Marker-Assisted Reduction of Redundancy in Germplasm Collections: Genetic and Economic Aspects

Rob van Treuren[1] and Theo J.L. van Hintum
Centre for Genetic Resources, The Netherlands
DLO Foundation
P.O. Box 16, 6700 AA Wageningen, The Netherlands
[1]Corresponding author
Email: r.vantreuren@plant.wag-ur.nl

Keywords: collection management, genebank economics, genetic diversity, plant genetic resources, molecular markers, optimization, rationalization

Abstract

Germplasm collections invariably contain duplicate accessions, both within and between genebanks. These redundancies are a burden for curators because they do not contribute to the diversity in the collection, but do require genebank budget for maintenance. Thus, both from a genetic and economic point of view, identification and elimination of redundancies should be an important genebank objective. Molecular analysis can complement traditional approaches towards identifying duplications. Interpretation of molecular data is, however, by no means straightforward because various genetic relationships among potential duplicates may occur. Since in all collections, even for self-fertilizing crops intra-accession variation is often observed, the question is not so much whether two accessions are identical, but rather whether they are sufficiently different to consider them distinct. To address this question, statistical methods can be applied to estimate and test molecular variance components. Genetic issues related to the answering of this question are illustrated by experimental data from redundancy studies carried out in various crop collections of the Centre for Genetic Resources, The Netherlands. In addition, economic aspects of rationalization studies are discussed. It is concluded that, considering the high costs of molecular analyses, the potential economic benefits of rationalization with the help of molecular technology should not be taken for granted and indeed strongly depend on the crop involved. In contrast, the added value obtained by molecular characterization in improving insight in the genetic composition of collections and the quality of genebank operations can be regarded considerable. Examples of spin-off from such studies are presented.

INTRODUCTION

Germplasm collections established and maintained by genebanks provide for the present and future utilization of crop genetic resources. In the initial phase of developing germplasm collections, the focus has been mainly on the acquisition of material due to the necessity of conservation. Many genebank collections have started as working collections for taxonomic studies and plant breeding activities. In many cases genebanks expanded their collections by including obsolete varieties and material obtained from expeditions to natural distribution areas of crop-related wild relatives. To extend the collection, it has also been common practice to request material from colleague genebanks. In many cases, this has resulted in collections of considerable size, rather unbalanced composition and unknown value. It has been estimated by the FAO that the number of genebank collections exceeds 1300, collectively containing more than six million accessions (FAO, 1996). As a result, genebanks are nowadays more concerned with gaining knowledge about their current collections and with the improvement of the efficiency of genetic resources management, rather than with the enlargement of their collections. Research aiming at the optimization of the composition of collections is therefore of particular interest, both from a genetic and economic point of view.

One of the possibilities to improve the composition of a collection is to identify and eliminate redundancies. Germplasm collections invariably contain duplicate accessions, both within and between genebanks. Duplicates may occur within collections for various reasons. For instance, transformation of information and common errors may give rise to the presence of identical material registered under different identifiers. Accessions originating from breeding activities may involve material derived from a common or identical genetic background. In the case of wild material, expeditions may have been organized to identical collection areas, or sampling may have been carried out without sufficient prior knowledge about the distribution of genetic variation across the natural areas. Duplication between genebanks, other than intended as safety backups, may have originated from the acquisition of material from common sources or from the exchange of material between genebanks. Therefore, international co-operation is indispensable to improve the efficiency of genetic resources management within the network of genebanks. The level of duplication encountered can be quite substantial. Based on passport data, a duplication study within and among four main lettuce collections revealed an average of 12% internal duplication, whereas the percentage of accessions unique to single collections ranged only from 33 to 54 (Hintum, 2000). Evidently, redundancies are a burden for curators because they do not contribute to the diversity conserved, but do consume genebank budget for maintenance. Thus, both for genetic and economic reasons, identification and elimination of redundancies are important options for improving genebank efficiency.

Probable duplicates can be identified by comparison of passport data (Hintum and Knüpffer, 1995; Hintum, 2000). These analyses may reveal accessions with identical or similar names, samples originating from identical collecting areas, or material with a common breeding history. Suspected redundancies can then be verified with additional information, such as data from morphological or evaluation studies. However, such studies may be hampered by a limited number of suitable characters that can be scored, the restricted level of variation that is often observed, and the sensitivity of many characters for environmental influences. Presently, molecular marker technologies can be applied as a useful tool in many aspects of plant genetic resources management (Bretting and Widrlechner, 1995; Brown and Kresovich, 1996). Therefore, redundancy studies are increasingly being carried out with the help of molecular characterization data (Waycott and Fort, 1994; Virk et al., 1995; Oliveira et al., 1997; Phippen et al., 1997; Cervera et al., 1998; Zeven et al., 1998; McGregor et al., 2002).

