5

Molecular Markers and the Management of Genetic Resources in Seed Genebanks: a Case Study of Rice

B.V. Ford-Lloyd[1], M.T. Jackson[2] and H.J. Newbury[1]

[1]School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; [2]International Rice Research Institute, P.O. Box 933, 1099 Manila, Philippines

5.1. Introduction

Plant germplasm has been accumulated over many decades, and is now stored in genebanks in countries around the world. The United States National Plant Germplasm System reputedly holds more than 380 000 different accessions of some 8700 species of plant, within which, for example, there are more than 110 000 accessions of wild and cultivated cereals held in the Small Grains Collection in Idaho (National Research Council, 1991). The Consultative Group on International Agricultural Research (CGIAR) centres conserve about 500 000 germplasm samples of more than 30 crop species and wild relatives, mostly as *ex situ* seed collections, but also as field genebanks and *in vitro* collections. The International Rice Genebank at the International Rice Research Institute (IRRI) has a collection of more than 80 000 samples of rice alone from more than 110 countries (Jackson and Huggan, 1993) comprising landrace varieties, breeding lines and commercial varieties of *Oryza sativa*, landrace varieties of *O. glaberrima*, and all 20 wild species of the genus *Oryza*. Since 1973 over 740 000 packets of rice have been distributed throughout the world for use in basic and applied research, and this germplasm has contributed to improvements in many characteristics of new rice varieties (Jackson, 1994). Pressure for germplasm distribution will increase over the next 30 years as plant scientists strive to meet the demands for a 70% increase in rice production by the year 2025.

The management of such collections is difficult due simply to their vast size, and there is a clear requirement for the development of

 Biotechnology and Plant Genetic Resources (eds J.A. Callow, B.V. Ford-Lloyd and H.J. Newbury)

procedures which utilize fast and reliable methods for the identification of material, the measurement of diversity, or for the determination of redundancy, in order to facilitate the organization and prioritization of germplasm (Virk *et al.*, 1995a). Even more important may be the need to rapidly and efficiently identify the most appropriate rice germplasm for research and crop improvement.

In recent years there has been an explosion of new DNA-based marker methods such as restriction fragment length polymorphism (RFLP) analysis, and those utilizing the polymerase chain reaction (PCR) such as RAPD (random amplification of polymorphic DNA) (Williams *et al.*, 1990). RAPD technology has been used successfully for measuring diversity in plants, and the patterns of variation observed have been shown to closely resemble those obtained using more classical characters (Howell *et al.*, 1994; Virk *et al.*, 1995a). A range of other PCR-based techniques suitable for the measurement of diversity have been developed over the last few years. Some of these target repeat regions of the genome to produce markers. Single primers complementary to minisatellite repeats provide multilocus markers (Matsuyama *et al.*, 1993; Neuhaus *et al.*, 1993). More widespread is the use of pairs of primers complementary to sequences flanking microsatellite repeats. This strategy produces single locus markers which, because of the hypervariability of microsatellite repeat lengths, is an efficient method for the detection of polymorphisms (Morgante and Olivieri, 1993; Senior and Heun, 1993; Cregan *et al.*, 1994). Yang *et al.* (1994) used PCR technology to identify microsatellite polymorphism across landraces and cultivars of rice, whilst Wu and Tanksley (1993) have reported the identification of microsatellite alleles that are specific to either indica or japonica rices. Recently, an additional PCR-based marker technique – amplified fragment length polymorphism (AFLP) analysis – has been developed, which results in the selective amplification of restriction fragments from within a total digest of genomic DNA to yield typically 50–100 dominant marker bands per polyacrylamide gel track (Vos *et al.*, 1995). This method seems certain to find wide application in the study of plant diversity.

Germplasm collections in many parts of the world, like that at IRRI, are faced with challenges related to the various activities which must take place within the genebank, and these are exacerbated by the size of some of the collections and the number of accessions which are conserved. It is now clear that some of the constraints could be alleviated by the application of the molecular marker technology which is becoming more and more readily available. Examples in which molecular markers may be suitably employed to assist genebank management, organization and the way that material is accessed include:

- The accurate identification of germplasm.
- The routine maintenance of germplasm, which is a continuous process

involving seed viability testing, rejuvenation and replenishment of stocks, which will be streamlined by the identification of duplicates and the development of core collections.
- The selection of germplasm for safety storage at other genebanks.
- The choice of germplasm for use by breeders and other researchers involved in making crosses, and mapping, identifying and isolating genes of interest.

Molecular markers are, of course, being used very successfully for the assessment of genetic diversity amongst genetic resources of an increasing number of crop species and wild relatives, and this may take place either before germplasm is accepted into a genebank for storage, or after storage has been initiated. This molecular assessment of 'biodiversity' is dealt with extensively in other chapters.

