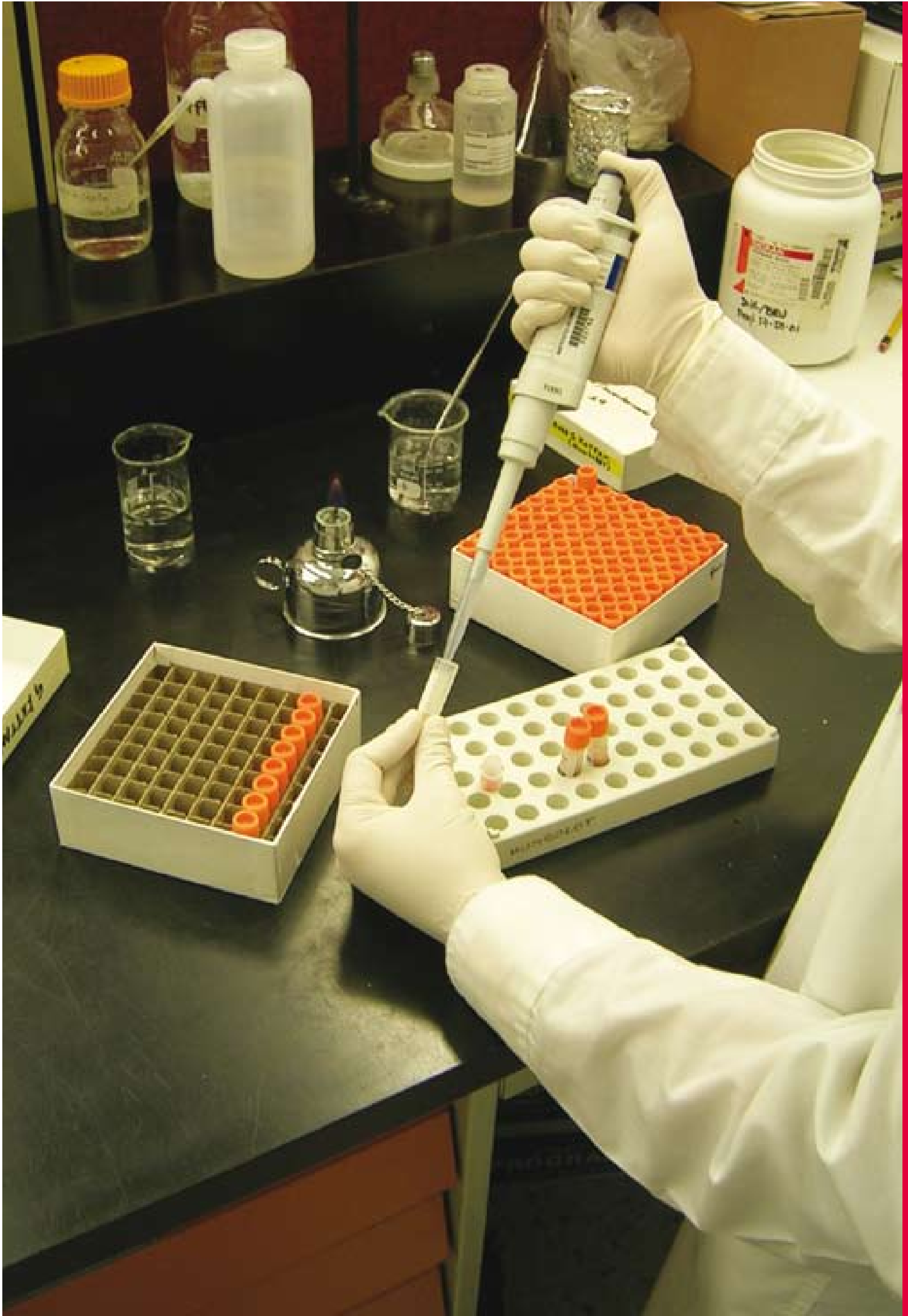
TOPICAL REVIEWS IN AGRICULTURAL BIODIVERSITY

IPGRI

DNA banks— providing novel options for genebanks?

