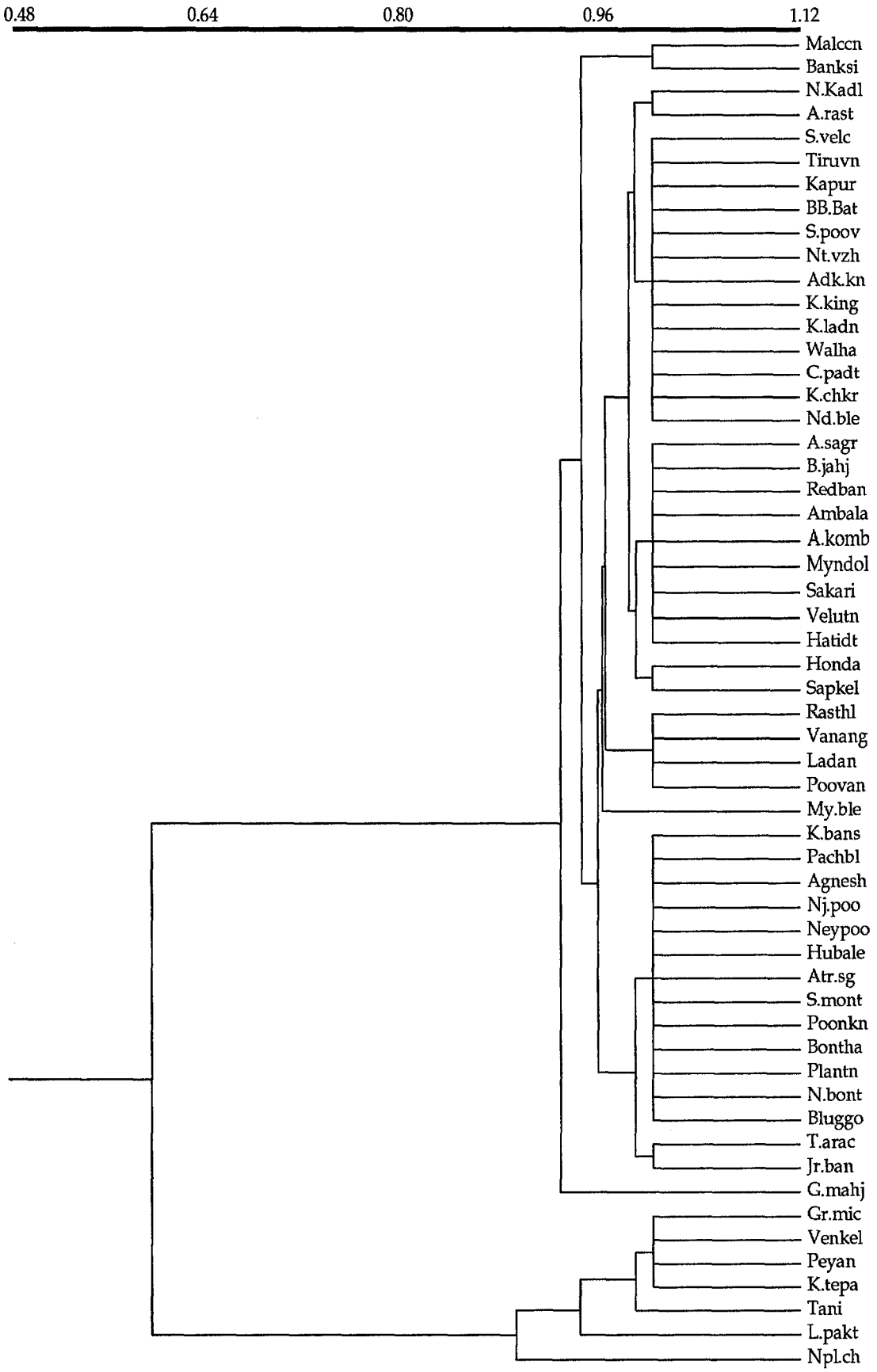

Fig. 1. Phenogram resulting from the analysis of 605 RAPDs depicting relationships between 57 *Musa* accessions. The key to abbreviations are in Table 1.

similarities among the cultivars and species examined (Figure 1). The two diploid species namely *M. acuminata* (A genome) and *M. balbisiana* (B genome) accessions were placed farthest apart in the phenogram. The remaining cultivars were placed in the clusters in between the two progenitor species and their relative closeness to the A or B genome cluster depended on their genomic composition. The cultivars with AAA, AAB and AB genomic constitutions were closer to *M. acuminata* ssp. *malaccensis* whereas cultivars with ABB and ABBB genomic composition were placed in the same cluster as *M. balbisiana* accessions. *M. velutina* predictably did not group with any of the cultivars studied. Most prominent exceptions observed were for the AAA genome triploid cultivars viz. Amrit Sagar, Bar Jahaji, Ambala Kadali, Anai Komban and Red Banana. The sub-clusters containing these cultivars were closer to the sub-cluster with AAB genome cultivars. Moreover, the sub-cluster comprising AB genome cultivars Pachabale, Njali Poovan, Safed Velchi and Hu Bale and the sub-cluster comprising Nalla Bontha, Rasthali, Poovan, Vanang, Poon Kunnan and Adakka Kunnan (all AAB) were closer to *M. acuminata* ssp. *malaccensis* than AAA genome cultivars. The plant material analyzed included 21 unclassified cultivars. These cultivars were placed in different sub-clusters along with those previously classified thus helping in the identification of their genomic composition. The notable exception was a cultivar supplied as Gros Michel, which has been shown to be A genome triploid arising from polyploidization of the diploid species *M. acuminata* ssp. *malaccensis*. This cultivar is supposed to closely resemble the A genome progenitor species. However, in the phenogram it was placed in the sub-cluster made up of B genome diploids and ABB triploids. Therefore to check the reliability of RAPD analysis, alternate techniques were used.

Nuclear DNA RFLPs

RFLP analysis is a highly reliable technique which is repeatable between laboratories. RFLPs were studied by hybridizing random genomic DNA clones picked from genomic DNA library in pUC 19 vector (Bhat *et al.* 1994). The properties of selected nuclear DNA probes used for hybridizations to detect RFLPs are presented in Table 4. High polymorphism was prevalent among the species and cultivars analyzed. The number of alleles detected varied from two each with the probes M-9, M-14 and M-24 to 14 each with M-20 and M-21. The number of cultivars distinguished with each probe differed from two in the case of M-11 and M-24 to 46 with M-21. As evidenced by the RFLP patterns, only M-20 and M-21 were multiple copy sequences and all other probes were single copy sequences.

The cluster analysis of RFLP data separated out the cultivars into six distinct groups (Figure 2). The clustering pattern obtained was comparable to the phenogram from RAPD analysis. The positioning of the diploid progenitor species, namely *M. acuminata* ssp. *malaccensis* (A genome) and *M. balbisiana* (B genome) were similar to that of RAPD analysis. Unlike in the RAPD phenogram, the A genome diploids clustered together as A genome triploids and the AB genome diploid cultivars were in a separate sub-cluster along with four AAB group cultivars. All the ABB cultivars clustered with *M. balbisiana* accessions, the exceptions being Venkel, Local Pakte, Nepali Chinia, Kapur, Sambrani Poovan and Booditha Bontha Batheesa which were grouped in sub-clusters comprising predominantly AAB group cultivars. The Gros Michel cultivar as in RAPD phenogram did not cluster with A genome diploid or A genome triploid. To resolve the identity of this cultivar, chloroplast DNA RFLPs were studied.

Table 4. Nuclear DNA RFLPs in 57 *Musa* cultivars detected by random genomic probes from cv. Maiden Plantain.