5.2. Operations Within a Seed Genebank

Confining the discussion only to the primary concerns of a genebank manager, we can identify four principal areas of activity. First is the acquisition of seeds, which involves receiving material as a result of exploration and collection activities or exchange, its preparation for storage, documentation and quarantine procedures. Secondly, seeds are conserved in the form of active/working or base collections, or duplicate safety collections. Thirdly, there are seed management activities including multiplication, regeneration, viability testing, characterization and seed distribution. Finally various activities may take place related to the utilization of germplasm such as detailed evaluation or enhancement.

Taking a more detailed example, the International Rice Genebank at IRRI has as a primary aim the conservation and continued availability of genetic resources for rice improvement worldwide. The germplasm is freely available on request, and is used to continually restore valuable material which has been lost in the country of origin. Base and active collections are maintained, and provision is also made for duplicate 'black box' storage of germplasm at the National Seed Storage Laboratory, Fort Collins, USA. Incoming germplasm is examined by the Seed Health Unit at IRRI, and viability testing is carried out. As and when necessary, germplasm is rejuvenated and multiplied to produce high-quality seed for long-term conservation (Kameswara Rao and Jackson, 1996a,b) and for distribution in response to requests. Wild species receive particular attention and are always grown in pots in a quarantine screenhouse, and perennial species are often maintained as living plants for prolonged periods if seeds are difficult to produce. Seed samples – 10 g each for cultivated rices, but only 10 seeds for the wild rices – are routinely sent out on

request, with more than 740 000 since 1973. When no specific samples are requested, judgements often have to be made as to which material is most appropriate and will most effectively satisfy the demands of the recipient.

No genebank should be a 'museum collection'. The material conserved must be characterized not only to distinguish species, taxonomic groupings and varieties, but also to facilitate preliminary selection of germplasm by end-users. To this end, morphological and agronomic characters are scored in small field plots at IRRI using standard descriptors which will allow the rational choice of material for exchange and distribution, while the maintenance of passport data permits selection of germplasm on an ecogeographical basis. In addition to such routine characterization, IRRI scientists have also screened thousands of accessions for resistance to pests and diseases, and tolerance for different abiotic stresses, an evaluation process which requires more or less elaborate testing of germplasm in the laboratory or field trials.

5.3. Taxonomic Identification as part of Germplasm Characterization

The accurate identification of material held in any genebank is arguably the most essential part of the germplasm characterization process, for without such information breeders will have no means of selecting material for crosses and entry into breeding programmes. Taxonomic identification is an essential first step to determine whether any germplasm is part of the primary, secondary or tertiary gene pool of the crop concerned. While such identification may be undertaken using traditional taxonomic characters, this is not always possible or indeed accurate. A very useful summary of examples of the way molecular markers have contributed to our understanding of crop gene pools is given by Gepts (1995).

In Asian rice (*Oryza sativa*), six crossability groups have been recognized comprising the bulk of the primary gene pool. It is of particular concern to breeders that unambiguous identification of indica and japonica rices can be achieved, as these are currently the focus of plant breeders' attention for crossing and for the development of the 'new plant type', despite the fact that there is difficulty in hybridization and recombination between these two types. Breeders attempting to utilize the considerable variation represented by these groups face increasing difficulty in distinguishing material from these groups, and have regularly used isozymes to help make the necessary identifications (Glaszmann, 1987, 1988). For example, the lines Azucena and PR 304 have been classified as indica using morphological characters, whereas they behave as japonica types in crossing studies (Gurdev Khush, personal communication).

When analysed using RAPD they are clearly revealed as japonica rices (Virk *et al.*, 1995a). Such discrepancies were apparent in another experiment by Virk *et al.* (1995a). Forty-four rice accessions which had been previously classified as indica or japonica on morphological grounds were found by cluster analysis of RAPD data to divide into two major groups. All 31 accessions of one group had been classified as indica; however, eight of the other group had been designated as either japonica or javanica, while the other five had been classified as indica. Clearly the RAPD classifications do not always correlate exactly with classifications based on morphology, but they do accord well with the classifications of rice based upon crossability and isozyme data.

Six *O. sativa* isozyme groups can now be identified using RAPD; these groups reflect precisely those defined by crossability. The same can also be achieved by other molecular marker strategies including RFLP (Wang and Tanksley, 1989; Zhang *et al.*, 1992; Zheng *et al.*, 1994), and microsatellites (Wu and Tanksley, 1993). The ease with which this discrimination can be made using RAPD markers is illustrated in Fig. 5.1.