M.C. de Vicente and M.S. Andersson (editors)



SGRP

DNA banks— providing novel options for genebanks?

M.C. de Vicente and M.S. Andersson (editors)

Contents

IX. Opportunities, limitations and needs for DNA banks

Andreas W. Ebert[1]*, Jawahir L. Karihaloo*[2] *and Márcio Elías Ferreira*[3]

[1]Tropical Agricultural Research and Higher Education Centreer (CATIE), Turrialba, Costa Rica
[2]National Research Centre on DNA Fingerprinting, NBPGR, New Delhi, India
[3]EMBRAPA Genetic Resources and Biotechnology, Brasília, Brazil

Concern over the dramatic loss of agrobiodiversity in farmers' fields and in nature has led to efforts to conserve plant germplasm in several *ex situ* collections maintained by national and international research centres throughout the world. Apart from conservation, routine operations of genebanks include germplasm collection and acquisition (enrichment), characterization, evaluation, documentation and supply of germplasm to users and other genebanks. Many national and regional genebank collections, however, have deficits in characterization and evaluation data of the accessions. This lack of information drastically reduces the possibility of utilizing the conserved germplasm, by both plant breeders and farmers alike. It also limits the correct judgement of the representativeness of a genebank. Characterization and, where possible, evaluation of accessions is, therefore, central to strategic germplasm conservation. DNA banks might represent an option to speed up the characterization, evaluation and utilization of genetic resources. In fact, molecular characterization programmes pursued by genebanks could become catalysts for the creation of DNA banks since these routinely generate excess DNA samples that are often stored for future reference and research.

Opportunities for DNA storage in genebanks, especially in connection with molecular characterization, are becoming more and more apparent. There are a few areas where one could already foresee an impact of DNA banks in the next few years, including its use as: 1) a resource for high throughput germplasm characterization and improved genebank management; 2) a substrate for association genetics and marker-assisted selection (MAS); 3) a promoter of germplasm information exchange, including novel services such as exchanging DNA samples and sequence information; 4) a reference basis for evolutionary and comparative genomic studies; and 5) a complementary preservation procedure aiming at gene and genome conservation. Some of these opportunities are discussed below.

Opportunities for DNA storage in genebanks

High throughput germplasm characterization and improved genebank management

Molecular or genetic markers are seen as descriptors that offer reproducible complementary information to the classical morphological descriptors and phenotypic data used in the characterization and agronomic evaluation of genebank accessions (de Vicente et al. 2004). Detailed genetic information of the accessions in a given collection improves genebank management in several aspects as described by Karp et al. (1997): 1) it allows for the detection of gaps in the collection, guiding new collection missions and the exchange of germplasm; 2) it provides valuable knowledge concerning molecular diversity and genetic relationships through systematic genetic fingerprinting, within and between genepools; 3) it allows for the identification of unique genotypes of special importance to the genebank, as well as the identification of duplicate accessions by genetic fingerprinting (Duplicate accessions can be bulked to prevent loss of alleles [Sackville-Hamilton et al. 2002]); 4) it also allows for monitoring the genetic stability and integrity of the accessions, detecting genetic drift, natural selection and contamination during regeneration cycles; and 5) it benefits the potential users, allowing them to identify valuable traits and types quickly based on the genetic information on the accessions.

Large genebank collections are often poorly characterized, making it difficult for plant breeders to access and exploit such collections. Using marker technologies, the genetic diversity in such collections can be systematically assessed. This information is then used to establish core collections (Hodgkin et al. 1995). These core collections or subsets of large collections contain a limited number of accessions that capture most of the genetic variability present in the entire collection. These core subsets facilitate the management of the collection and contribute to an increased utilization of the germplasm. DNA banks as hubs of molecular marker application could, therefore, play a prominent role in germplasm characterization, facilitating genebank management and germplasm use.

