Issues in Genetic Resources No. 11 August 2004

The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics

M. Carmen de Vicente (editor)

Abstract

The role played by genebanks as repositories of plant genetic resources has evolved since their inception because of the need to adapt to the changing demands of their different clients. The past 20 years have witnessed a significant unravelling of important pieces of genetic knowledge. As a result, the organization of DNA is now well understood, information on how genes function has increased, and the relationship between phenotype and genotype is better documented, forcing us constantly readjust the value we assign to genetic resources and tap new ways of better exploiting the wealth they represent. Genebanks may need to revise their principles, not because former tasks should be abandoned but because of the new light shed on genetic resources as well as the new clients that emerge. Contemporary clients seek expertise that ranges from traditional breeding to molecular biology and even state-of-the-art genomics. Genebank managers must accordingly offer a broader range of services, and staff must be capable of covering an overarching array of disciplines. Thus several issues must be addressed, such as the attributes that a genebank should maintain, the convenience of networking to outsource certain types of expertise and procedures, the gap this entire situation may create among genebanks, and the benefits genebanks offer in countries with varying levels of development. This paper addresses most of these concerns based on a consultation held with genebank curators, breeders, molecular biologists and geneticists. It aims to present not so much solutions as arguments that might steer a constructive exchange of ideas in coming years, so that a balance may be found between the need to maintain genetic resources and the required infrastructure and those for providing the additional services that modern science demands.

FUTURE HARVEST
<www.futureharvest.org>

IPGRI is a Future Harvest Centre supported by the Consultative Group on International Agricultural Research (CGIAR)

Comments and queries on the contents of this paper are welcome and should be addressed to:
Volume Editor: M. Carmen de Vicente
Genetic Resources Science and Technology (GRST)
IPGRI, Via dei Tre Denari, 472/a, 00057 Maccarese, Rome, Italy

Issues in Genetic Resources is an occasional series of papers published by IPGRI on important topics of interest to the genetic resources community.

You may download any Issues paper as a portable document format (PDF) file from http://www.ipgri.cgiar.org/publications. Hard copies can be obtained from IPGRI's publications distribution office (ipgri-publications@cgiar.org).

Partners in this publication
The papers in this publication are based on discussions mediated at the SGRP–IPGRI Expert Consultation Meeting held in November 2002 in León, Spain. We extend our grateful thanks to the following entities, which collaborated with organizing the meeting and/or publishing this paper:

The International Plant Genetic Resources Institute (IPGRI) is an independent international scientific organization that seeks to advance the conservation and use of plant genetic diversity for the well-being of present and future generations. It is one of 15 Future Harvest Centres supported by the Consultative Group on International Agricultural Research (CGIAR), an association of public and private members who support efforts to mobilize cutting-edge science to reduce hunger and poverty, improve human nutrition and health, and protect the environment. IPGRI has its headquarters in Maccarese, near Rome, Italy, with offices in more than 20 other countries worldwide. The Institute operates through three programmes: (1) the Plant Genetic Resources Programme, (2) the CGIAR Genetic Resources Support Programme and (3) the International Network for the Improvement of Banana and Plantain (INIBAP).

The international status of IPGRI is conferred under an Establishment Agreement which, by January 2003, had been signed by the Governments of Algeria, Australia, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Cameroon, Chile, China, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, Greece, Guinea, Hungary, India, Indonesia, Iran, Israel, Italy, Jordan, Kenya, Malaysia, Mauritania, Morocco, Norway, Pakistan, Panama, Peru, Poland, Portugal, Romania, Russia, Senegal, Slovakia, Sudan, Switzerland, Syria, Tunisia, Turkey, Uganda and Ukraine.

Financial support for IPGRI's research is provided by more than 150 donors, including governments, private foundations and international organizations. For details of donors and research activities please see IPGRI's Annual Reports, which are available in printed form on request from ipgri-publications@cgiar.org or from IPGRI's Web site (www.ipgri.cgiar.org).

The geographical designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of IPGRI or the CGIAR concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries. Similarly, the views expressed are those of the authors and do not necessarily reflect the views of these organizations.

Mention of a proprietary name does not constitute endorsement of the product and is given only for information.

The CGIAR System-wide Genetic Resources Programme (SGRP) joins the genetic resources programmes and activities of the Future Harvest Centres in a partnership whose goal is to maximize collaboration, particularly in five thematic areas. The thematic areas—policy, public awareness and representation, information, knowledge and technology, and capacity-building—relate to issues or fields of work that are critical to the success of genetic resources efforts. The SGRP contributes to the global effort to conserve agricultural, forestry and aquatic genetic resources and promotes their use in ways that are consistent with the Convention on Biological Diversity. IPGRI is the Convening Centre for SGRP. The Inter-Centre Working Group on Genetic Resources (ICWG-GR), which includes representatives from the Centres and the Food and Agriculture Organization of the United Nations, is the Steering Committee.

The U.S. Agency for International Development (USAID) is an independent federal government agency that provides economic, development and humanitarian assistance around the world in support of the foreign policy goals of the United States. With headquarters in Washington, DC, USAID's strength is its field offices around the world, carrying out its work in close partnership with private voluntary organizations, indigenous organizations, universities, American business companies, international agencies, other governments, and other US government agencies. The advancement of US foreign policy objectives is done by supporting economic growth, agriculture and trade; global health; and, democracy, conflict prevention and humanitarian assistance.

Fundación Española para la Ciencia y la Tecnología (FECYT), Spain. The Spanish Foundation for Science and Technology (FECYT) is an independent, not-for-profit organization, aiming at providing a continued and flexible service to the Spanish system of science-technology-corporation. FECYT contributes to the identification of needs and opportunities suggesting, at the same time, modes of operation to the agents of the scientific

research and technological innovation system. The strategic objectives of FECYT are: (1) to contribute to the course of the scientific and technological progress of the Spanish system in order to improve the communication among all agents in the system; (2) to act as a platform of encounter, analysis and interdisciplinary and intersectorial debate for the scientific, technological and business communities in the country; and (3) to promote the dissemination of knowledge in matters related to science and technology with the goal of creating a scientific and technological culture among citizens.

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Spain. The Spanish National Institute of Investigation of Agrarian and Food Technology (INIA) is an autonomous research organization of the Spanish Ministry of Science and Technology, and has a mandate to work on research, development and innovation in agriculture, foods, forestry and environment. INIA's objectives concern the programming, coordination, allocation of resources, monitoring and evaluation of scientific and research activities, in addition to the execution of research and technological development activities, including technology transfer in the areas of food and agriculture.

Food and Agriculture Organization of the United Nations (FAO). The Food and Agricultural Organization of the United Nations (FAO) is the lead agency for agriculture, forestry, fisheries and rural development. An intergovernmental agency, FAO has 180 Member Nations plus a member organization, the European Community. Since its inception in 1945, FAO has worked to alleviate poverty and hunger by promoting agricultural development, improved nutrition and the pursuit of food security—the access of all people at all times to the food they need for an active and healthy life.

Citation:
de Vicente, M.C. 2004. The Evolving Role of Genebanks in the Fast-developing Field of Molecular Genetics. Issues in Genetic Resources No. XI, August 2004. International Plant Genetic Resources Institute, Rome, Italy.