Of equal importance to genebank management is the ease with which germplasm of closely related species can be identified. In *Oryza*, and particularly within the *Oryza sativa* complex of AA genome species, there is often uncertainty with regard to the allocation of germplasm to several of the species. *O. rufipogon* and *O. nivara* from Asia, and even *O. glumaepatula*, which does seem to have a fairly distinct geographical distribution in South America, have all posed problems of identification using morphological characters. Much more precise identification can be achieved using RAPD markers (Fig. 5.2). RAPD markers have even revealed the true identity of material entering the genebank with the designation *O. meridionalis* that had been misidentified when compared to holotype material (Martin *et al.*, 1997).

5.4. Identifying Duplicates

The race against genetic erosion of crop gene pools has yielded many thousands of accessions safely stored in genebanks. However, genebanks have a finite capacity, and it is apparent that they often conserve more than one sample of the same genotype. In other words, plant genetic resources collections contain duplicate materials, yet the scale of the problem in seed genebanks cannot be determined with certainty. For many vegetatively propagated species, this situation is much more easily addressed.

From a purely management point of view, there are distinct advantages in identifying duplicate accessions, and thereby focusing most effort on unique genetic materials for conservation. Until now the identification

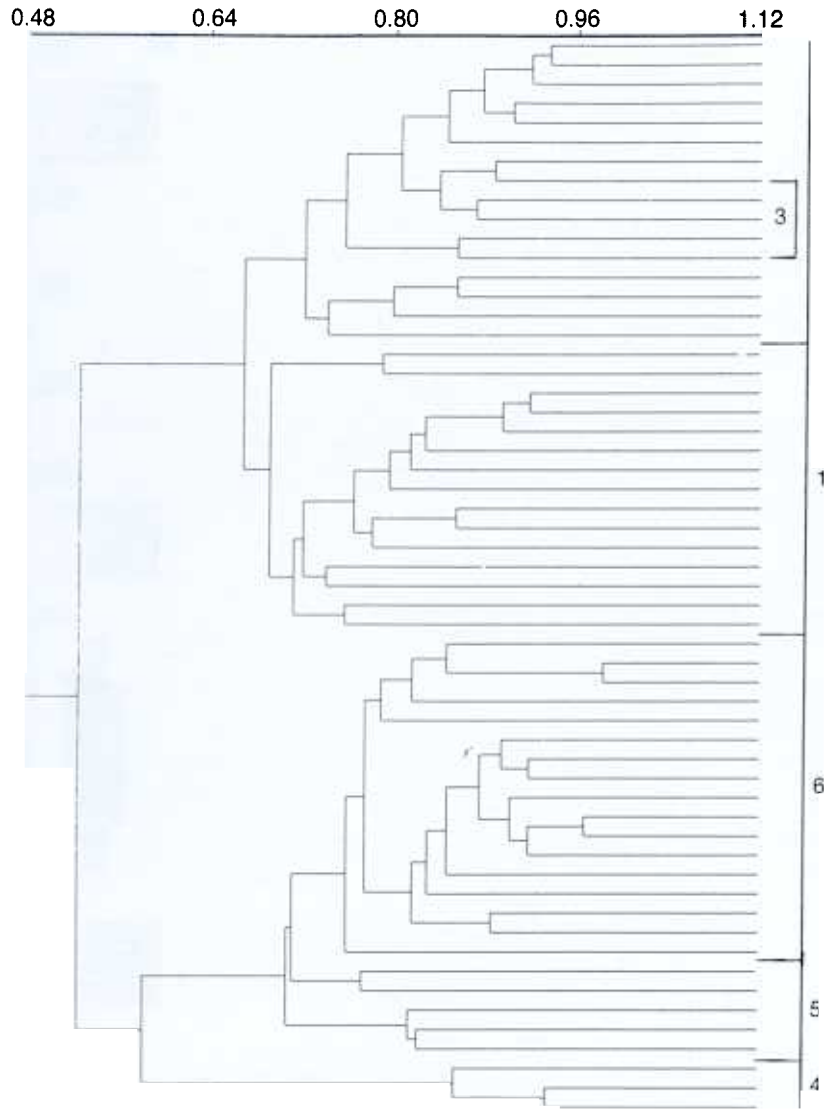

Fig. 5.1. Clustering of *Oryza sativa* accessions according to crossability group, based upon RAPD data. 1–6: rice crossability groups I–VI.

of duplicate accessions has had to rely on comparison of morphological characters, some of which are subject to environmental variation, together with passport data including (amongst others) variety name and origin. Identification of duplicates of vegetatively propagated species, such as potato, is more straightforward than for seed-propagated crops such as