Association genetics and marker-assisted selection

The application of molecular tools to identify genes controlling specific traits in accessions of cultivated species and their wild relatives constitutes an important new role of genebanks to enhance germplasm utilization. The introgression of identified genes into genotypes with a more desirable genetic background using marker-assisted selection (MAS) has been a component of this approach (Tanksley and Nelson 1996; Xiao et al. 1996; Ortiz and Engels 2004). The MAS scheme discussed by Causse et al. (2001) for the transfer of the five most important quantitative trait loci (QTLs) involved in the organoleptic fruit quality of tomato is an example of such an approach.

The analysis of genetic variation in germplasm collections and the proper documentation of the number and types of useful polymorphisms offer genebank curators the opportunity to estimate the value of the conserved accessions and enable them to offer specific accessions with desired characteristics to plant breeders and users in general. These users can then make an informed choice and select only those genotypes that best fit their objectives (Ortiz and Engels 2004). In a paper discussing the evolving role of genebanks, de Vicente (2004) stressed the increasing demand for specific genes and alleles at QTLs.

The rate of discovery of nucleotide variation at QTLs contributing to phenotypic variation of complex traits is expected to increase with the adoption of linkage disequilibrium and candidate gene strategies for fine mapping and cloning of QTLs (Rafalski 2002; Morgante and Salamini 2003). This approach would eliminate the requirement for structured segregating populations and genetic studies could be directly performed on the accessions deposited in the genebank (Graner et al. 2004). The feasibility of association mapping between DNA markers and agronomic traits has been successfully demonstrated in a genebank collection of 600 potato varieties. Highly significant association with QTLs for resistance to late blight and plant maturity was detected with PCR markers, specific to a major gene for resistance to late blight (Gebhardt et al. 2004). Thus, in the near future, genebanks might be asked to provide not only seed, but also DNA samples and the corresponding information on both marker and sequence data.

Exchange of genetic information and DNA samples

The exchange of DNA samples will certainly facilitate genetic or genomic studies on accessions in a given genebank. It will be a lot easier to exchange DNA samples, rather than seed or vegetative propagules. Transboundary movement of seed and other plant material requires seed inspection, phytosanitary certificates and quarantine testing to ensure it is free of undesirable diseases and pests. Exchanging DNA samples instead of seeds avoids the need for these time-consuming and costly procedures. The risk of pathogen contamination will be simply circumvented. Moreover, transportation costs of DNA samples would be much lower than that of seed or vegetative material.

The transfer of genetic material in the form of DNA samples would, therefore, be preferable in programmes focusing on genetic and genomic studies, rather than agronomic performance. Upon the conclusion of a study based on the DNA samples and having identified a specific germplasm accession possessing a gene of interest, the user could request only that accession, thereby reducing risk, cost and time.

Reference basis for evolutionary and comparative genomic studies

DNA markers can be used for the taxonomic determination of plant genetic resources for any given genus to at least the species level. In some cases, the classical taxonomy had to be revised based on new DNA and sequence data (Graner et al. 2004). Aligned nucleotide sequences can be used to make inferences about the ancestral relations between them in molecular phylogenetics (Hartl 2000). These studies shed light on patterns of species evolution at the molecular level due to natural and artifical selection. In studies with cultivated wheat using microsatellites, Khlestkina et al. (2004) demonstrated that modern wheat breeding caused a qualitative shift in genetic diversity over the past 50 years, rather than a quantitative one. They concluded that it is necessary to maintain the existing *ex situ* collections and to collect new material, in order to exploit the whole range of allelic variation.

Genetic diversity can be utilized beyond the species-boundary of the primary genepool using genomic approaches. The development of new technologies, such as microarrays and libraries of expressed sequence tags (ESTs) facilitate the detection of genes of special interest within germplasm collections (Richards 2004). The information available on one intensively studied species or crop can be used to predict and locate genes in a completely different species. The application of comparative genomics in cereal crops of the *Poaceae* family has shown that there is a considerable conservation in gene content and gene order over 60 million years of evolution in species such as wheat, maize and rice (Richards 2004). This homeology suggests that the available sequence and linkage data in one species can be useful for mapping orthologous genes in other species. Comparative genomics also offers the possibility of exploiting the potential of wild relatives of crops in genebanks more efficiently.