The application of molecular marker analysis, however, introduced new issues related to redundancy studies. First, molecular marker technologies are used because of their higher resolving power, compared to traditional methods. As a result of the higher levels of variation that are usually detected, probable duplicates will rarely found to be completely identical. This may even be the case for self-fertilizing species for which intra-accession variation is often observed (Treuren and van Hintum, 2001). Therefore, verification of probable duplicates using molecular markers is not simply a matter of determining whether two accessions are identical or not, but rather whether they are sufficiently different in order to consider them distinct accessions. In that respect, the use of the term “duplicate” should preferably be avoided, and the term “redundancy” be used instead. The question, however, is how to decide whether populations are sufficiently different. This will require quantification of variation within and among samples. Statistical tests, such as AMOVA, may thereby be used to obtain necessary decision criteria (Excoffier et al., 1992; Phippen et al., 1997). Second, the variation that is measured by the majority of molecular markers will generally be neutral in nature, unless markers are used that are targeted to specific genes. Users of genetic resources will generally be interested in variation in a few specific characteristics. An important question in characterization studies therefore is to what extent molecular marker data are representative for functional diversity. This may vary with the kind of material analyzed, i.e. cultivars, landraces and wild material. For example, cultivars may appear redundant based on molecular data, but nevertheless differ in a single important selected character.

From a user perspective, available data on such characters should therefore always be included when taking decisions on redundancies. On the other hand, accessions of wild material may show similarity in currently relevant characters, e.g. due to natural selection, but nevertheless differ based on molecular data. From a conservation perspective such accessions should be maintained because the observed differences may reflect variation for characters that are presently neutral but may become important in the future. Third, the majority of molecular marker techniques are still costly to perform. Obviously, removing redundancies has financial advantages because of a reduced number of accessions that need to be stored, regenerated, documented and distributed. However, from an economical viewpoint the question is how these savings relate to the costs to identify redundancies. This will depend on several factors, including the crop investigated, the molecular technique applied, the level of accuracy required and the level of redundancy encountered. Apart from data on redundancy, molecular marker studies often generate important spin off, useful to genetic resources management but difficult to express in financial terms. Theory about genebank economics is still in its infancy, but is receiving increasing attention (e.g. Swanson, 1996; Pardey et al., 2001).

At the Centre for Genetic Resources, The Netherlands, several redundancy studies using molecular markers have been carried out during the last few years in a variety of organisms. In the present paper, these studies will be outlined with special reference to the genetic and economic aspects.

CLONALLY PROPAGATED ORGANISMS

Enset (*Ensete ventricosum* Welw. (Cheesman)) has its center of origin and diversity in Ethiopia. Enset is an important multi-purpose (co)staple crop for a large part of the Ethiopian population. During the last decades the total available genetic diversity of enset has been reduced due to the fact that the local farming systems in which the crop is maintained have become endangered. Since the conservation of a clonally propagated crop, like enset, is rather complex and relatively expensive, assessment of clonal diversity was considered essential to assist a conservation programme. Morphological characterization of enset is hampered by the lack of suitable characters that can be scored, and phenotypic data provided by farmers are only limited. Furthermore, a complicating factor is that Ethiopia consists of many ethnic groups that often modify the names of clones following exchange of material. Names of clones may also be changed when clones are being used for different purposes. Altogether, germplasm diversity in enset is difficult to determine without additional research.

Therefore, an AFLP study was undertaken to investigate genetic relationships among 146 enset clones from five different regions in southern and southwestern Ethiopia. Four AFLP primer combinations were analyzed, for which a total number of 180 bands were scored. Within the total sample, 104 bands (58%) appeared to be polymorphic. The study revealed 21 duplication groups consisting of a total number of 58 clones (Negash et al., 2002).