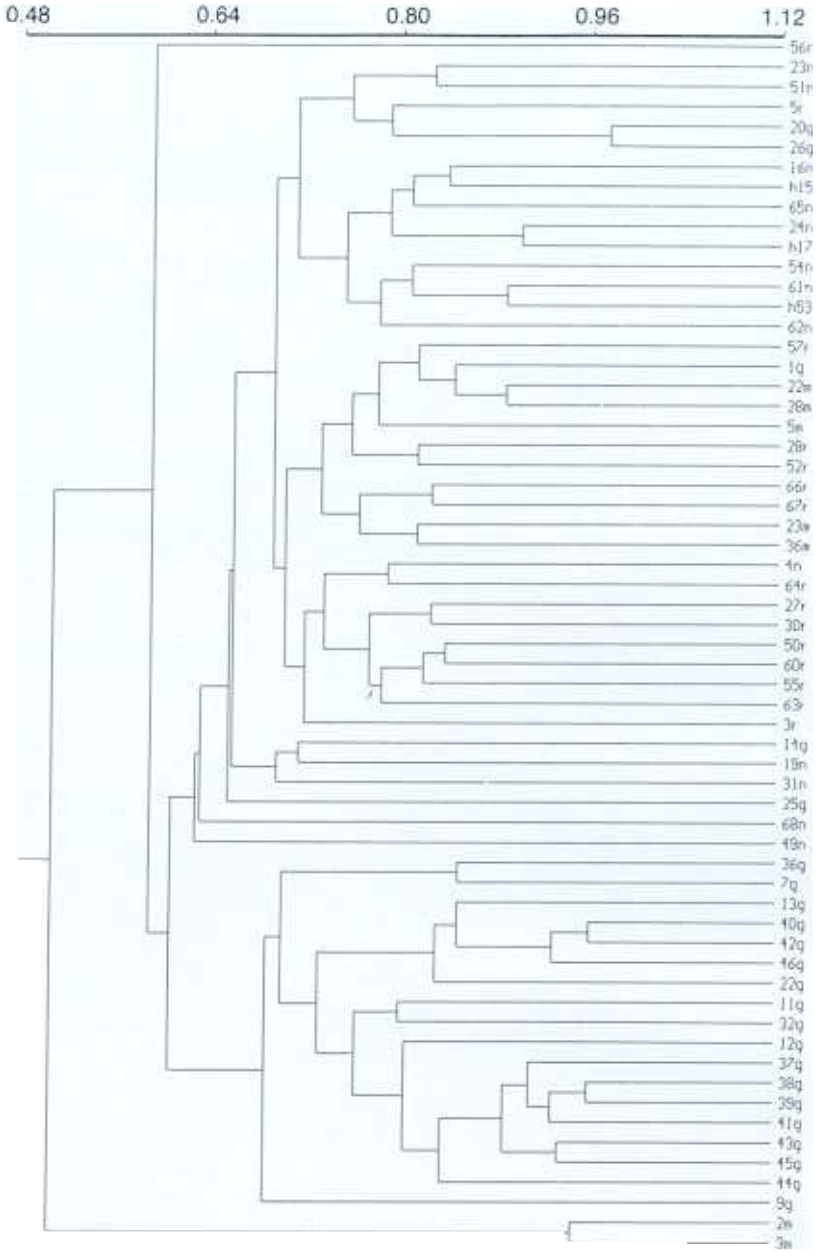

Fig. 5.2. Identification of wild rice species using cluster analysis (simple matching coefficient and UPGMA clustering) of RAPD data. (n = *Oryza nivara*; r = *O. rufipogon*; g = *O. glumaepatula*; m = *O. meridionalis*). Some accessions (e.g. 5, 22, 23, 28 and 36) were subsequently found to have been misidentified.

rice. At the International Potato Center (CIP) duplicate accessions have been routinely identified for some time by comparison of tuber proteins separated in polyacrylamide gels, and complemented by field observations of morphological characters. Duplicate clones have been eliminated from the collection of Andean potato varieties, reducing its size to more manageable proportions.

In more recent work, germplasm samples of rice from IRRI including known and suspected duplicates, as well as closely related germplasm, have been subjected to molecular analysis by Virk *et al.* (1995b). Their results demonstrate that an accurate discrimination of these categories of germplasm samples, including the identification of true and suspected duplicates, can be achieved. Two procedures have been proposed for identifying such duplicates. These differ in the method used in the initial stages, and the relative merits of either method would have to be balanced by those persons charged with conserving a collection. Local situations will vary considerably with regard to the relative costs of field work compared to molecular biology, and the expertise available to carry these out.

Procedure 1

- Select potential duplicate accessions from the collection following examination of available 'passport' data.
- Undertake initial morphological characterization of the suspected duplicates.
- Undertake a full molecular scrutiny of those germplasm samples that cannot be separated using these data.
- Designate as duplicates germplasm samples that cannot then be discriminated.