DNA banks as a complementary conservation strategy

DNA banking could constitute a complementary conservation strategy for safeguarding the genetic diversity of a crop's genepool, especially if combined with *in vitro* conservation or cryopreservation. DNA banks can also serve as backup or safety duplicates of the physical seed, field or *in vitro* collections, in case of catastrophic losses. Although it is not (yet) possible to recover a plant from a DNA sample, the storage of entire genomes (total DNA) or genome fragments (genomic libraries) would permit the preservation of its valuable genetic information, thus, contributing to the objective of gene or genome conservation (Andersson 2004). With the impressive advances in molecular genetics, these preserved genes or genomes might be of high relevance in the future.

Genome conservation could play a major role for species that are currently under threat of extinction. This applies in particular to those in densely populated tropical regions that are under severe threat. Of the approximately 17 000 vascular plant species reported to exist in India, nearly 15% are under threat (Ahmedullah 1999). This is aggravated by the high level of endemism existing in several ecologically vulnerable regions of the world. In Sikkim state of India, 60% of the plant species are endemic (Myers 1988), while 33% of the total Indian flora is endemic to the country (TERI 2004). In Brazil, only 8% of the original Atlantic Forest, a highly biodiverse tropical biome that covered the coast from north to south, has been preserved. This area has been subjected to intense human pressure for the last 500 years. Central America and Mexico together form one of the Vavilov centres of origin of cultivated crops. According to Zeven and de Wet (1982), 225 domesticated plant species have their origin in this centre of diversity, representing roughly 9% of the total number of 2489 domesticated species worldwide. Alarmingly high deforestation rates in the range of 2.1% per year have been reported for Central America by FAO (1993). This single factor presents a major threat to the wealth of economically important species in this region. It is unlikely that the current *in situ* or *ex situ* conservation efforts will be adequate to guarantee the survival of all these vulnerable species. DNA conservation, on the other hand, at least offers the possibility of ensuring the availability of the genome long after the plant has become extinct.

Limitations for DNA storage in genebanks

Plant recovery from stored DNA

A major limitation of DNA banking is the fact that technologies to regenerate plants from stored DNA are not (yet) available. Hence, DNA conservation cannot be considered as a substitute for conventional conservation strategies, but can only be seen as a complementary strategy. In general, the inability to recover living plants from stored DNA may discourage the curators of genetic resources to support investment in DNA conservation. Furthermore, if the national programmes do not envision comprehensive DNA-based research, justification of DNA conservation would not be apparent.

Plant DNA extraction procedures

Several plant species are recalcitrant to the commonly used DNA extraction protocols. Those with high concentrations of polysaccharides, proteins, tannin and lipids need special treatment. Thus, there is a need to refine and improve existing protocols to overcome these limitations. It is possible that general extraction protocols applied to specific taxonomic groups would be refined to serve as a reference to groups of species.

Life-span of stored DNA

Another limitation of DNA banking is the relatively short life-span of stored DNA, making it necessary to replace the DNA at frequent intervals. However, the life-span can be easily extended if various factors are optimized, such as extraction and purification procedures and the management during storage. A life-span of 9 years has been observed with DNA extracted from coffee germplasm at the Tropical Agricultural Research and Higher Education Centre (CATIE) in Costa Rica (C. Astorga, personal communication). The Royal Botanical Gardens, Kew, which has in storage over 13 000 samples of DNA from a diverse range of plant species, has been largely successful in terms of long-term storage (Royal Botanical Gardens, Kew 2004). Their extraction protocols and storage procedures could serve as a basis for the standardization and improvement of applied protocols.