ISBN 92-9043-623-9

IPGRI
Via dei Tre Denari, 472/a
00057 Maccarese
Rome, Italy

Contents

Foreword
Emile Frison **6**

Preface **7**

I. Introduction
Maria Carmen de Vicente **8**

II. Molecular technologies for managing and using genebank collections
Christopher Richards **13**

III. Genebank management and the potential role of molecular genetics to improve the use of conserved genetic diversity
Rodomiro Ortiz and Jan Engels **19**

IV. Plant genetic resources: benefits and implications of using molecular markers
Andreas Graner, Klaus J. Dehmer, Thomas Thiel and Andreas Börner **26**

V. Connecting plant germplasm collection and genomic centres: how to better link curators, molecular biologists and geneticists?
Serge Hamon, Emile Frison and Luis Navarro **33**

VI. Capacity-building and training
Theresa Fulton and Stephen Kresovich **46**

Acronyms and abbreviations **50**

Participants at the 2002 SGRP–IPGRI Expert Consultation Meeting, León, Spain **52**

Foreword

It is more than 50 years since James Watson and Francis Crick published the now famous double helix structure of DNA, and almost 40 years since the plant breeding efforts that started the Green Revolution. In the meantime, much has changed both in molecular biology and in plant breeding. Molecular biologists can now read the genetic instructions of any organism almost at will. With the advent of novel tools such as DNA microarrays and massively enhanced computing power they will be able to move rapidly from the sequence to a good understanding of the organization and function of the genome underlying specific metabolic pathways. Plant breeders have developed unprecedented abilities to create novel genetic combinations, using both updated traditional breeding methods and more direct manipulation of the DNA. They have also adopted molecular tools such as marker-assisted selection to speed the development of improved varieties.

As plant breeding comes to make more and more use of molecular biology, it is becoming clear that the fundamental science may also have a much greater role to play in the use, conservation and management of plant breeding's raw material: genetic diversity. The inherited differences between individual plants or groups of plants are the basis of all improvements sought by plant breeders. Much of that diversity is threatened by changed farming practices and environmental changes and, perhaps ironically, by the spread of the very improved varieties whose existence depends on this diversity.

IPGRI has always been dedicated to the conservation of diversity, not for its own sake but so that it can be used by farmers, breeders and others to improve the productivity and sustainability of farming systems and thus to make a contribution to the well-being of people, especially in developing countries. Having helped others to establish and maintain collections of diversity—*ex situ* in genebanks and *in situ* and on farms—we now accept that in future the effective use of this diversity is going to assume greater importance. Molecular techniques have a lot to offer.

With this in mind, it seemed opportune to convene a meeting of experts from both ends and the middle, as it were, of the spectrum. Genebank curators and molecular biologists, and representatives of other interested disciplines, came together to exchange information, to listen to each other's requirements and capabilities, and to sketch out ways in which they could be useful to one another. This publication represents some of the first fruits of these collaborations, setting out paths and directions that participants can explore together.

I am confident that these collaborations will be of mutual benefit in the future, making both the use and conservation of genetic resources more effective. I hope too that when our successors 50 years from now look back they will see this meeting as something of a turning point too. Together, natural diversity and molecular methods will help national agricultural research systems, including their conservation programmes, as well as the Consultative Group on International Agricultural Research to achieve their goals: to increase food security, to alleviate poverty and to do so in a sustainable manner.

Emile Frison
Director General
International Plant Genetic Resources Institute
December 2003

Preface

This presentation is a product of three days of discussion held during an expert consultation meeting in November 2002 in León, Spain. The meeting was organized by IPGRI for the System-wide Genetic Resources Programme (SGRP). It brought together scientists from different CGIAR centres, national agricultural research programmes, universities and the private sector.

The meeting's participants agreed to make the topics discussed available to a wider public by publishing in IPGRI's series *Issues in Genetic Resources*. Together with Coosje Hoogendoorn and Jan Engels, we put together an outline covering the main topics and recommendations that had arisen in the meeting's discussions. Contact was made with participants interested in contributing a paper or willing to provide consultancy.

The presentation aims to cover those significant issues that developed as a result of introducing molecular technologies into the fields of germplasm and genebank management in such a way that they are likely to affect the future strategies and activities of these disciplines. We hope the ideas presented in this publication will help develop a vision of the genebanks' future role in an era of enormous development in molecular biology and genetics.

We would like to give special thanks to the contributors of the papers in this publication: C. Richards, R. Ortiz, J. Engels, A. Graner, K.J. Dehmer, T. Thiel, A. Börner, S. Hamon, E. Frison, L. Navarro, T. Fulton and S. Kresovich. We also extend our appreciation to those participants who, although they did not contribute papers, provided significant input by offering their perspectives to the discussions held during the meeting: J.I. Cubero, O. de Ponti, E. Dulloo, M.E. Ferreira, P. Freymark, C. Hoogendoorn, J.L. Karihaloo, J.M. Martínez Zapater, W. Roca, N.R. Sackville-Hamilton, B. Skovmand, V. Villalobos and X. Zhang.

We also extend our appreciation to IPGRI's Communications Services, who helped bring the publication into being, Lynn Menéndez at CIAT's Communications Unit and Elizabeth McAdam who helped with the production of this document.

I. Introduction

Maria Carmen de Vicente

Background

Ex situ collections of plant genetic resources (PGR) began with the botanic gardens of the Middle Ages and evolved during the late nineteenth to early twentieth centuries into so-called 'introduction stations'. At that time, no special criteria existed for conservation methodologies, given that the intention was essentially to use germplasm rather than conserve it.

The advent of modern plant breeding brought with it a greater need for diversity in breeding materials. Concurrently, novel varieties, particularly hybrids, became widely used because of their higher yields. This, in turn, led to large-scale replacement of traditional varieties, even crops, and rapidly diminishing on-farm sources of potentially valuable genetic resources. This genetic erosion raised concerns for the future availability of PGR.

In response, the introduction stations became genebanks, which gave plant breeders easier and quicker access to a wider spectrum of genetic resources. Genebanks began to take on the additional role of being repositories of threatened germplasm materials likely to be useful to humans. Several national genebanks were established, among which were the N.I. Vavilov Research Institute of Plant Industry (VIR, St Petersburg, Russia), first established in 1894; the Institute of Plant Genetics and Crop Plant Research (IPK, Gatersleben, Germany) in 1943; and the National Seed Storage Laboratory (NSSL, Fort Collins, Colorado, USA) in 1947.

In the 1960s, the first international genebanks, like those of the CGIAR, were created. In 1974, the IBPGR (now IPGRI) was established and, with it, a coordinated effort to collect threatened germplasm. A network of base collections was set up, and a start made to develop information systems to enhance germplasm use by facilitating its distribution and exchange.

Germplasm collection evolved from unsystematic, opportunistic and *ad hoc* approaches to well-planned, ecoregional or crop-oriented initiatives. Generalized germplasm explorations developed into well-informed, targeted collecting activities with specific purposes. Many countries set up national genebanks to service their public breeding programmes. Extensive *ex situ* collections became established for small, easy-to-store seeds, such as those of major cereals. Significant advances were made in developing storage methodologies for recalcitrant seeds and vegetatively propagated species. Even so, these collections are still relatively limited both in number and in the genetic variation they represent.