In the case of clonally propagated organisms, where an accession is usually a single clone, the decision whether two accessions are different seems quite straightforward. If two accessions show different AFLP profiles they can be considered genetically different. However, artifact bands may be generated during the AFLP procedures that erroneously are scored as genetic variants. A possible way to check and correct for artifact bands is to analyze multiple samples from the same individual. Differences between these samples can then be used to estimate an experimental error probability. Subsequently, samples with AFLP differences higher than the error percentage can be considered genetically different, whereas lower or equal values suggest similarity (e.g. Arens et al., 1998). In the enset study, four replicate tissue samples from a single individual were included in all experimental steps in order to estimate the frequency of artifact bands. Identical profiles for these samples were however found for all four primer combinations. Therefore, enset clones with a single AFLP difference could already be considered genetically different.

Another genetic issue is how many genetic markers should be analyzed to decide whether samples are different or not. If the four primer pairs used in this study were considered, it appeared that one primer pair could distinguish, depending on the choice, between 59 and 86 clones, two could distinguish between 97 and 103 clones, four between 105 and 107 and all four could distinguish 109 clones. Thus, two primer combinations are sufficient to discriminate between the majority of clones. Analyzing a third and fourth primer combination resulted only in limited additional information. The number of primer pairs required will obviously depend on the number of bands that are generated with each primer. In the case of enset, primer pairs with a total number of 5 selective nucleotides were used in order to increase the number of bands of the AFLP profiles (e.g. Treuren, 2001). For species with large genomes or when closely related material is investigated, it may be necessary to increase the number of primer pairs in order to obtain sufficient resolving power.

In the enset study, clones were considered different when already a single AFLP difference was observed. Since the aim in a conservation program is not to conserve AFLP variation, the question is how representative the molecular data are for agro-morphological diversity. In an accompanying study it was found that agro-morphological diversity and use value of enset clones correlated positively, but weakly with the molecular data. In particular, less duplication was identified, namely 11 duplication groups consisting of 23 enset clones. However, the majority of these clones were also identified as duplicate based on the AFLP data (Tsegaye et al., submitted). It was concluded that duplications of clones identified by both methods can be safely removed from a conservation program. Furthermore, the additional clones identified as duplicate only by the molecular data probably indicate close genetic relationship. A substantial structural reduction of up to 25% of the total conservation costs could therefore be achieved.

PREDOMINANTLY SELFING ORGANISMS

Characterization of an entire lettuce collection

When dealing with sexually reproducing organisms, multiple individuals collectively constitute the genetic identity of an accession. In the case of selfing organisms these individuals may all be genotypically identical but can also be different. For instance, modern cultivars tend to be more homogeneous compared to landraces and wild material. But even in modern cultivars of selfing species heterogeneity is frequently observed (e.g. Treuren and van Hintum, 2001). As a consequence of the presence of intra-accession variation, various genetic relationships may exist among accessions, ranging from complete identity to complete distinctiveness. Redundancy can be regarded as the situation where two or more accessions share genetic variation. However, the extent to which genetic variation is shared can be quite different (Fig. 1).

In the EU-funded project, entitled “Molecular markers for genebanks: application of marker technology for the improvement of *ex situ* germplasm conservation methodology”, CGN’s entire lettuce collection of about 2300 accessions was characterized by ten microsatellites and three AFLP primer pairs. The huge data set, consisting of a few million data points is still being analyzed, but concerning the methodology a number of relevant findings can be reported. First of all, the standard software appeared incapable of analyzing the huge data set. Second, all the types of genetic relationships, as illustrated in Fig. 1, could be found among accessions of the lettuce collection. Often, the situation was observed that heterogeneous accessions shared common genotypes. Third, the concept of redundancy could even become more complex if missing values are accepted in the data set. For example, accessions A and B may be duplicate, B and C may be duplicate, but A and C may be different (Hintum et al, submitted).

Breeder’s lines in a flax collection

Because intra-accession variation occurs, estimates of this variation need to be included in decisions about redundancy. A possible criterion to decide that accessions are distinct is when the variation among accessions is significantly higher than the variation

within accessions. In the absence of significant differentiation among accessions, they can be considered as drawn from the same population, and hence be regarded redundant. Combining such material into a single accession is one of the possibilities to reduce the amount of redundancy in a collection, and to improve the composition thereof. Variance components can be estimated from molecular data by an Analysis of Molecular Variance (AMOVA) that subsequently can be tested for significance (Excoffier et al., 1992). This approach was followed in a small-scale study in a set of breeder's lines of CGN's flax collection (Treuren et al., 2001).