Long-term conservation could also be achieved using a solid medium, such as paper, instead of a solution for DNA storage. The use of cellulose-based cards has been an efficient method of long-term storage of human blood cells. Its use for the conservation of plant DNA is also an efficient means of inactivating pathogens and protecting plant DNA from degradation. DNA can be stored directly in the treated paper after plant tissue disruption and transference of the substrate to the paper. It could also be stored as extracted DNA, after submitting a plant tissue to an extraction protocol and transferring the nucleic acid to the paper. The DNA maintained in a conservation paper can be stored at room temperature in a silica-gel-dried container. Identification is facilitated with the use of bar-coded tags that allows for a complete recovery of the sample information. DNA stored in paper would greatly facilitate sample exchange among institutions.

Resource and policy constraints

Besides the above-mentioned technical limitations, there are a number of cost and policy considerations that might limit the application of DNA storage as a genebanking option in some countries. Molecular marker techniques, and also DNA extraction per se, are generally quite expensive, especially in developing countries that depend on imported reagents and materials. Import prices can easily double for DNA extraction kits, enzymes and genetic markers, if all additional costs, such as customs handling and storage fees of goods are considered. In the case of importation of dangerous laboratory reagents to Central American countries, such as acetic acid, ethanol, chloroform, ethidium bromide and silver nitrate, additional costs of US$ 100 per product are levied by the airline carriers in the context of the new Bio-terrorism Act implemented by the United States (C. Astorga, personal communication). Additional problems are the bureaucratic hurdles to get imported perishable items such as enzymes out of customs without interruption of the cooling chain. In general, the costs of establishing a DNA laboratory and operating it may be quite prohibitive for many small, resource-poor developing countries. Though the consumable costs have been declining lately, DNA conservation may not be affordable for most national genebanks located in the tropics.

The availability of liquid nitrogen might also be a limitation in some countries and an important additional cost factor. Manufacturers of liquid nitrogen are often based in major cities, thus requiring transportation over long distances, if the genebank or laboratory is located in a remote place. Uninterrupted power supply is another major concern for many developing countries in Africa, Asia and Latin America, because the provision of energy often does not keep pace with the steadily increasing consumption due to rapid economic and population growth. To guarantee the safety of the DNA samples, it will be necessary to install backup power supply sources. A further frequent limitation in several developing countries is the unavailability or restricted availability of molecular geneticists to genebank curators and/or the lack of connection between them (Hamon et al. 2004). Staff members in genebanks are usually specialized in phenotypic characterization of accessions and have difficulties in adopting new genomic technologies due to a lack of training and, sometimes, lack of vision. Limited access to relevant information may also prevent global application of DNA conservation technology on a sustainable basis.

It is obvious that the use of marker technologies in genebank management requires significant additional funding. Many genebanks in developing countries are struggling to survive, have insufficient human and financial resources at their disposal to provide adequate germplasm management and simply cannot afford to invest in these new technologies. This might deepen the gap between rich and poor genebanks (Graner et al. 2004). It has to be stressed that the conservation of genetic resources and their proper management continues to be a very important task of genebanks, not only for the provision of basic material for the work of geneticists but also for the provision of phenotypic data, which are essential for the utilization of genetic information in breeding programmes. It is important to stimulate and assure the existence of genebanks in diverse environments and countries for the sake of plant diversity conservation and for the good of humanity.

Intellectual property and legal issues

The exchange of seed and other planting materials usually requires the signature of a material transfer agreement (MTA) by the beneficiary prior to the shipment. The MTAs regulate the intellectual property rights (IPR) of the requested material and related information, the conditions of its use and distribution to third parties, as well as benefit sharing. The MTAs usually ask for the submission of information and research results obtained with the genetic material to the dispatching genebank. In most cases, the MTAs are specifically designed for the exchange of seed or vegetative propagules and do not consider IPR issues in the event that DNA samples are exchanged. According to Andersson (2004), the following institutions make explicit reference to the exchange of DNA in their MTA: CATIE, Costa Rica; the National Institute of Agrobiological Sciences (NIAS), Japan, the Missouri Botanical Garden, USA; and the Royal Botanic Gardens, Kew, UK. However, even if the MTA covers the exchange of DNA samples, there are still different interpretations concerning the question whether this precludes the patenting of specific genes or not. Therefore, legal issues related to DNA exchange and transfer require immediate attention.