The role of molecular biology and genetics in modern genebank management

Since the late 1980s, the fields of molecular biology and genetics have undergone significant developments, most of which have offered new ways of solving biological questions in general and those of agricultural sciences in particular. At first, technologies to assess polymorphism—that is, the genetic variation among samples or individuals—were designed. They permitted analysis of genetic diversity and, thus, the creation of modern linkage maps, which became increasingly saturated, as more powerful marker tools became available.

Genetic maps led to the development of physical maps and the construction of many types of DNA libraries, all with the main goal of speeding up the cloning of individual genes of agronomic importance. Simultaneously, much progress was made with high-throughput technologies for gene expression analyses that helped uncover candidate genes involved in central metabolic pathways. DNA sequencing technology also improved from manual to sophisticated automated methods, facilitating increased throughput of efforts to decipher genomes of organisms from all kingdoms. Consequently, gene function could be better studied, and advances were made in bioinformatics, the discipline of

generating, collecting, storing and using data from genomic projects to achieve research objectives.

Knowledge has also expanded in areas such as understanding the potential of wild species to significantly contribute to traits of agronomic importance. A large array of possibilities has also opened up through the disciplines of comparative genetics and genomics. These new opportunities allow us to envisage a significant impact on germplasm conservation activities, which would certainly benefit from better-informed management and increased use of PGR.

When considering these circumstances, in 2002, IPGRI called for a meeting of experts to consult them on the implications that developments in molecular genetics may have for the future of *ex situ* germplasm collections. The meeting brought together 22 participants, all experts in germplasm, genebanks or molecular genetics. They came from universities, research institutes, national seed repositories, the private sector, IPGRI and four other CGIAR centres (CIMMYT, CIP, IITA and IRRI). The countries they represented were Brazil, China, France, Germany, India, Mexico, the Netherlands, Spain and the USA.

The meeting dealt with various aspects of the evolving role of genebanks with respect to developments in the molecular sciences, including: (1) extending the concept of 'client' to include not only the plant breeder, but also the basic scientist and molecular geneticist; (2) issues of intellectual property protection of genetic components of accessions; (3) the need to standardize molecular technologies; (4) the prospects of genebanks developing new services; (5) the range of germplasm materials that future users are likely to need; (6) comprehensive capacity-building; (7) the impact of scientific advances on the technological divide already existing between rich and poor countries; (8) networking needs; (9) the importance of continued phenotyping by genebank curators; (10) the need for increased documentation capacity; (11) access to information; and (12) the role of CGIAR in general and IPGRI in particular.

Experts' recommendations

At the end of the meeting the experts compiled 18 recommendations, which reflect the likely trends in future management of *ex situ* conservation of genetic resources. These recommendations are listed below, according to the four main areas as given in the meeting: research and management of genebanks, information technologies and bioinformatics, training and capacity-building, and policy and intellectual property rights (IPR).

Area 1: Research and management of genebanks

Molecular technologies

- The efficient and effective exploitation of the genetic diversity found in large genebank collections consumes considerable time and resources. Hence, molecular technologies should be used to create representative subsets of either the entire or a core collection. However, phylogenetic characterization should continue to be done for the whole collection, not just for the core or subset.
- Advances in genomics of major crops and model species should be applied to improve understanding of intraspecific variation in species, and to cover new species.
- Marker-based technology should be implemented in genetic resources management. Robust and affordable technologies should be promoted to improve genebanks' quality of output. In certain cases, the private sector could be approached to collaborate in crop-centred consortia and to provide information on reliable markers for their use in the public domain.
- In certain cases, it is advisable to wait before molecular techniques are implemented. This would allow costs to come down and throughput capacities to increase further, hence preventing loss of resources through investments made in acquiring technologies that might rapidly become obsolete.

Extent of genetic diversity coverage for ex situ *conservation*

• Studies should be carried out to define the total variation of a given species so that at least the smallest possible set of representative individuals or accessions is conserved. Such core collections, however, do not remove the need to maintain large, or larger, collections to cover variation at all levels (e.g. genes, individuals). In addition, priority in conservation should be given to materials that contain alleles that are locally common, but globally rare.

• Because (1) wild ancestors of cultivated species contain vast amounts of additional genetic diversity, and (2) the efficiency of transferring their desired genes is significantly increased by applied marker-assisted backcrossing, efforts to collect, characterize and evaluate accessions of wild relatives of crops should be substantially increased.

• Because neither conserving all living species *ex situ* nor converting the planet Earth into one protected area is feasible, well-defined regions in different parts of the world (i.e. the 'hot spot' zones) should be designated as protected areas. Better links must be established between *in situ* and *ex situ* conservation.

Phenotyping

• Precise and extensive phenotyping (i.e. the characterization and evaluation of morphological, agronomic, physiological, pathological and biochemical traits) should continue to be a major activity of genetic resources repositories. Molecular technologies must be adopted as a supporting tool that complements phenotypic screening, but does not replace it.

Collaboration

• No single organization can be expected to do everything. Hence, genebank curators must seek to implement new molecular technologies through collaboration with each other and with molecular geneticists, bioinformaticists and public–private crop-centred consortia.

• Consortium networks between genebanks, breeding and molecular genetics programmes must be promoted and established. Such collaboration is essential to efficiently exploit genetic diversity in wild relatives of crops. It is expected that genebanks will receive an increased number of requests for wild species.

Area 2: Information technologies and bioinformatics

Novel genebank services

• In the future, emphasis in requests is expected to increase on specific traits, quantitative trait loci (QTLs) and alleles. Genebanks must be prepared to expand their documentation and service tasks accordingly, for example, by providing not only information, but also DNA samples, as well as seed samples.

Integrating traditional and molecular data

• Substantial progress has already been made in bioinformatics outside the genetic resources community. Curators should seek collaboration with existing bioinformatics programmes to integrate molecular data (both marker and sequence data) with phenotypic and genebank management data.

• Genebanks should seek to cooperate with each other across national boundaries to develop central germplasm databases, including both traditional germplasm data and molecular data for individual crop genepools. Comparative genetics will most likely play an important role in germplasm management, characterization and use if this information is well organized and becomes publicly available.

• Action should be taken to promote the synthesis of ongoing work and implement findings and practices gathered through the use of model and globally important species.

Standardizing protocols and data

• Crop-centred consortium networks should be organized to promote the use of

common molecular technologies and their standards in genebanks. In addition, specific bioinformatics applications will be needed to create common standards of germplasm data management and documentation.

Area 3: Training and capacity-building

- Comprehensive training should be offered and implemented in developing countries to build up genebank human resources in molecular genetics. Courses should include basic principles of biology, genetic resources management, experimental design, data analysis and genomic sciences.
- Regional collaboration among institutions in the South should be promoted. The CGIAR, as a partner in consortia of germplasm and genomic projects, should provide leadership in the appropriate integration of modern genetics and genomic developments and tools into the conservation, characterization, evaluation and utilization of crop genetic resources.