Probe	Size(kb)	Alleles detected	Cultivars distinguishable
1. M-6	1.8	5	7
2. M-7	2.0	5	9
3. M-8	1.6	3	4
4. M-9	1.2	2	3
5. M-11	2.0	4	2
6. M-12	3.5	3	3
7. M-14	1.7	2	3
8. M-15	1.6	6	21
9. M-16	2.1	8	11
10.M-18	1.8	8	21
11.M-19	2.0	4	6
12.M-20	1.8	14	30
13.M-21	2.0	14	46
14.M-24	3.2	2	2
15.M-25	1.5	3	6
16.M-26	2.0	4	6
17.M-27	1.7	9	15
18.M-30	2.0	4	4
19.M-32	2.8	7	10

Chloroplast DNA RFLPs

The heterologous cpDNA probes from *V. radiata* used for cpDNA RFLP analysis hybridized intensely to total genomic DNA blots of *Musa* DNA. The differences in cpDNA from *M. acuminata* (A type) and *M. balbisiana* (B type) was evident from the hybridization patterns as the B type of chloroplasts were found to differ from A type chloroplasts in possessing a lower molecular weight fragment. The cpDNA RFLP patterns of the three unclassified cultivars namely Venkel, Local Pakte and Nepali Chinia and the cultivar labeled Gros Michel were similar to that of *M. balbisiana* accession. *M. acuminata* ssp. *malaccensis* and *banksii*, the AA diploid cv Naivedya Kadali and the cultivars with A type cytoplasm had similar cpDNA RFLP patterns. However, among the cultivars with A type cytoplasm the diversity prevalent was evident by the differences in the patterns of hybridization with different cpDNA probes. The phenetic analysis of the results of cpDNA RFLP analysis also support this observation (Figure 3). The phenogram shows the presence of four prominent clusters. All the cultivars with B type cytoplasm were grouped in a single cluster, whereas the cultivars with A type cytoplasm were distributed in three distinct sub-clusters.

These studies underline the need for using a range of molecular markers in genebanks for germplasm characterization, identification and classification. Complementary information from different kinds of analyses can meet the requirements of a conservationist effectively. These markers can also be useful for identifying core collections in genebanks. We have found RAPD and isozyme analyses suitable for characterizing and classifying genetic resources of a number of crop plants such as *Colocasia*, *Abelmoschus*, *Curcuma* and *Allium*. RAPDs, in spite of their drawbacks, seem to be highly adaptable for germplasm characterization.

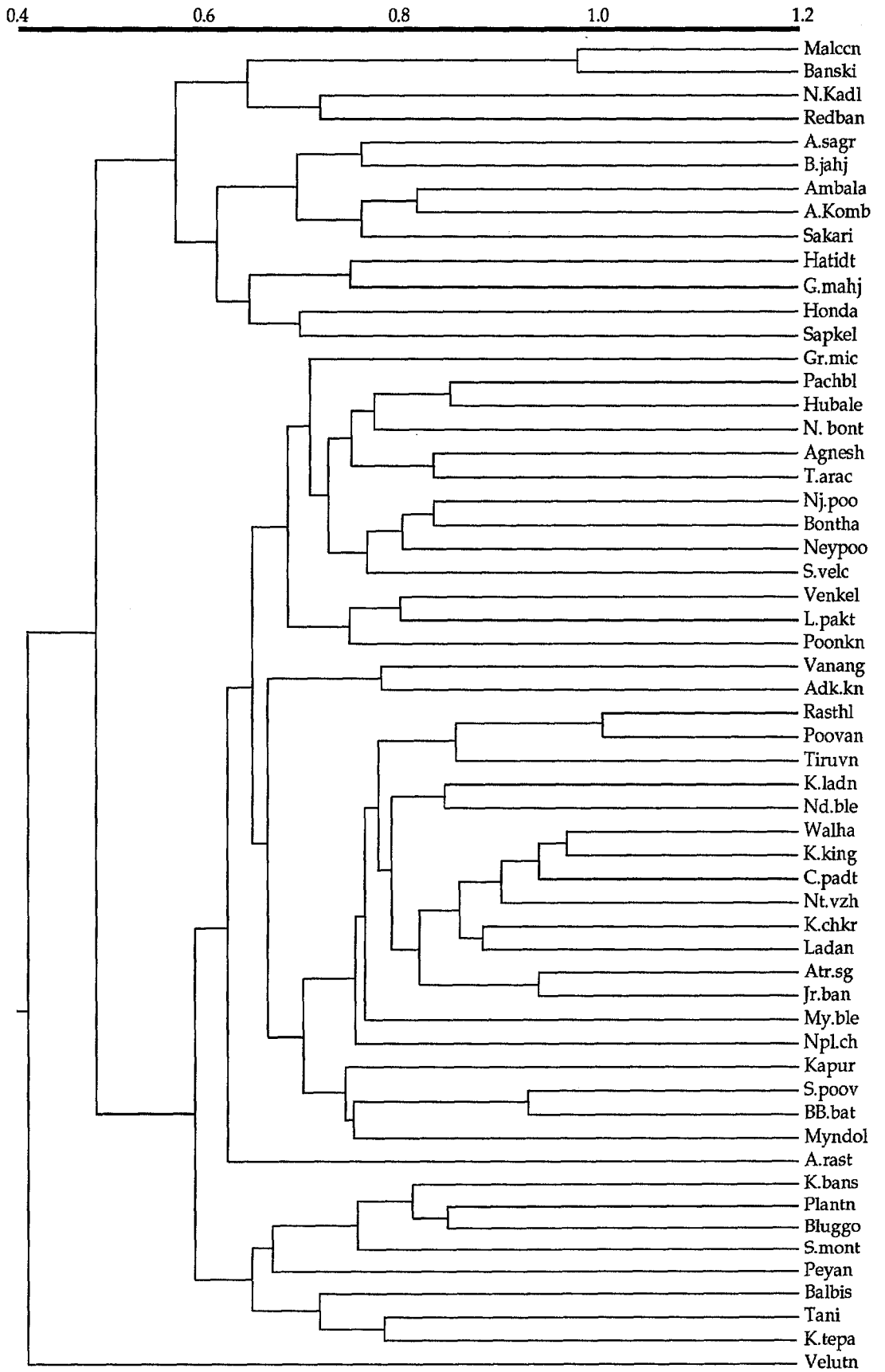

Fig. 2. Phenogram resulting from analysis of 107 RFLP alleles depicting relationships between 57 *Musa* accessions. The key to abbreviations of cultivar names is in Table 1.