The flax collection of CGN currently comprises about 1000 accessions. About 30% of this collection consists of material derived from research activities that has been classified as breeders' lines. For this material only limited passport data are available. Individual accessions can be distinguished based on a number, but can also be clustered into different groups based on a common group code that is part of the designation of the lines (e.g. M25-64, M25-221, M25-245, M25-330; 75Ru, 76Ru, 341Ru, 342Ru; 346Rm, 347Rm, 348Rm, 386Rm). These common group code might indicate a common genetic background, and suggested redundancies. Three of such groups, with a total number of 29 accessions were analyzed with two AFLP primer combinations on four individuals per accession. Twelve cultivars were included in the study as reference. Within the total sample, 144 polymorphic bands (59.8%) were scored. Contrary to expectation based on the predominant selfing habit of flax, high levels of intra-accession variation were observed for the majority of accessions, even for the cultivars investigated. To determine which accessions of breeder's lines could be bulked a procedure termed "stepwise bulking" was followed. First, a genetic distance matrix with corresponding significance values was generated for all pairs of accessions using AMOVA. Second, the pair with the smallest genetic distance was combined into a single accession. Third, a new AMOVA was carried out, followed by repeating the first two steps. The whole procedure was repeated until all elements of the matrix were significantly different from each other. This reduced the initial number of 29 accessions to 14, in other words a reduction of 52%. Interestingly, the bulking process was accompanied by only a 2.6% reduction in the among-accession component of variance (Table 1). It was therefore concluded that considerable reduction could be realized among the investigated material, while ensuring minimum loss of genetic variation.

Extrapolation of the results to the entire group of 317 flax breeder's lines would result in a reduction to 153 accessions. It was estimated that following rationalization about k€ 41 could be saved per generation, whereas the estimated costs to identify redundancies by AFLP fingerprinting were of the same order of magnitude (estimates based on rates from 1999). Return of investments can therefore be expected within one generation of flax conservation. It should be noted, however, that in this case study only four individuals per accession were used. However, the unexpected high levels of intra-accession variation observed suggests that sample sizes need to be increased to estimate variance components more accurately. This will of course increase the costs.

Wild potato species of the series *Acaulia*

If heterogeneity in accessions can be expected and large numbers of accessions need to be screened, a redundancy study may involve the analysis of huge numbers of samples. Consequently, the cost factor of identifying redundancies will be high. A possibility to reduce the number of samples to be analyzed is performing a pre-screening on a small number of samples per accession, followed by more extensive investigations on a subset of the accessions. This approach was followed for CGN's collection of wild potato species of the series *Acaulia*.

This collection currently consists of 314 accessions, comprising seed samples collected from South America. The majority of the accessions are well-documented, including exact data on collection sites. From these data it appeared that material stored as different accessions were sometimes collected from identical or nearby sites, suggesting redundancies. However, conservation of potato germplasm is costly because of the

expensive virus tests that are required prior to regeneration. Therefore, a redundancy study was undertaken directed to the optimization of the collection.

Two individuals of each of the 314 accessions were analyzed with two AFLP primer combinations, revealing 130 polymorphic bands. For 126 accessions it was found that the two individuals analyzed displayed identical AFLP profiles. For forty six of these 126 accessions the observed AFLP profile appeared identical to that of either a single plant or both plants from another accession. Subsequently, all of these samples were tested for a third primer pair to increase resolution. Based on the three primer pairs studied, accessions were considered potentially redundant when they had at least one genotype in common. This resulted in 15 potential redundancy groups collectively comprising 36 accessions (McGregor et al., 2002).

These groups, including one additional accession, were further analyzed in a follow-up study (Treuren et al., submitted). Because of the higher incidence of redundant germplasm observed across smaller geographic distances in the initial study, groups consisting of accessions that were collected less than 10 km apart were analyzed with five plants per accession, whereas for larger geographic distances sample sizes of 20 were used. Obviously, larger sample sizes allow the detection of less frequent variation, but this was considered irrelevant given the number of 20 plants that are used in standard regenerations of *Acaulia* germplasm. The total sample of 499 plants was analyzed with two AFLP primer pairs, revealing 137 bands of which 82 appeared polymorphic. For two potential duplication groups, all individuals displayed identical AFLP profiles. Within seven potential duplication groups it was found that the constituting accessions shared the major part of the genetic variation, whereas large genetic differentiation was observed among the accessions of six groups (Table 2). Apart from the accessions of the latter six groups, all accessions clustered tightly together within their potential duplication group, separately from all the other accessions of the collection, when the data were integrated with those of the initial study of McGregor et al. (2002). It was therefore concluded that 15 accessions could be considered redundant, involving about 5% of the total collection (Treuren et al., submitted).