Area 4: Policy and intellectual property rights

- New issues related to access and benefit sharing will undoubtedly arise as a consequence of research and information generated that result from the application of molecular technologies to germplasm materials, as well as tools. Consequently, timely workshops must be held to generate the appropriate knowledge for addressing these issues, thus helping to solve conflicts that may prevent germplasm materials from flowing.

The discussion papers

Many issues were discussed in our experts' meeting, which can be considered fundamental in defining IPGRI's approach to the role of genebanks in the next 10 years and, as such, are deeply embedded in IPGRI's overall strategy as an international centre with a mandate for conservation and use of PGR. The following five papers will be a valuable reference for those interested in genebanks and their relation with the modern, fast-moving fields of the molecular sciences. Brief summaries of the discussion papers that follow this introduction and that arose from the 2002 experts' meeting are presented below:

Chapter II aims to justify the incorporation of molecular technologies into genebank management activities as the only means by which to comply with the objective of making genetic diversity useful for plant breeding and basic research. It points out the significance of discovering meaningful variation in wild species, and the value of using marker techniques to organize germplasm. The effects of these techniques on guiding collection strategies and data distribution are described. The potential of comparative genetics, the need for well-established databases and data standardization are discussed, together with the role of comparative genetics in accelerating useful gene discovery. These activities should always be backed by good phenotyping and traditional breeding programmes.

Chapter III examines how molecular technologies may help discover the relevance of hidden characters in germplasm, especially in wild and weedy germplasm. Such discovery, in turn, determines the importance of maintaining germplasm in *ex situ* collections worldwide. From a wider perspective, the paper focuses on different aspects of germplasm management that, with the aid of molecular tools, will benefit decision-making on the extent and composition of collections.

By using real examples and figures, *Chapter IV* deals with novel solutions to traditional germplasm management practices such as taxonomic identification, duplication of accessions and verification of identity after regeneration or multiplication. The paper also points out the importance of taking into account those technical aspects involved in the effective management of molecular resources and data. Finally, the paper reflects on unsolved questions and issues such as the technological divide between developed and less

developed countries, the fact that most available financial resources are poured into only a few major crops, the need for appropriate guidelines and international coordination, the likelihood that genebank services will extend to provide DNA samples as well as traditional materials, the need to create links between different sources of genebank documentation and, lastly, bioinformatics.

Chapter V discusses the need to build networks, and indicates how important is the role of the genebank curator, who has a central position in relation to the community of users. Simultaneously, the difficulties of linking with different scientists and conveying appropriate information are recognized. Consequently, on studying recent examples in genomics, suggestions are made to reduce existing gaps and ensure desirable connections between curators, breeders and molecular scientists.

Capacity-building is the topic of *Chapter VI*, which demonstrates a clear need for it in light of the expanding genebank clientele. Comprehensive training for curators and related staff is essential to help close the technological divide. Networking is seen as having a special role, and as being effectively conducted by the CGIAR centres.

II. Molecular technologies for managing and using genebank collections

Christopher Richards
National Center for Genetic Resources Preservation, USDA-ARS, Fort Collins, Colorado, USA

In a real sense, genebanks are a study in contingencies. The fact that heritable genetic variation is the basis for adaptation (and crop improvement) has served as a rudimentary and diffuse guide to the assembly of the large *ex situ* collections developed over the last century. In many respects, the acquisition and maintenance of genetic diversity has been motivated more by the certainty of loss of local landraces and the vulnerability of wild populations in the face of increased habitat loss than from any specific evidence for agricultural worth. Moreover, the cost of storing and maintaining large collections is an investment that seems particularly cost-effective, compared with the risk of losing these valuable genetic resources (Pardey *et al.* 2001).

While Vavilov's original premise—that the collection of wild relatives is critical for agricultural improvement—has greatly influenced the scope of worldwide collecting efforts, the active use of these resources to develop modern varieties has been largely unsuccessful. Reasons may include the need for prebreeding lines that serve as a first step in backcrossing programmes and the inadequate characterization of germplasm. The net result is that, for most domesticated crop species, pedigrees can be traced to a handful of genetic lineages.

Genebanks have an obligation to make their genetic diversity useful and accessible for breeding and research. Large collections are under increasing pressure to become efficient at reducing redundancy, and documenting the variation they contain, lest they become relegated to the level of 'seed morgues'. Increasingly, this assessment has been augmented by the use of molecular markers that quantify the genetic diversity within and between accessions. Several works have detailed the philosophy behind optimizing collections to ensure diverse genetic representation through either the creation of core collections or some form of hierarchical sampling (Crossa *et al*. 1994; Brown and Marshall 1995; Hayward and Sackville-Hamilton 1997).

This paper aims to highlight several technical and analytical advances that have enormous potential in promoting the use of this diversity for breeding. It also identifies several points of synergy between large-scale genomic projects and genebank management and use.

The utility of wild germplasm

Mather and others (Mather 1941) put forward the paradigm that most complex traits are under the control of many genes with small additive effect. This paradigm has yielded to a more nuanced one of molecular characterization of phenotypic traits (Mackay 2001). Empirical data suggest that the allelic effects on a trait value are distributed as an exponential function, where a few genes control most phenotypic effects and other associated genes have an increasingly smaller effect.

High-density linkage maps have enabled researchers to identify genes having major effects (i.e. QTLs) on several complex plant traits (Tanksley 1993). Large-scale sequencing efforts to compile the complete *Oryza* and *Arabidopsis* genomes have provided more than just data on these species; they have also provided a framework for a whole host of associated inquiry that will have impact on the way genebanks provide information.

As large biological databases of genomic linkage maps and expressed sequence tag (EST) sequences become more accessible, the technical process of locating QTLs in model systems has become routine. These advances have proved to comprise a powerful new tool in characterizing plant genotype–phenotype relationships. Overall, empirical evidence suggests that much of the genetic variation useful to agricultural improvement is not recognizable in the plant phenotype. Instead of screening for promising phenotypes, the future may rely more on allele mining of wild germplasm guided by some information on phylogeny, population structure and genetic diversity. Most modern varieties have a very narrow genetic pedigree and use exotic germplasm.

Although labour intensive in a breeding programme, alleles of wild germplasm have the potential to produce substantial payoffs in phenotypic response.

Examples of breeding for a particular trait by searching for alleles of the underlying QTLs in phenotypically unpromising wild germplasm have underscored the critical importance of wild germplasm and the genetic diversity it represents to agriculture (Tanksley and McCouch 1997). While alleles of important QTLs are not restricted to the genepools of domesticated taxa, their phenotypic expression in wild genetic backgrounds may be variable. The field of quantitative genetics has moved from an era where complex statistics were used to physically locate these loci to the more complex task of dissecting their epistatic interactions in different genetic backgrounds (Mackay 2001).

Genebanks offer a key link in this process by providing the raw material for these kinds of breeding programmes. However, the relationship between genome projects and genebanks can be even more synergistic. The increased pace of plant genomic studies offers genebanks a set of useful, well-characterized loci to characterize more accurately standing diversity, and to detect historical relationships among lineages and patterns of molecular evolution and selection that have occurred at the DNA sequence level. Indeed, it is only through an evolutionary framework that the significance of these large genome initiatives will be realized (Charlesworth *et al*. 2001). Significant advances in the field of comparative genetics, molecular phylogenetics and coalescence theory will all affect the way genebanks collect and distribute data.