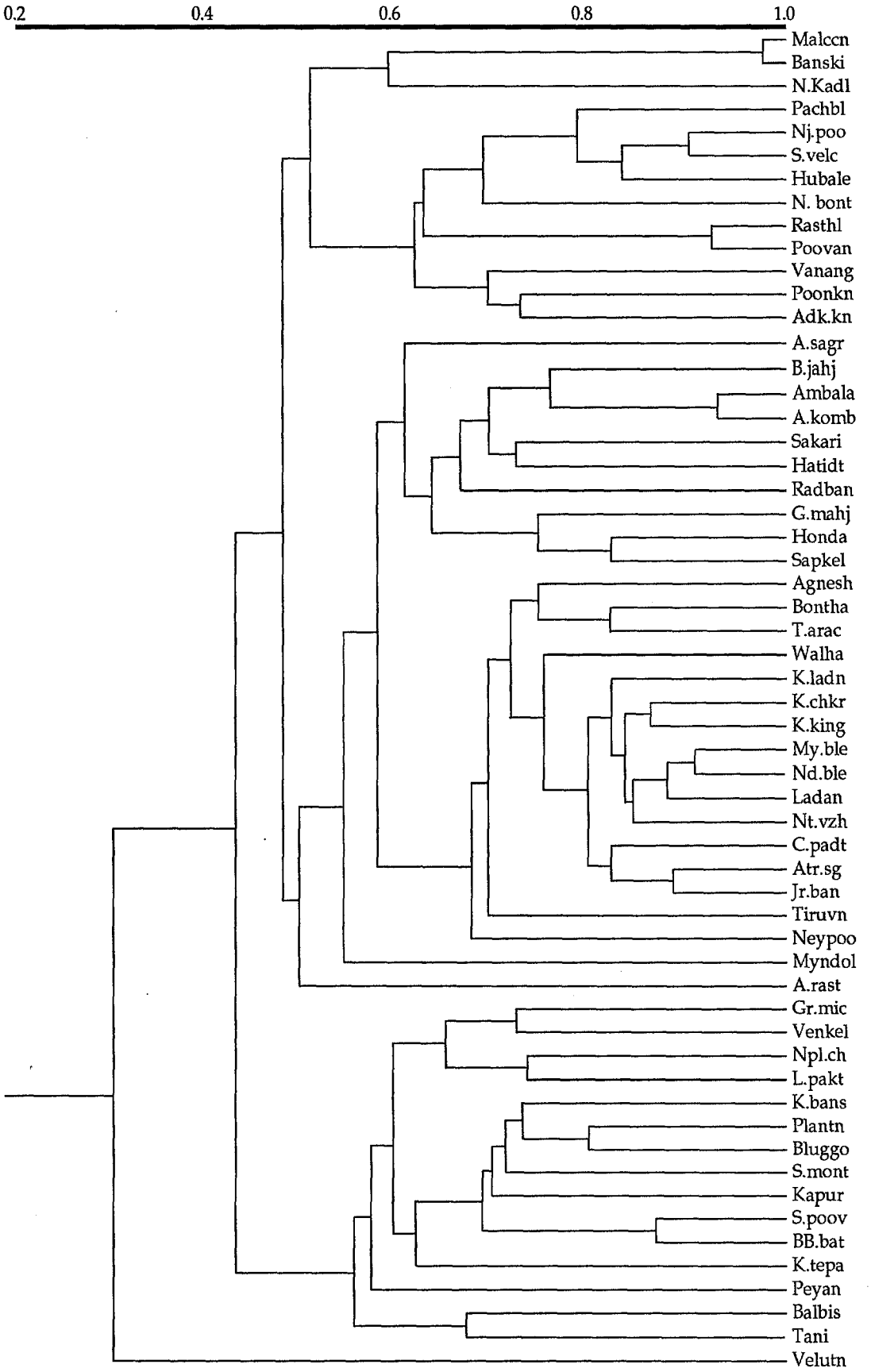

Fig. 3. Phenogram resulting from the analysis of cpDNA RFLP patterns of *Musa* accessions. The key to abbreviations of cultivar names is in Table 1.

Site characterized simple sequence repeat loci are now being screened in *Musa* so that germplasm characterization and classification might be carried out using 15 to 20 primer sets. This would greatly reduce the cost of germplasm screening once the start-up costs are absorbed.

Acknowledgements

The National Plant Tissue Culture Repository was established and is funded by The Department of Biotechnology, Government of India. The financial support provided by the Office for International Cooperation and Development (OICD), U.S.A., in carrying out a part of the characterization work is gratefully acknowledged.

References

Akkaya, M.A, A.A. Bhagwat and P.B. Cregan. 1992. Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Amalraj, V.A, K.C. Velayudan, R.C. Agrawal and R.S. Rana. 1993. Banana Genetic Resources, NBPGR Regional Station Report, Trichur, p. 67, NBPGR, New Delhi.

Bhat, K.V. and R.L. Jarret. 1995. Random amplified polymorphic DNA and genetic diversity in Indian *Musa* germplasm. Genet. Resources and Crop Evolution 42:107-118.

Bhat, K.V., S.R. Bhat, K.P.S. Chandel, S. Lakhanpaul and S. Ali. 1995. DNA fingerprinting of *Musa* cultivars with oligodeoxyribonucleotide probes specific for simple repeat motifs. Genetic Analysis Biomolecular Engineering 12:45-51.

Bhat, KV, R.L. Jarret and Z.W. Liu. 1994. Diversity in chloroplast and genomic DNA for germplasm characterization, clonal identification and classification of Indian Musa germplasm. Euphytica 80:95-104.

Gawel, N. and R.L. Jarret. 1991. Cytoplasmic genetic diversity in bananas and plantains. Euphytica 52:19-23.

Gawel, N., R.L. Jarret and A.T. Whittemore. 1992. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. Theor. Appl. Genet. 84:296-290.

Kaemmer, D., A. Afza, G. Weising, G. Kahl and F.J. Novak. 1993. Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp.). BioTechnology 10:1030-1035.

Lanaud, C., H. Tezenas du Montcel, M.P. Jolivot, J.C. Glaszmann and D. Gonzalez-de-Leon. 1992. Variation of ribosomal gene spacer length among wild and cultivated bananas. Heredity 68:148-156.

Shanmugavelu, K.G., K. Aravindakshan and S. Sathiamoorthy. 1992. Banana Taxonomy, Breeding and Production Technology. Metropolitan Book Co. Ltd., London.

Simmonds, N.W. 1966. Bananas, 2nd ed. Longmans Co., London.

Simmonds, N.W. and K. Shephard. 1955. The taxonomy and origins of the cultivated bananas. J. Linnean Soc. (Botany) London 55:302-312.

Stover, R.H. and N.W. Simmonds. 1989. Bananas, 3rd ed. Longmans Co., London.

Report of the Working Group on increasing the use of plant genetic resources

- Currently, research groups are using molecular markers as tools for:
  - phylogeny
  - characterizing gene pools
  - validating core collections
  - identifying cultivars/genotypes
  - estimating gene diversity
  - screening for duplications in collections
  - monitoring genetic stability
  - selecting improved germplasm

- Several problems were identified throughout the discussion, such as:
  - sample sizes and number of markers to use
  - usefulness for establishing core collections
  - need for universal codominant markers (preferably primers)
  - analysis and reliability of the data being produced
  - importance of genome sampling

The Working Group made the following recommendations for IPGRI initiatives in this area:

a) IPGRI should ensure that two or three staff members visit a number of centres of excellence in any one year to develop familiarity with techniques and their applications.
b) IPGRI should attend UPOV BMTP meetings.
c) The work supported by IPGRI should focus on thematic issues, for example, validation of core collections, evaluation of germplasm, genetic informativeness of ecogeographic procedures.
d) IPGRI should seek to facilitate information exchange and support communication between conservation workers, breeders and molecular biologists (matchmaking) to enhance use of conserved material.
e) Assistance is needed in supporting national programmes in developing project proposals and support project submissions.
f) For international collections under FAO auspices:
   - FAO should take the initiative in calling for their molecular characterization.
g) IPGRI should support collaboration on the development of universal markers/primers.
h) In developing its programme, the Group recommended concentration on four to five crops (orphan crops) for IPGRI supported work. Support collation of information and gap analysis in four to five crops with substantial molecular information.

Technology transfer and application in developing countries

A comparative assessment of molecular techniques employed in genetic diversity studies (and their suitability in resource-limited settings)

Rashid A. Aman
National Museums of Kenya, Nairobi, Kenya

Introduction

Rapid advances in biochemical and molecular genetic technology are revolutionizing many fields of biological inquiry. Molecular biology techniques have had phenomenal impact on many scientific disciplines. The use of genetic information in biological macromolecules, such as proteins and DNA, to address numerous aspects of behaviour, life histories and evolutionary relationships has become routine. Integration of molecular data, especially on DNA, with information from such varied fields as ethology, field ecology, comparative morphology, systematics and paleontology has resulted in the enrichment and rejuvenation of these disciplines. Screening allozymes has been the dominant method for the analysis of genetic variation in natural populations since the 1960s and is still a method chosen for certain applications. However, the power of molecular biology techniques for directly examining variation with increased accuracy and resolution at the nucleotide level has facilitated the application of these techniques to the analysis of different types of questions relevant to population genetics and phylogeny. Molecular markers are increasingly employed to estimate population genetic parameters, of relevance to conservation biology, such as within population heterozygosity, between population gene flow and the genetic distinctiveness of taxonomic units. These indicators of a population's natural history and its prognosis provide valuable data which can form the basis for developing informed conservation management plans for species.

Several molecular techniques are available for the study of genetic diversity. These include karyotype analysis, comparative immunological methods, isozyme electrophoresis, DNA-DNA hybridization, restriction fragment length polymorphism (RFLP) analysis, random amplification of polymorphic DNA (RAPD) and DNA sequencing. All but the first three of these methods are DNA-based and offer direct examination of the genetic material at high resolution. A number of the DNA methods are anchored by the powerful techniques of molecular cloning and polymerase chain reaction (PCR) which, in themselves, are not inherently comparative. The emphasis of this discussion will be on DNA-based molecular techniques and how they can be applied in assessing the genetic diversity of genetic resources.