It seems likely that given the low sample size in the initial screening, the true level of redundancy has been underestimated. However, by focusing on accessions containing individuals that share identical genotypes, the probability of identifying true redundancies was optimized. Two types of errors can be made in redundancy studies. Accessions may erroneously be considered redundant (false positives), while redundant accessions may remain undetected (false negatives). The avoidance of false positives is most relevant in germplasm conservation. Therefore, this approach was followed for the *Acaulia* collection, thereby accepting a higher level of false negatives. Pre-screening of limited sample sizes, followed by more extensive analyses of accessions that have a high probability of being redundant is a possibility to reduce the necessary investments.

The total costs for the molecular study, including all labor costs, consumables and overheads were estimated at k€ 57.3. On the other hand, the costs to maintain 15 accessions during one generation cycle, including all labor costs, quarantine testing, germination testing, consumables and overheads were estimated at k€ 21.0. Thus, a simple cost-benefit analysis of the *Acaulia* study shows that the invested costs to identify redundancies are 2.7 times as high as the costs to maintain 15 accessions for one generation (Treuren et al., submitted). It was therefore concluded that, although the *Acaulia* collection is expensive to maintain, the necessary investments are not covered by the short-term benefits of a reduced collection. Return of investments may only be expected after three cycles of regeneration. The high costs of AFLP analyses are predominantly due to the high labor costs involved. These may be different in other countries. Reducing labor costs could also be achieved by hosting trainees and visiting scientists, the development of collections of extracted DNA samples, and a higher degree of automation in the future. Further optimization of the sampling strategy could also be achieved, e.g. by focusing on probable redundancies based on passport or evaluation data.

This would reduce the number of accessions to be sampled with a limited increase of the number of false negatives.

OUTCROSSING ORGANISMS

Although accessions of predominantly selfing organisms may show heterogeneity, the major part of the variation is distributed between accessions, rather than within accessions. The opposite is generally found for outcrossing organisms, often showing high levels of within-accession variation compared to diversity among accessions. Ideally, a germplasm collection consists of accessions for which the major part of the diversity is distributed among accessions, and not within. The question therefore is how many accessions to maintain, while ensuring coverage of the major part of the variation. Specifically for outcrossing species, a related question is to what extent accessions may change in time genetically. Regeneration exposes heterogeneous accessions to the influence of genetic drift and selection, resulting in changes in genetic composition. The extent to which this occurs will depend on the specific regeneration procedures, including effective population sizes used and the methods of pollination and harvesting seeds. In a redundancy study, evaluating genetic differences among accessions can therefore not be disentangled from genetic changes in time.

CGN's *Brassica oleracea* collection includes, among others, a group of white cabbage cultivars. In the past, 93 original samples of this group have been reduced to 20 accessions (78% reduction) based on data about history, morphology, and the knowledge of crop experts. The bulking process was subsequently verified by an allozyme study that in addition suggested that the groups could have been even larger (Hintum et al., 1996). In a follow-up study, including a number of these accessions, 30 plants per white cabbage accession were investigated by two AFLP primer combinations (103 polymorphic bands). In addition, six accessions were compared genetically before and after a standard regeneration using 30 plants. It appeared that the genetic changes following regeneration were of the same order of magnitude as the smallest genetic differences among accessions within some of the groups. These data indicated that a single regeneration cycle may result in genetic differences that are comparable to those found among accessions within groups (Hintum et al., submitted). This finding corroborated the suggestion that some of the redundancy groups established earlier could indeed have been much larger.

DISCUSSION

The studies presented in this paper showed that marker-assisted reduction of redundancy in germplasm collections is not as straightforward as it may seem. Because genetic diversity may vary both in space and in time, several genetic issues need to be considered. Because the majority of molecular markers are still costly to perform also economical considerations play a role.