There are concerns in the developing world about the protection of its intellectual property rights on genetic resources accessed by the developed countries. Despite the provision of benefit-sharing in the Convention on Biological Diversity and the International Treaty on Plant Genetic Resources for Food and Agriculture, there are only a few examples of this having been effected. It remains to be seen how effective the MTAs are in enabling benefit-sharing between the genetic resources donor countries and the agencies and companies utilizing them for commercial benefits.

Infrastructure and capacity-building needs for DNA storage in genebanks

Basic laboratory infrastructure

A well-lit laboratory with working tables, storage space, running water and assured power supply is required for establishing a DNA bank. For DNA extraction, purification and storage, laboratory equipment should include a fume hood, laminar flow, reverse osmosis-based water purification system, hot air oven, microwave oven, horizontal gel electrophoresis unit with power supply, UV transilluminator, UV spectrophotometer, liquid nitrogen container, refrigerator, -20 °C freezer, -80 °C deep freezer and an autoclave. It is advisable to have access to a herbarium for the storage of reference samples with passport information and, if feasible, for replenishment of DNA stocks.

For the integration of fingerprinting data on accessions into the corresponding genebank documentation system and for the exchange of compatible data sets in networks, a computer with a printer and Internet connection and a skilled database administrator are absolute musts. There is, of course, also a need for technical staff trained in DNA extraction and purification and in preparation of herbarium samples.

Capacity building

Well-trained personnel in molecular biology and computer applications are necessary to organize and run the DNA bank. Comprehensive short-term (4–6 months) training workshops in well-known laboratories or genebanks applying long-term conservation of DNA and modern genebank management would fill this gap. The training content would comprise principles and applications of molecular biology, modern genetic resources management, genomic sciences, bioinformatics, experimental design, data analysis, interpretation and utilization, and intellectual property rights of DNA exchange. Educational and training materials could also be made available through the Internet, although this would not be a substitute for practical experience and laboratory training. A Web-based newsletter covering information on current developments in DNA technologies would be very useful.

While some globally reputed laboratories involved in DNA studies would be centres of choice for such training, it would be desirable to develop strategically located regional training centres in developing regions. Selected laboratories in these regions that are already involved in DNA research could be identified for catering to the training needs of neighbouring countries.

As these new technologies are very cost-intensive and require a high level of specialization, regional collaboration and networking of countries should be encouraged. The CGIAR Centres are in a good position to promote the appropriate integration of modern biotechnology tools into genebank management. Efforts in this direction, supported by national programmes, will certainly enhance the collection (enrichment), conservation, characterization, documentation and, above all, the utilization of plant genetic resources for the benefit of humanity (de Vicente 2004).

Conclusions

DNA banks may evolve as a strategic component of modern genebanks providing the basis for improved genebank management and facilitating high throughput germplasm characterization, association genetics and marker-assisted selection. DNA samples can be exchanged much more easily and at lower costs than can living plant materials, without the inherent risk of spreading diseases and pests. DNA banks serve as reference basis for evolutionary and comparative genomic studies and may offer a complementary conservation strategy for species under threat of extinction. They can also serve as safety duplicates for the physical seed, field or *in vitro* collections. However, there are also several limitations of a technical and legal nature to DNA storage in genebanks. The major limitations include the lack of technologies to regenerate plants from stored DNA and the relatively short life-span of stored DNA samples. Some limitations apply specifically to genebanks in resource-poor developing countries, which require support to build up and maintain the necessary infrastructure and to train human resources in DNA and molecular techniques. Regional collaboration and networking among countries is imperative, given the costs and the level of specialization required for these new technologies.

References

Ahmedullah A. 1999. Country report on red listing of plants in India. *In* Using the IUCN red list criteria at the national level: A regional consultative workshop for South and South Asia. Available from: http://www.biodiversityasia.org/ebook/html. Pp. 20–22.