Considerations in choice of technique(s)

Given this array of techniques, what are the major factors to be considered when choosing among them? Table 1 summarizes some of these considerations and the applicability of the various techniques.

Table 1: Factors determining choice of technique

Technique	Optimal Divergence Range	Resolution	Proportion of Genome Scanned	Complexity of Method	Cost	Time and Labour	Nature of Source Material	Sample Load	Comparability
Isozyme electrophoresis	Low/Med	Low	Med	Low	Low	Low	Tissues fresh or frozen	Med/High	Good
Immunological (MCF)	Med	Low	Low	High	Med/High	Med/High	Purified proteins frozen	Low	Fair
DNA-DNA Hybridization	Med	Low/Med	Med/High	Med/High	Med	Med/High	Large quantities of DNA	Low/Med	Fair
Restriction Enzyme Analysis	Low	Med/High	Low (single locus) Med (multilocus)	Med	High	Med/High	High MW DNA	Med	Good
PCR-based RAPDs	Low	Med	Med	Med	Med	Low/Med	Purified DNA	Med/High	Fair
Targeted PCR	Low	Med	Low	Med	Med	Low/Med	DNA, even degraded from variety of sources	Low/Med	Fair
DNA Sequencing	Low/Med/ High	High	Low	High	High	High	Any purified DNA	Low	Excellent

Genetic divergence

The objectives of any study on genetic variation will determine which one of these methods will be appropriate for the level of analysis sought. It should be recognized that genetic diversity is hierarchically arranged along a spectrum. This ranges from individuals, family units, extended kinships and geographic population structures within species to a graded scale of genetic differences among reproductively isolated taxa that have been separated phylogenetically for various lengths of evolutionary time. Each of these levels will require different approaches. For example, population genetic studies invariably involve low divergences and could be profitably carried out using methods such as RFLP analysis, RAPD or DNA sequencing. Each of these methods will, of course, have its own limitations in terms of resolving power.

Resolving power

Most of the DNA based methods of analysis have high resolving power because they examine DNA variation directly. Sequencing has the highest resolution power as it provides for the comparison of nucleotide sequences of homologous loci.

Proportion of genome scanned or analyzed

As far as this factor is concerned, it is evident that a method such as DNA-DNA hybridization provides for scanning of the entire genome albeit at a lower level of resolution. Methods such as isozyme electrophoresis can scan variation over a reasonable proportion of the protein-coding genome if a sizable number of loci are examined. The RAPD method, on the other hand, is capable of scanning a sizable portion of the total genome due to the random nature of the process.

Complexity of the method

Complexity here refers to the level of technical detail and preparation required to carry out these techniques. While some of the techniques may seem routine and repetitive once the conditions of assay have been set, considerable technical knowledge and detail are required for optimizing experimental conditions and trouble shooting when things go wrong in the case of PCR-based methods and sequencing.

Cost

Cost considerations will involve setting up and maintaining the scientific hardware infrastructure and providing consumables and supplies as well as institutional costs in the form of salaries and overheads. With the exception of isozyme electrophoresis, most of the other methods, and in particular DNA-based methods, require substantial investments in setting up and maintaining laboratories carrying out these analyses.

Labour/Time

Methods will vary in the level of labour intensiveness required which will also be determined by the scope of the particular project. For example, analyzing only a handful of samples for phylogenetic purposes may benefit from a labour intensive method such as DNA sequencing while a large population study may call for a less labour intensive method.

Nature of biological sample

The various techniques differ in the type, state and amount of tissue required for analysis. The sample size may be limited simply because of the lack of availability of the source or the amount of material that can be obtained from an organism. In certain cases, these limitations may call for steps to be taken to increase the amount of material

available such as the tissue culture of fibroblast cell lines for studies using RFLP methods, or the germination of seeds in the case of plants. The methods of sample collection, handling and storage will be dictated by the technique to be employed. Field collection of wild samples from plants and animals present special difficulties which have to be considered when choosing the technique for analysis. Some techniques will require fresh or frozen materials while others will work even with archival specimens.

Comparability

It is important that a data set produced by a particular technique can be easily compared with data from other similar studies. This implies that the method must be consistently reproducible within and between laboratories. In this regard, DNA sequencing affords the greatest advantage as it determines discrete character states of nucleotide sequences which can be easily compared with other independently obtained sequences from the same or homologous locus. Other methods such as RAPD can pose serious reproducibility problems, even within a laboratory, unless the conditions of experimentation are meticulously worked out and standardized within and between laboratories.

Sample load

This refers to the number of independent samples that a technique can accommodate without extending the resources available. Some techniques, such as isoenzyme electrophoresis and a number of the PCR-based methods, allow for the analysis of a sizable number of samples whereas others, such as DNA sequencing, permit the analysis of only a limited number of samples at a time.

Advantages and limitations of these techniques

Isozyme electrophoresis

This method has enjoyed widespread popularity in molecular evolutionary genetics studies since its advent in the 1960s. The method permits the generation of a wealth of Mendelian data on populations in a relatively short time. It can be applied to studies at the population or species levels and, in some instances, can also provide useful information on deeper divergences at the genus or family levels (e.g. *Triticeae*). Starch gel electrophoresis of proteins, coupled with histochemical visualization of locus-specific allozymes, offers a relatively inexpensive and fast method of analyzing single locus variation in natural populations of any life form. The method can handle large sample sizes and uses simple unsophisticated equipment. Standard electrophoretic and staining conditions have been established for numerous loci throughout major groups of organisms so that data can be easily compared. Some of the limitations of the method, however, are its low resolving power and the fact that it needs fresh or appropriately frozen tissue material in order to maintain enzymatic activity which is the basis of the detection method. This makes sample collection and storage difficult in field and uncontrolled situations. The analysis of a variety of tissue types from an organism, while increasing the number of genes that can be assayed, often means sacrificing the organism to obtain these tissue types (destructive sampling).

Immunological methods

Immunological methods rely on the antigenic properties of proteins and their reactions with the specific antibodies that they elicit. The degree of reactivity between antibodies and antigens from different species is measured using various direct or indirect

properties of the antigen-antibody binding process. The difference in antigen-antibody reactivities in tests involving homologous versus heterologous antigens provides a measure of the genetic relationship between proteins from the species being compared. A number of immunological techniques have been used in comparative studies but the most widely used has been microcomplement fixation (MCF) in which antibodies against purified proteins form the basis of the refined estimates of antigen-antibody reactivities. The method has been the basis of many classic studies of molecular evolution. The cross-reactivity registered in immunological tests depends on the affinity and specificity of the antibodies to the antigen, properties which are functions of the immunization protocol as well as the genetic relationships between these molecules. Standardization in techniques is, therefore, important in this method. The method is quite labour intensive, requiring biochemical expertise and facilities for protein purification and raising of antibodies. Sizable tissue and serum samples are required. Although the large body of immunological data in existence will continue to foster interest in this method, its use has become limited since the advent of DNA-based technologies.

DNA-DNA hybridization

DNA solution hybridization studies have been widely used in the analysis of genome complexity and organization. By revealing important aspects of genomic structure such as amounts and lengths of repetitive DNA and interspersion patterns among repetitive and low-copy sequences, solution hybridization techniques have had quite an impact on molecular genetics. Charles Sibley and Jon Ahlquist pioneered the large scale application of this technique to phylogenetics in their classical studies on avian systematics. The method involves measuring the reassociation kinetics of homoduplexes and heteroduplexes in pairwise comparisons of species. The measured difference in the thermal stability provides a quantitative estimate of the genetic distance or divergence between the two species being compared. The method permits scanning of a large proportion (primarily the single copy portion) of the genomes being compared and has been promoted as a powerful source of phylogenetic information. The major criticism of the method is that the raw data consist solely of distance values and not directly observed molecular character states such as nucleotides. Furthermore, factors such as differences in base composition, DNA fragment size and genome size that influence the kinetics of hybridization, have not been completely understood although there have been attempts to standardize some of these. Other drawbacks of the method include the fact that relatively large quantities of DNA are required and pairwise comparisons need to be made between all samples. The method is not suitable for genetic studies at the population level.