Exploiting genetic resources through comparative genomics

Although relatively few plant taxa have yet been intensively sequenced and mapped, *Brassica, Populus* (poplar), *Medicago* and *Lotus* are being sequenced. *Oryza* and *Arabidopsis* serve as primary templates for the plant genome. The completed sequences have yielded a wealth of molecular marker polymorphisms, including insertions of transposable elements, small insertions or deletions, tandem repeats and single nucleotide polymorphisms that are useful for fine-scale mapping of phenotypic traits in these species.

A powerful tool of genomics, and the hallmark of bioinformatics, is the ability to efficiently compare the sequence of any gene to that of any other. Efficient query and comparison software, increases in database sequence holdings and large increases in computer-processing speed have all contributed to substantial changes in the way data are annotated, analysed and disseminated. One profound result of the mapping effort in cereal crops, including wheat, maize, rice and other grasses, is the remarkable conservation of gene content and gene order through the 60 million years of speciation events in the *Poaceae*. The exact collinearity of these genomes is perturbed by translocations, deletions, duplications and other mutational events over time, but large regions of chromosomes show high degrees of similarity (Gale and Devos 1998; Paterson *et al*. 2000).

This homeology (vestiges of direct sequence homology from these species' common ancestor) suggests that intensive sequence analysis in one species can have beneficial effects on the mapping of other species. The comparative method can be used to infer physical map locations of genes in other taxa (Brueggeman *et al*. 2002). As genetic maps from a variety of species are aligned, the patterns of similarities and differences emerge. These comparisons have important implications for understanding the evolutionary trajectories that shape genetic diversity in plants.

Research initiatives such as the development of comparative database structures systematically exploit molecular linkage map and sequence data from different taxa to understand several evolutionary mechanisms,

including the rate of heterogeneity among lineages, convergent evolution among genes and functional genomics of orthologous genes in different taxa (Ware *et al*. 2002). The prospect of regions of synteny among diverse taxa bodes well for gene discovery and may greatly affect the speed and efficiency of mapping studies in wild taxa that previously had no genomic analysis.

One particularly profitable use of comparative genomics contrasts wild and domesticated sister taxa to identify genes that may have undergone strong selection during domestication (Vigouroux *et al*. 2002). Domesticated lineages have several selected phenotypic traits that may become manifest in genomic scans for selection. These types of comparisons are possible only with sophisticated analytical tests for neutrality of polymorphisms at the sequence level. When studying molecular evolution, one assesses changes in sequence divergence to accept or refute patterns consistent with neutral variation at equilibrium between drift and mutation. Inferring processes that give rise to observed patterns of polymorphisms is a central objective of molecular evolutionary biology.

Selective, demographic and random processes can all play important parts in shaping DNA sequence polymorphisms. Several tests have been developed to detect the effects of some of these processes (Tajima 1983; McDonald and Kreitman 1991; Fu and Li 1993). In addition, a stochastic model of genealogical descent known as the coalescent provides a framework for analysing the polymorphism data currently observed. This analytical approach puts forward a null model based on a continuous-time Markov process for generating random genealogies under certain population parameters.

Gene genealogies produced in this way are random outcomes of an underlying evolutionary process that can be compared with the observed sequence polymorphisms. This enables the researcher to accept or refute specific models of evolutionary change and, as such, the genealogies comprise a powerful simulation tool for hypothesis testing and exploratory analysis. In the study mentioned above by Vigouroux *et al*. (2002), both neutrality tests and coalescent simulations were used on simple sequence repeat (SSR) polymorphisms within EST sequences in maize and teosinte. Their data identified several genes that were implicated in the domestication of maize and therefore had important agronomic value.

It is noteworthy that in crop species under artificial selection, numerous QTLs have been attributed to regulatory genes (not just structural genes), demonstrating the ability of plant regulatory genes to influence quantitative phenotypic variation, in addition to their previously demonstrated impact on discrete traits (Doebley and Lukens 1998). By extension, in natural systems, intraspecific phenotypic variation important to ecological interactions may also be controlled by regulatory loci. Regulatory alleles affecting ecologically important traits such as flower shape and symmetry, inflorescence architecture, corolla pigmentation pattern, flowering phenology and fruit size have been identified. Experiments are being conducted to examine the micro-evolutionary processes affecting the distribution of these alleles (Purugganan 2000).

Molecular genetics and *ex situ* management

Molecular marker technologies are increasingly being used in the genetic resources arena to quantify the levels of genetic diversity within and among accessions and to increase the efficiency of collection management (van Treuren *et al.* 2001). These techniques usually rely on the frequencies of neutral markers such as those generated through random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLP), SSR and other sequence-based methods that seek to quantify variation

at variable loci. This distinctly prospective method is used to cluster the current genetic diversity in a way that facilitates representation in *ex situ* collections (Bretting and Widrlechner 1995; van Hintum and van Treuren 2002).

The data from these studies have had significant impact on our ability to infer genetic relatedness, and demographic histories of populations and species. In wild lineages, the pattern of neutral variation captured through these markers may provide a framework for *ex situ* management but the functional 'ecotypic' variation—the potentially adaptive variation among accessions—may not be so easily revealed (Reed and Frankham 2001; but see Merilä and Crnokrak 2001).

In a sense, neutral variation has been widely used as a tractable surrogate for the functional genetic variation that underlies quantitative traits associated with adaptation and future evolutionary potential. Although molecular characterization of the plant phenotype is still emerging (Purugganan and Gibson 2003), comparative approaches and the development of large bioinformatics data sets in model systems will enable the execution of population genetic studies of ecologically or agronomically important traits.

In terms of genebank management, emphasis will increasingly be given to maintaining genetic diversity at key loci that control important traits of agronomic interest. Molecular tools may not only influence the assessment of broad patterns of genetic diversity but may also be critical in modelling gene genealogies of key functional loci that have been shaped through selection and drift in natural systems (Hey and Machado 2003). As comparative studies increase and more traits are genetically dissected, the panels of functional genes used to screen germplasm will increase. Analytical techniques for revealing the signature of selection at these loci may not only be important in choosing unique accessions to include within an *ex situ* collection, but may also be important tools for ensuring that diversity is maintained at an acceptable level over time.

Because seeds deteriorate during storage, samples must be regenerated before viability becomes critically low. Each round of regeneration exposes the accession to sampling error, causing genetic drift (loss of alleles), possible selection (changes in specific allele frequencies) and contamination (novel alleles introduced). Comparisons between diversity measures across regeneration cycles are few (Wu *et al*. 1998; Chebotar *et al*. 2002), but they demonstrate that collections are not static and even accessions of domesticated species can show substantial genetic changes within a few generations. The challenge to genebank managers is to disentangle the stochastic and deterministic elements in these dynamics.