Genetic aspects

With respect to the genetic issues, the most important question to be answered is how reliable the decisions are that are taken in redundancy studies. Many factors can be expected to affect this reliability, including the resolution of the marker technique applied, the number of polymorphic markers studied, and the sample size used for the analysis. Obviously, the higher the number of markers and the larger the number of individuals, the higher the probability to detect variation. The question of course is how far one should go. What are needed are estimates about the probability of making the incorrect decision, given parameters on e.g. numbers of markers and sample sizes. In practice, these parameters are generally based on practical and financial limitations and on common sense. Nevertheless, development of statistical theory is needed to obtain insight in the reliability of decisions about redundancy. These should preferably include data on the distribution of genetic variation within the collection and data on the genetic changes that may be expected over time. A complicating factor thereby is that germplasm may be regarded differently by conservationists and users. For example, plant breeders

tend to focus more on specific characters that are presently relevant, while conservationists also need to consider the variation that may be of limited current use but has potential relevance in the future. Redundancy may therefore have a different meaning, depending whether it is approached from a user or conservation point of view.

Economic aspects

Redundancy studies are not only performed to improve the composition of the collection, but also to reduce the costs of conservation. Maintenance of redundant germplasm wastes money that can also be spent on other genebank operations or on novel accessions. The question of course is whether it is profitable to invest in the identification of redundancies. This will depend on the crop involved and on the level of redundancy expected. Crops that can be simply multiplied and stored will require minimum costs for maintenance. Investments to identify redundancies will therefore rarely outweigh the costs of maintaining the material. Economic theory about genetic resources conservation is still developing. It addresses issues such as the costs of different genebank operations and the optimal allocation to achieve the goals of a genebank (Pardey et al., 2001), but also focuses on the value of conserved material and the information thereof (Swanson, 1996). In the simple economic analyses presented in the present paper, the costs of the molecular analyses were compared to the short-term benefits of a reduced collection. However, the long-term benefits were not considered, neither the relevant spin-off that was generated in the studies. The latter are generally difficult to express in financial terms.

Spin-off

In the redundancy study on flax breeders' lines, the major part of the variation appeared to be distributed within accessions, rather than among accessions. This finding was difficult to reconcile with the predominant selfing nature of flax, and suggested that flax is more outcrossing than generally assumed. This would have implications for the regeneration of flax accessions that at CGN are performed without any precautions to avoid outcrossing between accessions. Therefore, the redundancy study in flax also resulted in recommendations to improve the regeneration protocols of flax by taking safety measures to avoid contamination (Treuren et al., 2001).

CGN's wild potato germplasm collection of the series *Acaulia* consists of the species *Solanum acaule* and *Solanum albicans*. Within *S. acaule* four subspecies can be distinguished, namely ssp. *acaule*, ssp. *aemulans*, ssp. *palmirense* and ssp. *punae*. In the redundancy study in the *Acaulia* collection, the AFLP data appeared very useful for taxonomic purposes. Within the total sample, 16 misclassifications were identified, including four cases that did not belong to the series *Acaulia*. In addition, the taxonomic classification could be elucidated of 97 *S. acaule* accessions for which the subspecies was unknown prior to the study. Furthermore, it appeared that accessions could consist of plants belonging to different taxonomic levels (McGregor et al., 2002). The AFLP data of *S. acaule* ssp. *acaule* in that study were also related to the geographic origin of the accessions. These analyses revealed that sampling sites that are less than 20 km apart increases the probability to collect identical or very similar genotypes. This finding was suggested as a general guideline to maximize the diversity sampled in future collection missions for this species. The follow-up study in the series *Acaulia* (Treuren et al., submitted) was carried out using increased sample sizes per accession, and therefore revealed more accurate data about intra-accession variation. For several accessions, none or only very limited intra-accession variation was observed. It was suggested to reduce the number of plants used for regeneration of such homogeneous accessions. Homogeneity of accessions allows the number of regenerated plants to be based on the number of seeds required for storage and distribution, rather than on minimizing effects on the genetic integrity of accessions. Accessions could also consist of distinct genotypes. In those cases, depending on the frequency of the genotypes, splitting up accessions was suggested as an option to avoid the loss of diversity. In summary, from the redundancy

studies in the series *Acaulia* also data were obtained to reveal documentation errors, to improve the quality of passport data of accessions, to derive guidelines to optimize sampling strategies, and to optimize regeneration procedures.

The redundancy study in white cabbage included the comparison of the genetic constitution of accessions before and after a standard regeneration (Hintum et al., submitted). An interesting result was that sometimes highly significant shifts in allele frequency could be observed at AFLP loci. Since these could not be attributed to methodological inconsistencies, and random effects were unlikely to underlie this finding, it was suggested that plants might be subjected to selection during regeneration. These shifts could for example be explained by variation in flowering time when some AFLP loci are linked to genes affecting this trait. In that case, single plant harvesting could be employed to improve the regeneration protocol.