Restriction site analysis

Assays of DNA restriction fragment length polymorphism (RFLP) have found widespread application in molecular genetics. The power of the method lies in its ability to make direct comparisons of DNA fragments or restriction sites from a given locus thereby providing information on the nature as well as the extent of variation at the locus. While estimation of sequence variation is possible from fragment data alone, sequence length variation in the region studied can bias estimates of nucleotide substitution. The site mapping approach is, therefore, more appropriate when length variation is a possibility such as when comparing higher levels of divergence. The fragment data technique is relatively easy to use since it is of moderate biochemical complexity and permits the routine analysis of a large number of samples. The various methodological approaches to restriction site analysis in existence differ primarily in

the type and proportion of genome target being analyzed as well as in the levels of complexity involved. Methods may be targeted at analyzing nuclear or organellar (mtDNA, cpDNA) genomes and may examine single copy or repetitive elements within these genomes. The target locus may be studied in crudely or highly purified genomes or in isolation, following amplification from these genomes by PCR prior to restriction site analysis. Fragments generated may be visualized either directly by ethidium bromide staining or after transfer onto membranes by isotopic or non-isotopic techniques. In most routine cases of restriction analysis, relatively substantial quantities of high molecular weight DNA will be required depending upon the number of restriction enzymes to be analyzed. In restriction analysis of organellar DNA, highly purified mtDNA or cpDNA is required as a probe and this often calls for availability of source tissue rich in these organelles. Because certain aspects of the method hinge on techniques such as molecular cloning and PCR, the cost involved in this technique and its various permutations can be substantial.

Random amplification of polymorphic DNA

This PCR-based technique uses short primers of arbitrary sequence to produce random amplification of DNA fragments from the genome being studied. Random amplification is achieved by using single short (9 to 10 bp) primers of arbitrary sequence and annealing at a lower temperature, both of which lower the specificity of the reaction so that a number of anonymous, but often reproducible, fragments are generated from most complex genomes. The method does not require any prior characterization of the genome to be analyzed unlike standard PCR where sequence information is a prerequisite for designing primers. All PCR-based methods enjoy the tremendous advantage of requiring only small amounts of DNA. The RAPD method is simple, fast and permits analysis of a large number of individuals at reasonable cost. The method for detecting bands is simple and inexpensive requiring staining of agarose or polyacrylamide gels with ethidium bromide or silver, respectively. Several other refinements of the RAPD method bearing fancy acronyms have sprung up but the basic principle remains oligonucleotide-directed random amplification. To ensure reproducibility, the method demands careful replication of reaction conditions to achieve consistent amplifications. The RAPD technology finds its greatest application in detecting polymorphisms in closely related organisms (low divergence) such as those that compose a species complex, different populations of a single species or individuals within a population. RAPD markers have already been extensively used in gene mapping research, in individual and strain identification and in those issues in ecology and population biology requiring genetic analysis of relatedness or identity. While the technique offers a number of advantages, it should be recognized that, because of the random nature of the amplifications, one cannot be certain that all comigrating bands seen are homologous (related by evolutionary descent) in all samples analyzed. The problem of uncertain homology becomes serious at higher taxonomic levels where it is likely that only a few shared bands are generated. For this reason, the technique is limited to closely related organisms where total sequence divergence is low and many RAPD bands are shared making the inference of homology stronger. Another limitation of the method is that a good number of the RAPD markers show dominance/recessive inheritance in diploid organisms whereby a particular fragment is amplified from some individuals but not from others, making it impossible to distinguish heterozygotes from homozygotes for the dominant allele. Concern has also been raised on use of the method in parentage analysis because of the occasional appearance of nonparental bands in offspring of known parentage.

Nucleic acid sequencing

The power of this technique derives from the fact that it reveals the discrete order of nucleotides (characters) which are the basic units of genetic information encoded in organisms. While the potential size of data sets amenable to DNA sequencing is vast, it is not logistically possible to gather sequence data from all genes or the entire genome of an organism (although exceptional efforts are ongoing to sequence the entire human genome and the genomes of a number of other scientifically important taxa from representative individuals). In most studies examining genetic variation, logistic constraints restrict sequence acquisition to several hundred base pairs from only one or a few genes in a given study and from relatively small numbers of individuals. The high resolution of sequence data, therefore, comes at the expense of sacrificing a broad base of loci and individuals which might be more desirable in certain cases. On the other hand, because sequence data is of absolute character state, alignable sequences generated by different laboratories can be freely compared and independently generated data sets of homologous loci can be combined. DNA sequencing is an expensive and labour-intensive procedure requiring a fairly high level of sophistication in a laboratory setup. In the period before the advent of PCR technology, preparation of template DNA for sequencing entailed extensive molecular cloning procedures that rendered it prohibitively expensive and time-consuming for population-level studies comparing large numbers of individuals. The coupling of PCR and sequencing has greatly facilitated the application of sequencing to population studies making it possible to generate hundreds of base pairs of sequence from homologous genes from large numbers of individuals or taxa. Recent developments in automation of sequence determination and capture of data have greatly enhanced the speed with which sequence data can be acquired but this is still out of reach for most laboratories because of high costs.

RAPD as choice technique in resource-limited settings

In so far as the use of molecular genetic techniques (i.e. DNA-based techniques) in the study of genetic diversity is concerned, it seems that a prudent choice would be versatile methods that can investigate genomes at a level of complexity above the primary sequence but below the level of karyotypic or gross DNA similarity comparisons; methods which at the same time should be able to support high throughput i.e. analyze large sample sizes rapidly at moderate cost. Given these considerations, it seems that the best choice would be PCR-based methods which also have the enormous advantage of requiring very small amounts of DNA, something that could in itself be limiting in the first place. PCR-based methods capable of scanning substantial portions of genomes such as RAPD (and its various modifications) are particularly appealing. The method is also favoured since in most cases of genetic diversity assessment, particularly in plants, individuals in populations or closely related organisms will be the targets of study. In plant breeding, for example, where molecular markers are required to establish the genetic basis of agronomic traits or in plant gene banking where molecular characterization of accessions can facilitate efficient management of germplasm collections, the RAPD technique is very suitable. DNA fragments generated by the RAPD technique can be easily visualized on agarose or polyacrylamide gels which is a great advantage in resource-limited situations.

Resource limitations

The most important resource limitation in most cases will be the lack of funds to set up facilities and support research in such facilities. Even for existing research facilities, continued funding is a major constraint. Other limitations in resources include inadequacy of scientific infrastructure to support research and lack of trained personnel to carry out such work. Molecular genetic research is highly sophisticated and the technologies involved are advancing at a rapid pace. Without an established scientific infrastructure and culture, it is almost impossible to engage in and keep abreast of developments in molecular biology research. Therefore careful assessment of the resources available or likely to be required in any sustained programme in molecular genetics is very important, particularly in resource-limited situations. Drawing from my personal experience of having established and operated a molecular biology laboratory in a developing country over the past six years, I would like to highlight some of the special constraints or limitations that one is faced with in such settings.

Equipment

Specialized and sophisticated equipment is required for molecular biology research. The array of instrumentation required is large and no single laboratory can possess all of them. The type of research that can be carried out is, therefore, going to be dictated by the instrumentation available. Sharing of equipment, particularly if it is large and expensive, within and between institutions is, therefore, encouraged. Equipment needs to be maintained and serviced regularly to ensure long life. In developing countries, this is a major problem because the technical expertise required to carry out this work is simply not available. Manufacturers of equipment do not have adequate representation in these countries because the markets are not sufficiently attractive for their products. Consequently, once equipment is acquired, often through second or third line vendors, continued backup service is not available. Replacement parts for equipment are not readily available and it is not uncommon to see broken down equipment sitting idle for months on end as parts are awaited to be shipped from overseas. Compounded with the lack of funds, such problems can cripple laboratories.