While breeding structure and demographic histories (such as rapid changes in effective population sizes) are expected to affect all markers in similar ways, functional loci (expressed or regulatory), subject to selection, show high rates of gene-to-gene variation. Comparisons among species with different breeding systems and life history traits may offer genebank managers important conceptual guidelines about which type of accessions are particularly prone to genetic erosion and which are not. The benefits of undertaking genetic diversity studies in functional genes in model systems may be better management practices that maintain long-term accession integrity (van Tienderen *et al*. 2002).

The acceleration of genomic data puts genebanking in a particularly vital position for future gene discovery and agricultural improvement. The linkage of model systems to wild germplasm can be thought of as an iterative process. Model systems provide a framework for gene identification and synteny. These loci will become increasingly incorporated in marker panels to monitor and maintain diversity in a variety of accessions. Both neutral marker and specific loci are used to infer the historical genealogies

of these loci and also to refine phylogenetic relationships between taxa. The use of novel QTL alleles retrieved from wild germplasm may implicate additional loci critical to some phenotypic trait such as another gene in a metabolic network that can be subsequently mapped and characterized in the model system.

With increasing sequence-based information of QTLs of agronomic importance, genotyping panels for genebanked accessions will greatly increase the accessibility and the efficient maintenance of large *ex situ* collections.

The critical importance of plant phenotype

The future will bring an increase in genotypic data, but the critical importance of plant phenotyping and traditional plant breeding will not diminish. Without expert evaluation for traits in the field, mapping studies cannot proceed. In addition, the critical need for prebreeding lines in which to evaluate the effect of novel QTL alleles will only increase. Molecular marker-assisted breeding can be highly efficient, but it is breeding all the same and requires the skill of researchers with whole plant and breeding experience.

The use of genomic data in genebank management also puts a heavy priority on bioinformatic solutions that incorporate data from sequences to field characters and descriptor data. The ability to link heterogeneous data sets requires a high degree of data standardization. Sequence data may become the gold standard, rather than genotypes based on fragment analysis. Similar database structures are currently being implemented in medical applications where the connections between clinical and genetic data are vital to therapeutics and gene discovery.

In the plant germplasm community, it may only be a matter of time before these database structures are used routinely to weave together data from the NCBI, Gramene, SINGER and GRIN databases. Whatever the eventual data retrieval system, molecular technologies will continue to play an important role in making genebanks accessible for agronomic improvement.

References

Bretting, P.K. and M.P. Widrlechner. 1995. Genetic markers and plant genetic resources. Plant Breed. Rev. 13:11–86.

Brown, A.H.D. and D.R. Marshall. 1995. A basic sampling strategy: theory and practice. Pp. 75–91 *in* Collecting Plant Genetic Diversity: Technical Guidelines (L. Guarino, R. Rao, V. Reid and R. Reid, eds.). CAB International, Wallingford, UK.

Brueggeman, R., N. Rostoks, D. Kudma, A. Killian, F. Han, J. Chen, A. Druka, B. Seffenson and A. Kleinhofs. 2002. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. Proc. Natl Acad. Sci. USA 99:9328–9333.

Charlesworth, D., B. Charlesworth and G.A.T. McVean. 2001. Genome sequences and evolutionary biology, a two-way interaction. Trends Ecol. Evol. 16:235–242.

Chebotar, S., M.S. Roder, V. Korzun and A. Börner. 2002. Genetic integrity of ex situ genebank collections. Cell. Mol. Biol. Lett. 7:437–444.

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

Doebley, J. and L. Lukens. 1998. Transcriptional regulators and the evolution of plant form. Plant Cell 10:1075–1082.

Fu, Y. and W. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Gale, M.D. and K.M. Devos. 1998. Comparative genetics in grasses. Proc. Natl Acad. Sci. USA 95: 1971–1974.

Hayward, M.D. and N.R. Sackville-Hamilton. 1997. Genetic diversity—population structure and conservation. Pp. 49–76 *in* Biotechnology and Plant Genetic Resources (J.A. Callow, B.V. Ford-Lloyd and H.J. Newbury, eds.). CAB International, Wallingford, UK.

Hey, J. and C.A. Machado. 2003. The study of structured populations—new hope for a difficult and divided science. Nat. Rev. Genet. 4:535–543.

van Hintum, Th.J.L. and R. van Treuren. 2002. Molecular markers: tools to improve gene bank efficiency. Cell. Mol. Biol. Lett. 7:737–744.

McDonald, J.H. and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351:652–654.

Mackay, T.F.C. 2001. The genetic architecture of quantitative traits. Annu. Rev. Genet. 35:303–339.

Mather, K. 1941. Variation and selection of polygenic characters. J. Genet. 41:159–193.
Merilä, J. and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. J. Evol. Biol. 14:892–903.
Pardey, P.G., K. Bonwon, B.D. Wright, M.E. Van Dusen, B. Skovmand and S. Taba. 2001. Costing the conservation of genetic resources: CIMMYT's *ex situ* maize and wheat collection. Crop Sci. 41: 1286–1299.
Paterson A.H., J.E. Bowers, M.D. Burow, X. Draye, C.G. Elsik, C.X. Jiang, C.S. Katsar, T.H. Lan, Y.R Lin, R.G. Ming, and R.J. Wright. 2000. Comparative genomics of plant chromosomes. Plant Cell 12:1523–1539.
Purugganan, M.D. 2000. The molecular population genetics of regulatory genes. Mol. Ecol. 9: 1451–1461.
Purugganan, M.D and G. Gibson. 2003. Merging ecology, molecular evolution and functional genetics. Mol. Ecol. 12:1109–1112.
Reed, D.H. and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095–1103.
Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.
Tanksley, S.D. 1993. Mapping polygenes. Annu. Rev. Genet. 27:205–233.
Tanksley, S.D. and S.R. McCouch. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066.
van Tienderen, P.H., A.A. de Haan, C.G. van der Linden and B. Vosman. 2002. Biodiversity assessment using markers for ecologically important traits. Trends Ecol. Evol. 17:577–582.
van Treuren, R., L.J.M. Soest, Th.J.L. van Hintum. 2001. Marker-assisted rationalisation of genetic resources collections: a case study in flax using AFLPs. Theor. Appl. Genet. 103:144–152.
Vigouroux, Y., M. McMullen, C.T. Hittinger, K. Houchins, L. Schulz, S. Kresovich, Y. Matsuoka and J. Doebley. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc. Natl Acad. Sci. USA 99:9650–9655.
Ware, D., P. Jaiswal, J. Ni, X. Pan, K. Chang, K. Clark, L. Teytelman, S. Schmidt, W. Zhao, S. Cartinhour, S. McCouch and L. Stein. 2002. Gramene: a resource for comparative grass genomics. Nucleic Acids Res. 30:103–105.
Wu, X., N. Wu, X. Qian, R. Li, F. Huang and L. Zhu. 1998. Phenotypic and genotypic changes in rapeseed after 18 years of storage and regeneration. Seed Sci. Res. 8(Suppl. 1):55–64.