In general, marker-assisted redundancy studies generate data about genetic relationships among accessions that subsequently can be used to structure a collection. For example, marker data can be used to create core collections, i.e. a subset of the collection that contains the major part of the variation (Hintum et al., 2000). Also in the formation of core selections (Hintum, 1999) marker data are useful, e.g. to prioritize between accessions that cannot be distinguished based on passport or evaluation data. Structuring a collection will be relevant to genebanks as well as to users of the germplasm.

CONCLUSIONS

Molecular marker techniques are useful tools to assist in the reduction of redundancies from germplasm collections. However, further development of statistical and economical theory is still needed to draw conclusions concerning the extent of redundancy and the feasibility of tracing and reducing it. Furthermore, the added value obtained by molecular characterization in improving insight in the genetic composition of collections and the quality of genebank operations can be regarded considerable and should be included in the 'feasibility equations'.

Literature cited

Arens, P., Coops, H., Jansen, J. and Vosman, B. 1998. Molecular genetic analysis of black poplar (*Populus nigra* L.) along Dutch rivers. Mol. Ecol. 7:11-18.

Bretting, P.K. and Widrlechner, M.P. 1995. Genetic markers and plant genetic resource management. Plant. Breed. Rev. 31:11-86.

Brown, S.M. and Kresovich, S. 1996. Molecular characterization for plant genetic resources conservation. p.85-93. In: H. Paterson (ed.), Genome Mapping of Plants, Academic Press, San Diego.

Cervera, M.T., Cabezas, J.A., Sancha, J.C., Martínez de Toda, F. and Martínez-Zapater, J.M. 1998. Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theor. Appl. Genet. 97:51-59.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.

FAO, 1996. FAO State of the World's Plant Genetic Resources for Food and Agriculture. Food and Agriculture Organisation of the United Nations, Rome.

Hintum, Th.J.L. van 1999. The core selector, a system to generate representative selections of germplasm accessions. Plant. Genet. Res. Newsl. 118:64-67.

Hintum, Th.J.L. van 2000. Duplication within and between germplasm collections. III. A quantitative model. Genet. Res. Crop Evol. 47:507-513.

Hintum, Th.J.L. van, Boukema, I.W. and Visser, D.L. 1996. Reduction of duplication in a *Brassica oleracea* germplasm collection. Genet. Res. Crop Evol. 43:343-349.

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

Hintum, Th.J.L. van and Knüpffer, H. 1995. Duplication within and between germplasm collections. I. Identification of duplication on the basis of passport data. Genet. Res. Crop Evol. 42:127-133.

Hintum, Th.J.L. van, van de Wiel, C.C.M., Visser, D.L., van Treuren, R. and Vosman, B. The distribution of genetic diversity in a *Brassica oleracea* genebank collection related to the effects of regeneration, as measured with AFLPs. Theor. Appl. Genet. (submitted).

Hintum, Th.J.L. van, Verbakel, H., Boukema, I.W., de Groot, E.C. and Peleman J. Wild germplasm in genebank collections: Molecular genetic diversity of wild *Lactuca* in the CGN collection. Theor. Appl. Genet. (submitted).

McGregor, C.E., van Treuren, R., Hoekstra, R. and van Hintum, Th.J.L. 2002. Analysis of the wild potato germplasm of the series *Acaulia* with AFLPs: implications for *ex situ* conservation. Theor. Appl. Genet. 104:146-156.

Negash, A., Tsegaye, A. van Treuren, R. and Visser, L. 2002. AFLP analysis of enset clonal diversity in South and Southwestern Ethiopia for conservation. Crop Sci. 42: 1105-1111.

Oliveira, J.A., Lindner, R., Bregu, R., García, A. and González, A. 1997. Genetic diversity of westerwold ryegrass landraces collected in Northwest Spain. Genet. Res. Crop. Evol. 44:479-487.

Pardey, P.G., Koo, B., Wright, B.D., van Dusen, M.E., Skovmand, B. and Taba, S. 2001. Costing the conservation of genetic resources: CIMMYT's *ex situ* maize and wheat collection. Crop Sci. 41:1286-1299.

Phippen, W.B., Kresovich, S., Candelas, F.G. and McFerson, J.R. 1997. Molecular characterization can quantify and partition variation among genebank holdings: a case study with phenotypically similar accessions of Brassica *oleracea* var. *capitata* L. (cabbage) 'Golden Acre'. Theor. Appl. Genet. 94:227-234.