Procurement of materials and supplies

This problem is related to the one discussed above. Molecular biology work requires a steady supply of consumables ranging from glass and plastic ware to chemicals and biochemicals. Most important of all, this kind of work relies heavily on the use of enzymes and biological products which are perishable if not adequately handled and stored. Where radioactivity is being used, procurement poses special problems owing to the relatively short half-lives of commonly used isotopes. Ensuring that such materials are shipped properly in a timely fashion and that upon arrival at destination they are appropriately handled and stored until safe delivery to the laboratory is a major task in itself. In developing countries where facilities for and knowledge of proper storage conditions at airports may not exist, ensuring that losses due to improper handling do not occur can be difficult. Red-tape bureaucracy for imported materials does not help the situation.

Unreliability of essential services

Essential services in any laboratory include power and water supply. In developing countries, it is not uncommon to experience periodic power interruptions and blackouts as well as seasonal water shortages. These can have a discouraging effect on research activities unless the institution is adequately prepared for such failures by,

for example, having a back-up generator and alternate supplies of water. Although not commonplace, such interruptions can cause havoc with experiments and stored biological materials, as well as taking a toll on equipment as a result of power surges. Not to mention the waning of scientific enthusiasm that this can cause.

Conducive environment

The existence of a conducive environment for research is essential for its success. Providing for scientists to interact and exchange ideas both at local and international levels is important for intellectual nourishment. In developing countries where "research communities" are scarce and far apart, isolation can set in and this can be very detrimental in this age of rapid technological advancements particularly in the field of molecular biology. The lack of adequate library facilities and the inability in many cases to benefit from electronic exchange on the information superhighway only worsen the situation. In this age of DNA research, access to worldwide data banks is absolutely essential.

Support services

Centralized support services such as genebanks, field genebanks, *in vitro* storage facilities, animal holding facilities, radiation handling and disposal facilities, and computing facilities, to name but a few, are essential for any serious biological research programme. Such support services are difficult to maintain in resource-strapped situations.

Above all, no research programme can flourish without sustained financial support.

Conclusion

In conclusion, the state of the art in molecular genetics today offers vast opportunities for research, in both developing and developed nations, for addressing issues relating to ensuring food security, basic health and conservation of biodiversity and the environment. Ways must be found to ensure that developing nations are able to participate in high level research despite the constraints that they may face. The ideals of the United Nations Convention on Biological Diversity and the sharing and transfer of technology must be relentlessly pursued.

Meeting training needs in developing countries

David F. Marshall
Wolfson Laboratory for Plant Molecular Biology, School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Background

For many years at Birmingham, we have run two postgraduate courses which have offered training in Plant Genetics and Plant Genetic Resources to a large and diverse number of students from around the world. Over this period we have seen the development of a growing range of molecular techniques which offer considerable potential as tools in plant genetics, plant genetic resources and plant breeding. In addition to participants on these courses, we also have a regular turnover of research workers from around the world as well as our own Ph.D. programme students and post doctoral fellows who require training in these disciplines. We have to continually assess what level of importance these techniques should have in our training programmes and, in turn, how we should approach the actual details of the training programmes themselves. A key feature is the wide diversity of backgrounds of those we are required to train.

Training for awareness and critical understanding

It is important to identify two needs with respect to training in these molecular techniques. The first of these is the need to provide a sufficient level of understanding about the range of currently available and developing techniques to enable the student to make informed judgements about the value of using molecular techniques in programmes for which they have a direct responsibility or, alternatively, to be able to critically judge molecular data or conclusions based on statistics derived from such data. The ability to make a value judgement of when **not** to use molecular techniques must also be seen to be a valuable outcome of training in this area.

Training for independent research

The second need is to be able to train participants to a sufficient level to enable them to carry out their own experimental programmes using molecular techniques.

The level of training required to enable participants to be self-sufficient in such molecular technologies is not always readily apparent. It depends on several critical factors. First of all the background of the participants is of great importance. How much general molecular and laboratory experience do they already possess? The training problems encountered range from simple understanding of the concept of molarity and the preparation of solutions to detailed understanding of biochemistry and molecular biology. The second major factor is the crop species group. There is a considerable difference between training someone to a level at which they can be self sufficient in work with maize, barley or Brassica crops - where they can rely on the support of a vast international literature and experience - compared with, for example, a high polyploid tree legume which, though it may be locally of major significance,

does not have such available resources. Under such circumstances, researchers are forced to carry the full cost and responsibility of developing molecular techniques almost entirely on their own. I will return to this aspect later. The final and, in many ways, the most important point of concern with regard to future research is the actual working environment in which the trained personnel will eventually work.

This concern applies not just to the equipment resources available and the quality of the laboratory facilities, though these are clearly of importance, but also to the intellectual environment and the quality of access to biochemical supplies. All these have a major bearing on the cost-effective operation of molecular laboratories. Such problems of the working environment are not uniquely associated with establishing molecular technology in relatively poor developing countries. We have experienced directly the wide range of difficulties which arise when transferring molecular technologies to small isolated laboratories even within the UK or elsewhere in Europe. If researchers are working in isolation, especially if they are relatively inexperienced, a whole series of normally minor problems assume enormous magnitude. It helps greatly to have colleague(s) with similar experience in order to discuss procedural problems. It can often be extremely difficult to pinpoint whether a technological problem is due to 'operator error' or to a faulty batch of enzyme or primer etc. In a well-founded laboratory, with many similar projects, the easy solution to this problem is to simply to borrow someone else's solutions, batch of enzyme etc. In contrast, for the isolated worker, months away from access to a new batch of enzyme, pinpointing the problem may require an unproductively expensive factorial experiment and a lengthy pause in their experimental programme. We have helped a number of researchers from small labs in the UK resolve problems of this type.

Cost effectiveness of molecular techniques

Any training programme which is going to be of value must focus not simply on the technology itself but also on the cost- effectiveness of the techniques involved. We really have to train people to be aware of the investment costs that arise in developing the techniques and setting up a laboratory, as well as the recurrent consumable costs of day to day use. Finally, such costs have to be evaluated in terms of the value of the data they provide. This clearly means that training in molecular technology cannot be divorced from developing an understanding of the application of the techniques and the relationship between the molecular data and the variation for agronomically valuable traits. We must never loose sight of the fact that molecular techniques for the measurement of diversity are a **substitute** for the direct analysis of traits that the germplasm resource may eventually be expected to provide to the breeder. In this context, they can only be judged as cost-effective if they are a useful predictor of diversity for these traits themselves and provide an advantage in terms of cost and/or time, or in terms of producing a general picture of diversity in a given source of germplasm.

There are many well documented examples in the literature where molecular diversity is shown to be a relatively poor predictor of agronomically important traits or where even different molecular techniques give a different view of the partition of diversity in a range of germplasm. We know, to our own intense disappointment, that *Nicotiana rustica* which has been the experiment model for much of the quantitative genetics theory developed at Birmingham by Mather, Jinks and their followers, has almost no easily -detectable isozyme or RFLP variation even though it is extremely variable for a wide range of quantitative characters. We have had similar experience with a range of *Nicotiana tabacum* material.

The cost effectiveness of any given marker technology, judged in this way, may vary significantly from country to country and from crop to crop, since trialling costs are a major factor which can vary by orders of magnitude around the world. For example, the difference in costs between trialling a major perennial plantation crop such as oil palm or rubber and small rapidly grown vegetable such as calabrese is considerable.

The historical context

In order to appreciate the value and significance of these molecular techniques, it is appropriate to begin with a review of the range of marker systems that were available before isozyme systems were developed in the sixties and seventies. Clearly the important features that distinguish those early, mostly morphological or physiological, marker systems were that they had gross effects on phenotypes, they were normally recessive and they were individually relatively rare. This meant that their systematic use was strictly limited by the need to breed appropriate marker genotypes into the germplasm under investigation unless fortune presented one with appropriately marked germplasm. Either way, the use of such markers was limited to structured breeding or genetical experiments. However, the one major advantage of these early systems was that they almost invariably required little in the way of sophisticated resources or expensive consumables for their use.