III. Genebank management and the potential role of molecular genetics to improve the use of conserved genetic diversity

Rodomiro Ortiz† and Jan Engels‡
†International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria and ‡IPGRI, Rome, Italy

Genebank curators are interested in *ex situ* conservation of plant genetic resources because genetic diversity is continuously being lost in farmers' fields and in nature. Genebanks are reservoirs of biodiversity and sources of alleles that can be relatively easily retrieved for genetically enhancing crops (Ortiz 2002). Efforts have been made to collect threatened landraces, cultivars that were becoming obsolete, genetic stocks and, increasingly, wild relatives of cultivated species. All these materials are important for crop improvement because breeding gains rely largely on access to genetic variation in the respective crop genepools. If genes available in wild species are to be put into a usable breeding form, then the long-term research agenda must include the development of advanced breeding lines with the desirable genes in a suitable genetic background (i.e. prebreeding must be carried out).

Managing *ex situ* collections in genebanks

Ex situ collections are usually established either through collecting or assembling through exchange with existing collections, followed by rejuvenation or regeneration of seed and other propagules. Traditionally, the routine operation of genebanks also includes activities such as characterization, evaluation and documentation (Engels and Wood 1999; Ortiz 1999; Engels and Visser 2003). Frequently, though, characterization and especially evaluation were left to plant breeders and other users of germplasm. The lack of relevant information on the material for potential users has been judged to be the cause of the low use of accessions conserved in genebanks. This, together with the fact that accessions with proper characterization data increase the interest of the molecular geneticists, is leading to renewed attention by genebanks to characterization activities. To facilitate this work IPGRI, in collaboration with researchers from other organizations worldwide, has developed descriptor lists for about 90 crop species (for example *Allium*: IPGRI/ECP/GR/AVRDC 2001).

Analysis of genetic variation in germplasm collections generates an added value for genebanks, making this research a good investment. Well-documented analysis of the number and types of useful polymorphisms allows genebank curators to offer specific accessions with the desired characteristics to plant geneticists, who can then select materials tailored to their objectives and needs.

Despite the relatively high costs as well as the technical challenges involved, regeneration and multiplication are accepted as routine and essential activities by most genebanks. In more recent years, several genebanks have made significant investments to determine their collections' genetic diversity to improve germplasm management, including the establishment of core collections and the development of improved parents and sometimes even of new cultivars.

An adequate documentation system is an essential prerequisite for the effective management and subsequent use of genetic resources. Faster and more reliable computers allow researchers to manage and analyse larger amounts of data more easily, and publish catalogues and reports. Genebank documentation has been further enhanced with advances in information and geographic information systems (GIS) technology. Computerized documentation systems, and additional information obtained through GIS and/or from DNA marker technology, can help plant explorers search for sites where specific genes may be found. The increasing opportunities of linking different types of data from unrelated sources for one and the same accession or species greatly facilitate the use of conserved genetic resources for crop improvement and other research activities in general.

In summary, genebank curators can significantly contribute to the use of the conserved wild and cultivated genetic resources through adequate management practices. An increased application of molecular genetic

tools will further facilitate the use of germplasm in breeding efforts and add new value to the existing collections.

In situ conservation and genebanks

Over the past 15 years or so increased attention has been given to the conservation of genetic resources in their original habitats surroundings where the material obtained its distinctive characteristics, i.e. *in situ* and on-farm, respectively. In particular, the conclusion of the Convention on Biological Diversity in 1993 gave a boost to *in situ* conservation efforts. The conservation of crop genetic resources in farmers' fields allows continuing selection in diverse environments and with different selection pressures, and has relatively low direct costs. It also allows people to maintain control over their genetic resources.

In situ conservation can help preserve the co-evolutionary dynamics between crops and their wild relatives and pathogen populations of pests and diseases, which is maintaining the dynamic genetic interactions that permit micro-evolutionary changes in the host–disease system. Indeed, the co-evolution in a wild host and its resident pathogen population runs parallel to evolutionary changes in the pathogen population infecting crops. Such changes can be a response, for example, to the introduction of new cultivars containing introgressed wild resistance genes (Prescott-Allen and Prescott-Allen 1988). Likewise, pathogen biotypes from the wild alternative host can invade crops, eliciting a response reaction by wild resistance gene(s), already incorporated into improved cultivars, to the new crop pathogen population.

In view of the fact that most of the crop genetic resources are still being 'conserved' in farmers' fields and that the capacity of genebanks is usually limited, the linkages between on-farm conservation activities and genebanks are increasingly being recognized to be of critical importance. As a consequence, an active facilitation of germplasm movement between these two systems is of vital importance to both.

Wild relatives of crops and genetic enhancement

In the past, genetic diversity in wild relatives of crops and, to some extent, in wild species was predominantly the basis of the search for useful genes in resistance breeding (Lenné and Wood 1991), particularly when resistance levels to pests and diseases available in the primary (and sometimes secondary) genepool were low. Cooper *et al.* (2001) demonstrated the importance of using germplasm from wild relatives in base-broadening efforts through population management. Discovery and incorporation of new genes from wild relatives therefore provides perhaps one of the few means of sustaining crop improvement in the longer term. Although durability of resistance cannot be predicted (Johnson 1992), the use of increased genetic diversity through preventive breeding as part of the crop improvement effort may help buffer against crop losses arising as the pathogen population changes (McIntosh 1992).

Germplasm enhancement using genes from wild relatives is not an easy process, but many parents with wild genes have become available (Ortiz 2002). Backcrossing followed by selection has been the most common method for introgressing genes from wild germplasm into breeding materials. This activity has been termed 'prebreeding' or 'germplasm enhancement', an essential step in crop improvement, as well as the most promising route to increasing the use of wild germplasm. However, problems occur such as linkage drag, sterility, small sample size of the interspecific hybrid populations obtained and restricted genetic recombination in subsequent generations.

Despite its constraints, genetic enhancement using wild germplasm shows some success. For example, the ICRISAT genebank maintains about 2500 accessions (i.e. about 2% of all accessions) of wild relatives, and

wild and weedy species of sorghum, pearl millet, groundnut, chickpea and pigeon pea. Screening of this germplasm has identified several sources of resistance to important pests and diseases. Transfer of new cytoplasmic male sterility to pigeon pea and pearl millet, and the development of chickpea with enhanced yields and pigeon pea with high protein were achieved through conventional backcrossing at ICRISAT (Ortiz 2002).

Another good example is the conservation and use of wild and weedy genetic resources of rice at IRRI. Numerous disease-resistance genes have been incorporated, together with other traits, into breeding material that IRRI distributes to national rice research institutes worldwide for further use. All recently released breeding material from IRRI contains one or more genes from wild relatives of rice (R. Sackville-Hamilton, IRRI, pers. comm.).

The importance of wild and weedy germplasm for breeding programmes is demonstrated by the fact that more than 15.7% of germplasm accessions with known status and maintained by CGIAR centres are either wild relatives or weedy materials (S. Gaiji, IPGRI, pers. comm.). According to data from the SINGER database maintained by IPGRI (Table III.1), a significant flow of wild and weedy germplasm moves from the CGIAR genebanks to users worldwide. It is acknowledged that the relative importance of wild and weedy germplasm is rising, possibly because of the new opportunities that molecular genetic tools offer to exploit genetic diversity. In addition, marker-assisted backcrossing substantially increases the efficiency of this breeding approach for transferring desired genes and helps preventing linkage drag (Tanksley *et al.* 1989).