Procedure 2

- Select potential duplicates from the collection following examination of available 'passport' data as in Procedure 1.
- Carry out a pre-screen of pairs of suspected duplicates using a small number of molecular markers instead of morphological evaluation.
- Undertake a full molecular analysis of those germplasm samples which then cannot be separated.
- Designate as duplicates germplasm samples that cannot be discriminated.

Such results may provide information useful in the design of procedures that permit the routine identification of duplicates within a germplasm collection. Discussions concerning the number of marker bands that it is necessary to score before the designation of duplicates can take place are complex. It will never be possible to prove that two accessions are genetically identical without sequencing their entire genomes. Given that this is a practical impossibility, a decision must be made about the amount

of testing that will be performed before two accessions are accepted as (or 'designated' as) duplicates. This decision must be influenced by the number of potential duplicates that are to be tested. However, the results of Virk *et al.* (1995b) indicate that in rice, for one very similar pair of accessions, we can be 99% confident of detecting a difference between them if we examine a total of 86 RAPD markers. It would clearly be possible to use other types of DNA-based markers for this purpose, although it would be important to ensure that the variation defined using alternative markers was biologically valid in terms of taxonomy and genetics. Moreover, the number of bands to be scored may differ depending upon the sequence types represented by the markers.

The reliable identification of duplicate accessions will provide management options for the germplasm curator. Whether it will lead to the reduction in size of germplasm collections is debatable. In the case of CGIAR centres, their germplasm collections are held in trust, and many accessions are actually intentional duplicate materials of existing national germplasm collections. Clearly the CGIAR centres have an obligation to continue to conserve those germplasm accessions already accepted for safety storage. With the acquisition of new germplasm accessions, however, the situation is potentially different. The study of Virk *et al.* (1995b) suggests a novel procedure which would allow the level of certainty of identifying duplicate samples to be set before those samples became part of a germplasm collection, and before they were assigned unique accession numbers. This option is one which could have a significant impact on germplasm management, provided the PCR-based marker technology can be easily and economically utilized by germplasm curators.

5.5. Core Collections

A major issue in genetic resources for some time has been the size of germplasm collections in relation to their effective management and use. The large size of collections in genebanks may severely constrain many genebank practices. One answer to these problems is to develop a core collection, originally envisaged by Frankel (1984) as being a subgroup of accessions of any germplasm collection which would incorporate, with minimum redundancy, the genetic diversity of a crop species and its relatives. To all intents and purposes the core would form the 'active collection' for germplasm evaluation and distribution, with the remainder of the germplasm being kept as a 'reserve' collection. More specifically there are perhaps two rational and practical motives to develop core collections (Mackay, 1995). The first is to facilitate germplasm management, and the second to increase the use of germplasm by breeders. These two objectives may not complement each other exactly, and indeed the germplasm

curator may well be biased towards the former. However, both objectives could be achieved by the use of molecular markers.

The core collection approach has already been taken for barley, cassava, sorghum, wheat, coffee and *Phaseolus* (Hodgkin *et al.*, 1995). In rice, where the problem of collection size is as great as any, various steps have been taken towards the core approach which aim at fulfilling both of the objectives referred to above. A principal objective of the International Rice Genebank is to establish a core which can help in the safe duplication of accessions representing the broad diversity of the genus *Oryza*, in several locations around the world (Vaughan and Jackson, 1995). The more straightforward the development of this core, the better from the point of view of the germplasm curator. In addition, current knowledge of rice diversity based upon geographic, morphological, agronomic, biochemical and molecular characteristics has resulted in the development of a small core of about 270 accessions of *O. sativa* which represents the known diversity of rice (Glaszmann, 1987; Bonman *et al.*, 1990). Similarly, Vaughan (1991a,b) has designated a core collection of wild rices to enable researchers to evaluate this germplasm efficiently. Continued collecting and biodiversity studies mean that the composition of these core collections must be updated periodically. Molecular and biochemical markers can be used to determine the degree of differentiation that actually exists between wild species themselves and between them and *O. sativa*. Questions about precisely where allelic richness can be found and whether wild species really are sources of distinct alleles can be addressed in order to determine whether new accessions should be added to the core, or whether existing accessions in the core are largely redundant because they contribute little which is genetically unique. The use of the core collection in combination with marker data at IRRI has also enabled rapid identification of germplasm possessing some much sought-after traits. For instance, studies of allozyme data enabled accessions of the small rayada group of rices to be pinpointed for resistance to leaf scald (a seed-borne disease caused by *Rhynchosporium oryzeae*).

How then is the choice of core material to be made? Brown (1989a) has identified stratified sampling as being more efficient in establishing a core than purely random sampling. This relatively simple procedure involves dividing the collection into nonoverlapping groups, and then taking samples from each group. The way in which the groups are established will probably vary by crop species, but will depend upon taxonomy, passport data and ecogeographical information. In the face of uneven distribution of diversity and differentiation of accessions, this method will ensure that the allelic richness of a core will be maximized. One constraint to this approach, however, is the dearth of accurate passport data that unfortunately typifies the situation in many germplasm collections worldwide.