The difficulties and drawbacks of early genetic markers provide the ideal background against which to critically evaluate any new and promising marker techniques. In reality, the best way to judge techniques is to develop a feature list that can be used to describe the ideal marker system. It is important that such a list is couched in terms of the realities of the resources available to the students. This might seem to lead to a rich/poor nation dichotomy, but in our experience the reality is far from this. It comes down much more to terms of just what facilities and resources will be available to the trainee in the future and the extent of the molecular training/background encountered before the course. Our experience, in particular over the last few years, is that the variation in appropriate background experience we find with students of UK origin and training is not greatly different from what we find with the diverse range of overseas students we train. This, to some degree, reflects both the divergent range of botanical interests which will benefit from training in molecular techniques for diversity analysis and the convergence in education and botanical training that has come from global communication.

Just which marker systems should we promote in our teaching/training?

One of the biggest problems which faces anyone providing training in this area is just what marker systems we should talk about. Just a few years ago, it was all rather simple. We talked about isozymes, let everyone run a few gels and described RFLP technology. However, the rate of current change is something of a problem, not at the least because of the varying economic and patent complications associated with many of the promising new technologies. The approach I have adopted over the past few years has essentially been to review all the major technologies against a background of their suitability for a range of typical applications. It is important that researchers are made aware of the range of marker techniques and technologies even if they are unlikely to be in a position to utilize some of the more exotic (complex) techniques in

the near future. For example, the rapid development of efficient automated DNA sequencing apparatus gives some promise that automated microsatellite analysis may rapidly become a much more affordable technology. In particular, since human diagnostics and genome analysis of both crop plants and domestic animals are apparently converging on the use of microsatellite markers for a range of genetic marker applications, there may be considerable economies of scale available through the joint use of laboratory facilities for human, domestic animal and crop plant applications. Such benefits are actively being explored by a number of European organizations which see companies offering genotyping services based on microsatellite or other marker systems as the most efficient route to providing ready access to genotype data for breeding or diversity analysis.

Where to go from here?

As indicated earlier in this document, training is really required on two scales. The first major problem is to train genetic resource professionals to a level where they are able to critically evaluate both molecular and conventional techniques for diversity analysis in the light of the biological problem they wish to solve. This aspect of the training is amenable to a number of relatively simple solutions, some conventional and some novel. A simple course of lectures and perhaps demonstrations offers many advantages. It is a relatively portable format and provided it is taught by experienced practitioners can provide a flexible resource that can evolve with perceived needs. However, despite some new additions to the market, there is still no appropriate text book available around which to base an ideal course. Such a text should cover not only the technical aspects of molecular diversity analysis but should also be integrated with an introduction to the handling and analysis of molecular marker data. We know from our own experience that a lack of experience in critical handling of data is one of the major problems preventing a better understanding of the applications of molecular markers. A considerable volume of computer disk storage around the world is currently occupied by marker data that is either unanalyzed or un-analyzable. Many scientists fall into the common trap of undertaking their experimental work before they concern themselves with data analysis. However, as with many scientific endeavours the correct approach it to begin with a view of the correct analytical model that will be adopted. Which statistics will be used? How many populations? How many markers/loci? Are the numbers of loci/populations/individuals sufficient for the level of discrimination required?

An interesting alternative to the use of a simple text to cover this area would be to use a suitable hypertext medium. Currently the most interesting option would be to construct a text book in HTML (HyperText Markup Language) and make it globally available as World Wide Web document on the Internet via an appropriate Web Server. This approach has a number of interesting advantages. It can be kept 'current' by editing the text at suitable intervals without the need to 'publish' a new edition. The 'Forms' provision of recent HTML releases and extensions enables users both to give and receive feedback on the text and, with the use of appropriate software on the Web Server, students could automatically run analysis programmes on example data sets or their own data. Use of the World Wide Web would also enable pre-existing, relevant WWW resources such as crop genome databases to be incorporated via ***http*** links and useful software to be distributed via embedded ***FTP*** links. For those who do not yet have appropriate Internet links, the text could be distributed either via hardcopy or on CD-ROM.

In contrast, to enable students to be trained to a level at which they can be self - sufficient in their own laboratories requires a much more comprehensive period of training. Our experience indicates that this can only be fully achieved with an extensive period of practical placement of at least four to six months and, depending on the background of the student, they may also require appropriate course support. The resourcing of such placements is not a trivial matter in terms of both consumables and training and requires the full resources of an appropriately well founded laboratory and teaching environment to enable the appropriate breadth of experience to be gained. I suspect that we are really talking about Ph.D. level training, resourced by appropriate national or international funding agencies. Ideally, we also need to provide suitable follow-up support.

Conclusion

We can say that the main problems associated with training in developing countries relate to the diversity of background as well as the need to provide for the diversity of fates that trainees will eventually encounter. The rate of change of 'appropriate' technologies is a major issue as is the diverse range of molecular and analytical skills required for a full and balanced training. However, we are firmly committed to providing such training either at Birmingham or on appropriately resourced courses elsewhere, because we fully understand the importance of this area in underpinning a wide range of research in plant genetics, genetic resources and breeding as well as in ecology and biodiversity. This diversity of applications is of great importance, since it may provide the key to sharing the development costs of appropriate technology. One of the great take home messages is that there are (or there certainly should be) considerable economies of scale associated with marker techniques. The difference in the cost per data point decreases and the quality of the resulting data increases as we go from single person projects to large efficient laboratories - even if the scale of operation on any individual project remains the same. There are also a number of important truths that should be learned with regard to the resourcing of marker technology projects in genetic resources - or breeding for that matter. The key question that should be asked is whether the money that might be available for such a biotechnology spend will come from a fixed budget for genetic resources or breeding work or whether it comes from a development aide spend that can only be used for biotechnology. It is also important to stress the importance of producing data sets of high quality. It is a difficult lesson to both teach and learn, but it is inevitably true that it is better to reject data of dubious quality than preserve it in the data set in the hope that analytical techniques will magically resolve the associated problems. Analytical tools for either diversity or linkage analysis have a frightening ability to 'successfully' analyze some rather bizarre data sets unless some appropriate method of quality control or context is available against which to test the analysis.

It is this latter aspect that is perhaps the most fundamentally important training issue. We cannot teach molecular methodologies in isolation from suitable methods of experimental design and data analysis. This is no different from the trial design issues associated with more classical morphological traits. We have to equip people with the skill to evaluate the issues involved in the efficient partition of a given experimental resource between loci, accessions, populations, individuals etc. and the choice of the appropriate statistical analysis.

Report of the Working Group on technology transfer and application in developing countries

General considerations

It was noted that training was required at all levels to meet the needs for understanding, education, awareness raising, research and application work: Such training needs to have both informal and technical, as well as formal and degree elements. In addition, the different levels of technical, short course and research training help promote effective team establishment.

It is important for training to be organized with due consideration for the context in which it will be applied, the possibilities for take up and follow through, the integration of the training into the home institute's programme and the need for "after care", and refresher information, access to the literature etc. A regional approach is desirable, taking advantage of focal institutes. This promotes economy and more effective implementation and follow up. In developing courses in universities, an eye should be kept to the realities of practical training and implementation, taking advantage, where possible, of local research institutions as partners in the delivery of the informal or practical aspects of training. However, care should be taken not to raise expectations where local application is not feasible.

Gaps

Gaps exist both in a quantitative sense - how much training is available - and a qualitative sense - in which areas. A particular gap was noted in the application of new techniques to non-traditional areas, for example forestry, where there is a need for cross-fertilization from other fields. Information needs to be transferred between different players, from the research laboratory to the field, in order to translate research findings and raise awareness of potential at the grass roots level and to transmit concerns in the field to the researcher.

Resources

Training and technology transfer are resource intensive. An opportunist approach should be adopted, taking advantage of training opportunities wherever they may occur and seeking out all possible sources of funding, multilateral, bilateral etc. The importance of information exchange (for example on training opportunities) is emphasized as a way of making resources go further.