Andersson MS. 2004. Feasibility study for DNA storage and use: DNA banks—providing novel options for genebanks? A worldwide survey. IPGRI Report No. 6900/D10A. International Plant Genetic Resources Institute (IPGRI), Rome, Italy. 38 p.

Causse M, Leconte L, Baffert N, Duffé P, Hospital F. 2001. Marker-assisted selection for the transfer of QTLs controlling fruit quality traits into tomato elite lines. Proceedings of an International Symposium on Molecular Markers. Acta Hort 546:557–564.

de Vicente MC, editor. 2004. The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics. Issues in Genetic Resources No. 11. International Plant Genetic Resources Institute (IPGRI), Rome, Italy.

de Vicente C, Metz T, Alercia A. 2004. Descriptors for genetic markers technologies. International Plant Genetic Resources Institute, Rome, Italy. 24 p.

FAO. 1993. Forest resources assessment, 1990. Tropical countries. Food and Agricultural Organization of the United Nations Forestry Paper 112. FAO, Rome, Italy.

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schüler K. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol. Breed. 13:93–102.

Graner A, Dehmer KJ, Thiel T, Bömer A. 2004. *In* de Vicente MC, editor. The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics. Issues in Genetic Resources No. 11. International Plant Genetic Resources Institute (IPGRI), Rome, Italy. Pp. 26–32.

Hamon S, Frison E, Navarro L. 2004. Connecting plant germplasm collection and genomic centres: how to better link curators, molecular biologists and geneticists? *In* de Vicente MC, editor. The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics. Issues in Genetic Resources No. 11. International Plant Genetic Resources Institute (IPGRI), Rome, Italy. Pp. 33–45.

Hartl DL. 2000. A Primer of Population Genetics. 3rd ed. Sinauer Associates, Sunderland, MA, USA. 221 p.

Hodgkin T, Brown AHD, van Hintum ThJL, Morales EAV. 1995. Core Collections of Plant Genetic Resources. John Wiley and Sons, Chichester, UK.

Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. Technical Bulletin No. 2. International Plant Genetic Resources Institute (IPGRI), Rome, Italy.

Khlestkina EK, Huang XQ, Quenum FJ-B, Chebotar S, Röder MS, Börner A. 2004. Genetic diversity in cultivated plants—loss or stability? Theor. Appl. Genet. 108:1466–1472.

Morgante M, Salamini F. 2003. From plant genomics to breeding practice. Curr. Opin. Plant Biotechnol. 14:214–219.

Myers N. 1988. Threatened biotas: 'hotspots' in tropical forestry. The Environmentalist 8:1–20.

Ortiz R, Engels J. 2004. Genebank management and the potential role of molecular genetics to improve the use of conserved genetic diversity. *In* de Vicente MC, editor. The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics. Issues in Genetic Resources No. 11. International Plant Genetic Resources Institute (IPGRI), Rome, Italy. Pp. 19–25.

Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. in Plant Biol. 5:94–100.

Richards C. 2004. Molecular technologies for managing and using genebank collections. *In* de Vicente MC, editor. The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics. Issues in Genetic Resources No. 11. International Plant Genetic Resources Institute (IPGRI), Rome, Italy. Pp. 13–18.

Royal Botanic Gardens, Kew. 2004. DNA Bank Database. Available from: http://www.kew.org/data/dnaBank/homepage.html

Sackville-Hamilton NR, Engels JMM, van Hintum ThJL, Koo B, Smale M. 2002. Accession management: combining or splitting accessions as a tool to improve germplasm management efficiency. Technical Bulletin No. 5. International Plant Genetic Resources Institute (IPGRI), Rome, Italy.

Tanksley SD, Nelson JC. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191–203.

TERI. 2004. Hot spots in India. Available from: http://www.teriin.org/biodiv/hotspot.htm

Xiao J, Grandillo S, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L. 1996. Genes from wild rice improve yield. Nature 384:223–224.

Zeven AC, de Wet JMJ. 1982. Dictionary of cultivated plants and their regions of diversity; excluding ornamentals, forest trees and lower plants. CAPD, Wageningen, The Netherlands.