Schoen and Brown (1995) have gone further than this by undertaking a set of simulations. Utilizing allozyme marker data to demonstrate how core collections might be established, they identified how allelic richness could be achieved in cores developed by six different strategies, two of which (H and M) were described for the first time. All six strategies invoked stratified sampling from designated geographical groups, but the H and M strategies differed in that they utilized genetic marker data to guide sampling from within groups. When compared for allele retention the six strategies gave differing results, with the two strategies guided by marker data (M and H) consistently performing best, and the simplest approach, involving only random sampling, the worst.

To what extent molecular markers will be employed in the future via strategies of the H and M categories is arguable. Schoen and Brown clearly indicate that the way forward must be through stratified sampling of an entire collection using passport and other data to develop the groups to be sampled. With the increased use of different molecular marker techniques the limitations of allozyme information will be overcome, so that it will be increasingly possible to improve the selection process of material for the core from within the stratified groups of accessions. Selecting more accessions from groups of high marker gene diversity (H strategy), or targeting particular accessions that are both high in allelic richness and well differentiated (M strategy) can offer the possibility of further improvement over what can be achieved by simple stratified sampling.

In terms of the actual size of any core collection, statistical theory rather than any practical use of marker data has been used principally to determine what needs to be done. Brown (1989a) has argued that the core should consist of about 10% of the whole collection, up to a maximum of about 3000 accessions, for each species. He estimates that, at this level of sampling, the core will generally contain over 70% of the alleles present in the whole collection (Brown, 1989b). This seems to be a rule of thumb which many germplasm curators are adopting.

However, an alternative approach to this problem is provided by the work of Lawrence *et al.* (1995a,b), who considered the size of sample required to capture at least one copy of each allele at each of a number of independently inherited loci at a given probability. Their calculations indicate that, provided the sample size chosen gives a very high probability of conserving the alleles of a single locus, this size is also sufficient to give a high probability of conserving at least one copy of each allele at all other loci. Calculations based upon assumptions that the average genome of a species contains 40 000 structural loci (Nei, 1987), and that 40% of these loci are polymorphic (Hamrick, 1989), indicate that, even if the species is predominantly inbreeding, a sample size of only 172 will give a very high probability (>0.999999828) of conserving all of the alleles at all the polymorphic loci, even if the frequency of one allele at each locus is only 0.05.

In practice, therefore, it should be relatively easy to conserve all or very nearly all of the alleles of a population in a random sample of 172 plants!

Whether core collections are made up of samples taken from 172 plants or 3000, it would seem to be clear that the sampling strategy chosen to select the material which will make up those numbers is still of overriding importance and can be substantially assisted by using stratification and molecular markers.

5.6. Facilitating the Use of Germplasm

PCR-based molecular markers are increasingly being used to assess genetic diversity in germplasm collections. Sometimes, morphological data including those for quantitative traits of economic importance are also available from the genebank. There is an increasing desire to utilize this wealth of useful information, somehow to use molecular markers to assist identification of useful characteristics amongst conserved germplasm, and therefore to narrow the gap between genebank managers and plant breeders. One of the successes of the CGIAR conservation effort over three decades has been the close linkage between conservation and exploitation of germplasm. This has led to many outstanding examples of varietal release to alleviate hunger in developing countries.

Molecular markers are increasingly being used in marker-assisted selection programmes (Stomberg *et al.*, 1994). Both theoretical and experimental studies have shown that marker-assisted selection can be highly effective for producing improved genotypes. However, the success of such selection programmes is largely taken to depend on genetic linkage between markers and the relevant gene loci. Whilst such studies are invariably based on materials derived from planned crosses, could similar principles be applied to genetic resources held in germplasm collections? The work of Virk *et al.* (1996) has gone some way to achieving this by using multiple regression analysis to predict the performance of germplasm accessions of rice given the molecular marker genotypes of those accessions. From a large number of markers it was possible to pinpoint a handful which are significantly associated with a particular trait of interest. Subsequently they were able to make accurate predictions of field performance in a range of agronomic traits such as plant height, culm number, and days to flowering in a particular environment. It is known that associations can exist because of linkage disequilibrium as well as linkage (Hastings, 1990). It is also clear that for marker-assisted procedures to work for prediction or selection of complex inherited traits, it is of benefit that a high level of linkage disequilibrium must exist, particularly if the genome is not well saturated with markers (Stuber, 1990). This appears to be the case in the study of Virk *et al.* (1996), and if true, it appears that

associations between alleles at quantitative trait loci (QTLs) and at marker loci has been conserved throughout the period of diversification of rice germplasm in South and Southeast Asia.