Swanson, T., 1996. Global values of biological diversity: the public interest in the conservation of plant genetic resources for agriculture. Plant Genet. Res. Newsl. 105:1-7.

Treuren, R. van, 2001. Efficiency of reduced primer selectivity and bulked DNA analysis for the rapid detection of AFLP polymorphisms in a range of crop species. Euphytica 117:27-37.

Treuren, R. van and van Hintum, Th.J.L. 2001. Identification of intra-accession genetic diversity in selfing crops using AFLP markers: implications for collection management. Genet. Res. Crop. Evol. 48:287-295.

Treuren, R. van, Magda, A., Hoekstra, R. and van Hintum, Th.J.L. Genetic and economic aspects of marker-assisted reduction of redundancy from a wild potato germplasm collection. Genet. Res. Crop. Evol. (submitted).

Treuren, R. van, van Soest, L.J.M. and van Hintum, Th.J.L. 2001. Marker-assisted rationalisation of genetic resources collections: a case study in flax using AFLPs. Theor. Appl. Genet. 103:144-152.

Tsegaye, A., Negash, A., van Treuren, R. and Struik, P.C. Comparison of enset characterization based on farmers' knowledge and on AFLPs for effective conservation and utilization of genetic resources in Ethiopia. Crop Sci. (submitted).

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

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

Zeven, A.C., Dehmer, K.J., Gladis, T., Hammer, K. and Lux, H. 1998. Are the duplicates of perennial kale (*Brassica oleracea* L. var.: *ramosa* DC.) true duplicates as determined by RAPD analysis? Genet. Res. Crop. Evol. 45:105-111.

Tables

Table 1. Effects of stepwise bulking of accessions on molecular variance components using AMOVA on three series of flax breeder's lines. (data from Treuren et al., 2001)

	Before bulking	After bulking
Number of accessions	29	14
Variation among accessions	26.4%	23.8%
Variation within accessions	73.6%	76.2%

Table 2. Genetic relationships among accessions of 15 potential duplication groups within a wild species germplasm collection of the series *Acaulia*. Potential duplication groups were constituted based on AFLP analysis of two plants per accession in an initial study (McGregor et al., 2002), and verified using increased sample sizes in a follow-up study (data from Treuren et al., submitted).

	No. of potential duplication groups	No. of accessions involved
No variation	2	5
Limited differentiation	7	19
Large differentiation	6	13
Total	15	37

Figures

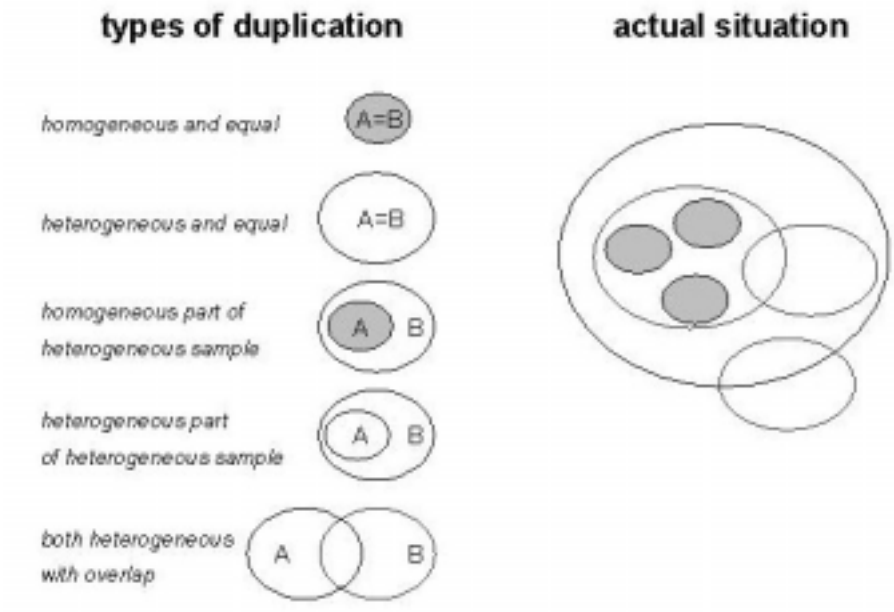

Fig. 1. Overview of the alternative types of redundancy in which two accessions share genetic variation to different extents.