IPGRI's molecular genetic capacity

Molecular genetic technology is developing and changing very rapidly and IPGRI needs to have access to the cutting edge of information. A phased approach is recommended for acquiring the necessary capacity; for the coming period of, say, two years, this could best be achieved by close contact with practitioners, through networks and platforms. Other mechanisms would include consultancies (taking advantage of identified, key resource people), honorary fellows and periodical convening of

specialist groups. In due course, the appointment of a specialist staff member should be reviewed. A larger meeting should be convened to promote the application of molecular genetic techniques to plant genetic resources problems. This could be attached, for example, to the International Genetics Congress. In keeping up to date with developments, IPGRI should look for opportunities for cross fertilization of ideas and techniques between the plant and animal worlds.

Recommendations

a) IPGRI should promote training involving molecular genetics for plant genetic resources at all levels within its capacity and in partnership with other institutions and donors.
b) IPGRI should look for new, non traditional partners to collaborate in training.
c) IPGRI should keep a balance in its programme between training in traditional and new technologies.
d) IPGRI should promote a balanced view and implementation by others.
e) IPGRI should encourage and support collaboration between developing and developed countries at the individual and laboratory levels.
f) IPGRI should facilitate the exchange of research information and materials to help, for example, with the exchange of probes.
g) IPGRI should promote/encourage the application of molecular genetic techniques to neglected plant genetic resource targets, while being aware of the strategic/political issues that may be involved.
h) IPGRI should collate and distribute information on training opportunities.
i) IPGRI should look for opportunities to raise awareness among other expert groups of the importance and potential of the balanced application of molecular genetic techniques in plant genetic resources work and also coordinate deliberations and outputs of such groups.

List of Participants

Dr. Ben Vosman
CPRO-DLO
P.O. Box 16
6700 AA Wageningen
The Netherlands
Fax: +31-317-415983
Tel. +31-317-476980
Email: B.Vosman@cpro.agro.nl

Dr. Rashid Abdi Aman
National Museums of Kenya
PO Box 40658
Nairobi
Kenya
Fax: +254-2-741424
Tel. +254-2-742131/4
or +254-2-742161/4
Email: RAman@tt.sasa.unep.no

Dr. Meredith Bonierbale
Centro Internacional de
Agricultura Tropical (CIAT)
Apartado Aereo 6713
Cali
Colombia
Fax. +57-2- 445-0273
Tel. +57-2- 445-0000
Email: M.Bonierbale@CGNET.COM

Dr. Fernando Gonzalez Candelas
Departament de Genetica
Universitat de Valencia
Dr. Moliner, 50
46100 Burjassot (Valencia)
Spain
Fax: +346-386 4372
Tel. +346-386-4505
Email: Fernando.Gonzalez@uv.es
or gonzalez@evalgb.geneti.uv.es

Dr. K. Venkataramana Bhat
NBPGR
Pusa Campus
New Delhi 110012
India
Tel. +91-11-573.9565

Dr. Angela Karp
IACR-Long Ashton Research Station
University of Bristol
Long Ashton, Bristol BS18 9AF
United Kingdom
Fax. +44-1275-394.281
Tel. +44-1275-392.181

Dr. Stephen Kresovich
USDA-ARS
Genetic Resources Conservation Unit
University of Georgia
Griffin, GA 30223-1797
USA
Fax. +1-770-229-3323
Tel. +1-770-228-7207 / 7254
Internet: skresov@gaes.griffin.peachnet.edu

Dr. Claire Lanaud
Head of AGETROP Laboratory
CIRAD, BIOTROP/AGETROP
2477, avenue du Val de Montferrand
B.P. 5035
34032 Montpellier cedex 1
France
Fax. +33-467.61.57.92
Tel. +33-467 61 58 29

Dr. David Matthews
Department of Plant Breeding
and Biometry
Cornell University
Ithaca, New York 14853
USA
Fax: +1-607-255-6683
Tel. Univ. +1-607-255-9951
Email:matthews@greengenes.cit.cornell.edu

Dr. Marcio Elias Ferreira
EMBRAPA/CENARGEN
National Center for Genetic Resources
and Biotechnology
SAIN - Parque Rural
Caixa Postal 02372
Brasilia DF 70849-970
Brazil
Fax. +55-61-274-3212
Tel. +55-61-273-0100, ext. 132
Email: ferreira@cenargen.embrapa.br

Dr. Takuji Sasaki
Leader, Rice Genome Research Program
NIAR - National Institute of
Agrobiological Resources
Kannondai 2-1-2,
Tsukuba, Ibaraki 305
Japan
Fax: +81-298-38-7468 (am)
or +81-298-38-2302 (pm)
Tel. +81-298-38-7441 (am)
or +81-298-38-2199 (pm)
Email: tsasaki@abr.affrc.go.jp

Dr. Suchitra Changtragoon
DNA and Isoenzyme Laboratory
Biotechnology Section
Silvicultural Research Division
Royal Forest Department
61 Phaholyothin Road
Chatuchak, Bangkok 10900
Thailand
Fax. 66-2- 5794730/3734714
Tel. 66-2-5614292-3, Ext. 440-1
Email: suchitra@mozart.inet.co.th

Dr. Stephen Smith
Coordinator of Germplasm
Conservation and Security
Pioneer Hi-Bred International
Johnston, Iowa
USA
Email: smiths@phibred.com
Fax: +1-515-253-2478
Tel.: 1-515-270-3353

Dr. David F. Marshall
Plant Genetics Group
School of Biological Sciences
University of Birmingham
Edgbaston, Birmingham B15 2TT
United Kingdom
Fax. +44-121-414-5925
or +44-121-414-5911
Tel. +44-121-414-5911
Email: D.FMarshall@bham.ac.uk

Dr. Robert McK. Bird
CIMMYT
Lisboa 27, Apdo. Postal 6-641
06600 Mexico, D.F.
Mexico
Fax. +52-5-726-7559/8; +52-595-544-25
Tel. +52-5-726-9091 / 726-7577
or +52-595-544-00 / 544-10

IPGRI Staff

Dr. Geoffrey Hawtin
Director General, IPGRI
Rome, Italy
Tel. +39-6-51892247
Email: G.HAWTIN@CGNET.COM

Dr. Masa Iwanaga
Deputy Director General - Programme
IPGRI, Rome, Italy
Tel. +39-6-51892200
Email: M.IWANAGA@CGNET.COM

Dr. George Ayad
Senior Scientist, Germplasm Collecting Strategies
IPGRI, Rome, Italy
Tel. +39-6-51892211
Email: G.AYAD@CGNET.COM

Dr. Jan Engels
Director, Germplasm Maintainance and Use Group
IPGRI, Rome, Italy
Tel. +39-6-51892222
Email: J.ENGELS@CGNET.COM

Dr. Thomas Gass
Director, Europe Group
IPGRI, Rome, Italy
Tel. +39-6-51892231
Email: T.GASS@CGNET.COM

Dr. Toby Hodgkin
Director, Genetic Diversity Group
IPGRI, Rome, Italy
Tel. +39-6-51892212
Email: T.HODGKIN@CGNET.COM

Dr. Lyndsey Withers
Director, Documentation,
Information and Training Group
IPGRI, Rome, Italy
Tel. +39-6-51892237
Email: L.WITHERS@CGNET.COM

Dr. Mark Perry
Project Leader, SINGER
IPGRI, Rome, Italy
Tel. +39-6-51892235
Email: M.PERRY@CGNET.COM

Dr. Abdallah Jaradat
Senior Scientist, Genetic Diversity
IPGRI WANA Office
Aleppo
Syria
Fax. +963-21-225 105 or 213 490
Tel. +963-21-247 485
Email: ICARDA-FAX@CGNET.COM

Dr. V. Ramanatha Rao
IPGRI Regional Office for
Asia, Pacific and Oceania
Tanglin
Singapore
Fax. +65-7389636
Tel. +65-7389611
Email: IPGRI-APO@CGNET.COM