One obvious benefit of obtaining information about molecular markers and quantitative traits would be the more efficient selection of putative parents for producing populations to map QTLs for a particular trait. Also, the procedure could be used as an initial screening method for the identification of QTLs. The established method for this is the selection of two parents that differ markedly in a particular quantitative character, and then the determination of associations between markers and that character in F_2 or backcross progeny. The apparent advantages of using diverse germplasm instead are: (i) that this could allow the detection of QTLs that vary across a wide spectrum of biodiversity rather than just between two parental lines; and (ii) that QTLs for any quantitative trait can be studied in the same investigation.

Regardless of the underlying causes of the associations which have been detected, the use of molecular markers, which are more or less randomly distributed across the genome, coupled with multiple regression analysis could substantially change and improve the way in which crop biodiversity is used in the future. The combination of techniques should allow the prediction of what a plant will look like in terms of quantitative agronomic traits prior to elaborate field trials. If a diverse test array of germplasm is scored for important traits requiring specialized assessment conditions (such as stress tolerances, for example) then marker data could provide an efficient means of predicting the value of additional germplasm for such characteristics. Such results may demonstrate the value of *ex situ* plant germplasm collections not just as repositories of useful genes, but also as sources of information about phenotypic characters.

5.7 Conclusion

One of the major criticisms regularly levelled at genetic resource conservationists over the last 40 years has been that they have frequently been unable to provide appropriate material for crop improvement programmes. However, with appropriate organization of conserved material and the application of current DNA-based marker technology, genebanks can more easily counter these criticisms and become much more valuable interfaces between the activities of conservationists on the one hand and those wishing to exploit germplasm for the benefit of humankind on the other. Often the separation of those who conserve germplasm from those who wish to use it is a barrier to effective exploitation of this valuable germplasm. The advent of molecular characterization and evaluation of germplasm opens another chapter in genetic conservation, and one which

will fundamentally change our perspectives on the nature, structure and value of crop gene pools.

References

Bonman, J.M., Mackill, A.Q. and Glaszmann, J.C. (1990) Resistance to *Gerlachia oryzae* in rice. *Plant Disease* 74, 306–309.

Brown, A.H.D. (1989a) The case for core collections. In: Brown, A.H.D., Marshall, D.R., Frankel, O.H. and Williams, J.T. (eds) *The Use of Plant Genetic Resources*. Cambridge University Press, Cambridge, pp. 136–151.

Brown, A.H.D. (1989b) Core collections: a practical approach to genetic resources management. *Genome* 31, 818–824.

Cregan, P.B., Akkaya, M.S., Bhagwat, A.A., Lavi, U. and Rogwen, J. (1994) Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants. In: Gresshof, P.M. (ed.) *Plant Genome Analysis*. CRC Press, London, pp. 47–56.

Frankel, O.H. (1984) Genetic perspectives of germplasm conservation. In: Arber, W.K., Llimensee, K., Peacock, W.J. and Starlinger, P. (eds) *Genetic Manipulation: Impact on Man and Society*. Cambridge University Press, Cambridge, pp. 161–170.

Gepts, P. (1995) Genetic markers and core collections. In: Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. and Morales, E.A.V. (eds) *Core Collections of Plant Genetic Resources*. Wiley, Chichester, pp. 127–146.

Glaszmann, J.C. (1987) Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics* 74, 21–30.

Glaszmann, J.C. (1988) Geographic pattern of variation among Asian native cultivars (*Oryza sativa* L.) based on fifteen isozyme loci. *Genome* 30, 782–792.

Hamrick, J.L. (1989) Isozymes and the analysis of genetic structure in plant populations. In: Soltis, D.E. and Soltis, P.S. (eds) *Isozymes in Plant Biology*. Chapman & Hall, London, pp. 87–105.

Hastings, A. (1990) The interaction between selection and linkage in plant populations. In: Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S. (eds) *Plant Population Genetics, Breeding, and Genetic Resources*. Sinauer, Sunderland, MA, pp. 163–180.

Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. and Morales, E.A.V. (eds) (1995) *Core Collections of Plant Genetic Resources*. Wiley, Chichester.

Howell, E.C., Newbury, H.J., Swennen, R.L., Withers, L.W. and Ford-Lloyd, B.V. (1994) The use of RAPD for identifying and classifying *Musa* germplasm. *Genome* 37, 328–332.

Jackson, M. (1994) Preservation of rice strains. *Nature* 371, 470.

Jackson, M. and Huggan, R. (1993) Sharing the diversity of rice to feed the world. *Diversity* 9, 22–25.

Kameswara Rao, N. and Jackson, M.T. (1996a) Seed longevity of rice cultivars and strategies for their conservation in genebanks. *Annals of Botany* 77, 251–260.

Kameswara Rao, N. and Jackson, M.T. (1996b) Seed production environment and storage longevity of japonica rices (*Oryza sativa* L.). *Seed Science Research* 6, 17–21.

Lawrence, M.J., Marshall, D.F. and Davies, P. (1995a) Genetics of conservation. I. Sample size when collecting germplasm. *Euphytica* 84(2), 89–99.

Lawrence, M.J., Marshall, D.F. and Davies, P. (1995b) Genetics of conservation. II. Sample size when collecting seed of cross-pollinated species and the information that can be obtained from the evaluation of material held in gene banks. *Euphytica* 84, 101–107.

Mackay, M.C. (1995) One core collection or many? In: Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. and Morales, E.A.V. (eds) *Core Collections of Plant Genetic Resources*. Wiley, Chichester, pp. 199–210.

Martin, C., Juliano, A., Newbury, H.J., Lu, B-R., Jackson, M.T. and Ford-Lloyd, B.V. (1997) The use of RAPD markers to facilitate the identification of *Oryza* species within a germplasm collection. *Genetic Resources and Crop Evolution* 44, 175–183.

Matsuyama, T., Oyama, M. and Akiyama, T. (1993) DNA fingerprinting in *Citrus* cultivars. In: Hayashi, T., Omura, M. and Scott, N.S. (eds) *Techniques on Gene Diagnosis and Breeding in Fruit Trees.*. Fruit Tree Research Station, Ibaraki, pp. 26–30.

Morgante, M. and Olivieri, A.M. (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant Journal* 3, 175–182.

National Research Council (1991) *Managing Global Genetic Resources. The US National Plant Germplasm System*. National Academy Press, Washington.

Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Neuhaus, D., Kuhl, H., Kohl, J.G., Dörfel, P. and Borner, T. (1993) Investigation on the genetic diversity of *Phragmites* stands using genomic fingerprinting. *Aquatic Botany* 45, 357–364.

Schoen, D.J. and Brown, A.H.D. (1995) Maximising genetic diversity in core collections of wild relatives of crop species. In: Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. and Morales, E.A.V. (eds) *Core Collections of Plant Genetic Resources*. Wiley, Chichester, pp. 55–76.

Senior, M.L. and Heun, M. (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. *Genome* 36, 883–889.

Stomberg, L.D., Dudley, J.W. and Rufener, G.K. (1994) Comparing conventional early generation selection with molecular marker assisted selection in maize. *Crop Science* 34, 1221–1225.

Stuber, C.W. (1990) Molecular markers in the manipulation of quantitative characters. In: Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S. (eds) *Plant Population Genetics, Breeding, and Genetic Resources*. Sinauer, Sunderland, MA, pp. 334–350.

Vaughan, D.A. (1991a) Choosing rice germplasm for evaluation. *Euphytica* 54, 147–154.

Vaughan, D.A. (1991b) Core collections: Providing access to wild *Oryza* germplasm. In: *Rice Genetics II*. IRRI, Manila.

Vaughan, D.A. and Jackson, M.T. (1995) The core as a guide to the whole collection. In: Hodgkin, T., Brown, A.H.D., van Hintum, Th.J.L. and Morales, E.A.V. (eds) *Core Collections of Plant Genetic Resources*. Wiley, Chichester, pp. 229–239.

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

Virk, P.S., Newbury, H.J., Jackson, M.T. and Ford-Lloyd, B.V. (1995b) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. *Theoretical and Applied Genetics* 90, 1049–1055.

Virk, P.S., Ford-Lloyd, B.V., Jackson, M.T., Pooni, H.S., Clemeno, T.P., and Newbury, H.J. (1996) Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* 76, 296–304.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23, 4407–4414.

Wang, Z.Y. and Tanksley, S.D. (1989) Restriction fragment length polymorphism in *Oryza sativa* L. *Genome* 32, 1113–1118.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18, 6231–6235.

Wu, K.-S. and Tanksley, S.D. (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. *Molecular and General Genetics* 241, 225–235.

Yang, G.P., Saghai Maroof, M.A., Xu, C.G., Zhang, Q. and Biyashev, R.M. (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Molecular and General Genetics* 24, 187–194.

Zhang, Q., Saghai Maroof, M.A., Lu, T.Y., Shew, B.Z. (1992) Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theoretical and Applied Genetics* 83, 495–499.

Zheng, K., Qian, H., Shen, B., Zhuang, J., Lin, H. and Lu, J. (1994) RFLP-based phylogenetic analysis of wide compatibility varieties in *Oryza sativa* L. *Theoretical and Applied Genetics* 88, 65–69.