Biotechnology and genetic resources conservation

Mendel discovered the principles of heredity at the end of the nineteeth century and, from the beginning of the twentieth century, the world has seen genetics rise as a scientific discipline. Among its outstanding discoveries were DNA as hereditary material (1944), the double helix structure of the DNA molecule (1953), cracking of the genetic code (1966), isolation of genes (1973) and application of DNA recombinant techniques (from 1980 onwards) (Ortiz 1998; Engels and Visser 2003).

Allozymes were available as the first biochemical genetic markers in the 1960s and were amply used by population geneticists in their early research. In the 1970s, restriction fragment length polymorphisms (RFLPs) and Southern blotting were added to the geneticists' toolbox. *Taq* polymerase was discovered in the 1980s, and the polymerase chain reaction (PCR) developed shortly afterwards. Since then, marker analysis, based on PCR, has become routine in plant genetic research (Ortiz and Crouch 2001).

Furthermore, new marker systems have been developed based on high-density arrays or 'gene chips', allowing thousands of genes

Table III.1. Distribution of wild and weedy germplasm from collections maintained by the CGIAR Centres

Period	Distribution of:			
	Wild germplasm (average no. of accessions per year)	Wild or traditional cultivars (%)	Weedy germplasm (total no. of accessions)	Weedy or traditional cultivars (%)
1985–1989	9 534	16	296	0.5
1990–1994	17 538	41	435	1.0
1997–2001	11 861	52	308	1.4

to be arranged in small matrices (or chips) that are probed with labelled cDNA from a tissue of one's choice. DNA chip technology uses microscopic arrays of molecules immobilized on solid surfaces for analysis. An electronic device connected to a computer may read this information and analyse it. We may speculate that in the future this technology will also facilitate genetic resources management in genebanks.

Genomics, bioinformatics and germplasm preservation for genetic enhancement

Genomics research integrates genetics with informatics and automated systems to elucidate the structure, function and evolution of past and present genomes. Among the most dynamic fields of agriculture and crop improvement are the sequencing of plant genomes, comparative mapping across species with genetic markers, and objective-assisted breeding after identifying genes or chromosome regions in accessions for further research.

Likewise, molecular markers are becoming 'descriptors' that offer reproducible results for characterizing genotypes. Molecular markers are important tools for genebank management, particularly because they can be used to estimate genetic relationships between accessions within a germplasm collection. Unique genotypes can be identified and preserved, or gaps in the collection identified with the aid of DNA markers, which can be used to optimize the management of genetic diversity. Moreover, as mentioned above, DNA markers provide a means of monitoring and facilitating the introgression of genes from wild species into cultivated genepools.

Furthermore, knowledge on conservation of gene order, advances in genomics and bioinformatics will allow a much better understanding of available genes and their function in well-studied crops or gene discovery in other research-neglected tropical crop species (Mahalakshmi and Ortiz 2001; Mahalakshmi *et al*. 2002). For example, researchers might be able to identify and characterize useful genomic regions conferring a specific trait in crops. Then, appropriate test materials would be chosen to assess the relevance of these genomic regions in each targeted crop in relevant environments.

As a result of the present knowledge, the concept of genepools now includes transgenes, as well as native and exotic genepools that are becoming available through comparative analysis of plant biological repertoires. Gene chips and transposon tagging will provide new dimensions for research on gene expression. Molecular biologists study not only individual genes but also how circuits of interacting genes in different pathways control the spectrum of genetic diversity in any crop species. Genomics may accelerate the identification of important genes available in genebanks, and facilitate their utilization through transformation, without barriers across plant species or other kingdoms of living things. Perhaps, one day, it will be possible for the genes providing extreme drought tolerance of pearl millet or cowpea to be introgressed into other cereals or legumes to achieve more water-efficient crops. This will have great consequences for the way genebanks operate.

Genebanking and *ex situ* germplasm collections

The sequencing of entire crop genomes opened new frontiers in the conservation of plant biodiversity and crop genetic enhancement. Recent advances in gene isolation and sequencing in many plant species seem to justify a futuristic vision that, within a few years or possibly decades, genebank curators will complement their large cold stores of seeds with crop DNA sequences that will be electronically stored and easily accessed by users through the Internet (Ortiz 1998, 1999). This form of characterization of plant genomes will ultimately create a true genebank, possessing a large and easily accessible

inventory of major genes of today's largely non-characterized crop genepools.

Nonetheless, collections of seeds and other propagules of comprehensively studied stocks should be maintained because plant geneticists, the main direct users of germplasm maintained by genebanks, need such germplasm for their work. In fact, known important genes are only a small percentage of the total genetic information that makes up a crop plant, including gene complexes that are not yet understood and will be difficult to 're-compose' if not available as genetic stocks in genebanks.

Likewise, finding new genes in not-yet-characterized germplasm that is maintained in one or more of the about 1300 genebanks or germplasm collections in the world adds value not only to that collection or collections, but also to the electronic sequence data that could make up the genebanks of the future. Genetic resources available in genebanks are the best source for gene discovery, especially if and when the traditional collections have been phenotypically characterized and additional relevant information is properly documented.

Duplicates and germplasm restoration

Unknowingly, and usually unintentionally, accessions can be duplicated within a collection, between collections and between genebanks. These duplicates should be identified if and when economically defendable to avoid waste of capacity (Engels and Visser 2003). Putative duplicates can be identified on the basis of passport data, but additional assessment or confirmation of the duplication status will be needed through phenotypic and genotypic characterization, using descriptor lists in the field and biochemical or increasingly DNA fingerprinting in the laboratory (Lund *et al*. 2003). Such duplicated accessions may need to be bulked to prevent loss of alleles in case the duplication is only partial or if absolute duplicates are to be eliminated (Sackville-Hamilton *et al*. 2002).

Networking will help genebanks to share responsibilities, resources and costs (Frison *et al*. 2003). For example, a national or regional genebank with limited financial resources can focus on the genetic diversity occurring in its own geographic domain, and/or it may agree to duplicate collections for reasons of safety in another, better endowed, national or international genebank.

Repatriation of originally native germplasm that is available only from 'foreign' *ex situ* germplasm collections can be an important activity for those genebanks with a national or regional mandate and determined to provide better services in germplasm of their own regions. This activity may be followed by germplasm restoration whereby such material is reintroduced to sites from where it was originally collected and has since been lost for *in situ* conservation or on-farm management.

Using new technologies to establish core collections

Many genebanks have large germplasm collections, which are often inefficiently managed and are therefore seldom accessed by plant breeders. A systematic assessment of the genetic diversity in such collections can help establish core collections. These subsets of large collections contain a limited number of chosen accessions that capture most of the genetic variability in the entire collection while representing, for example, about 10% of the total collection (van Hintum *et al*. 2000). Developing a core collection therefore improves the management and use of a germplasm collection.

A core collection is assembled by taking into account the hierarchical structure of the genepool. The entire collection can be stratified into groups sharing common characteristics according to taxonomy, geographic or ecological origin, and neutral or non-neutral descriptors. Samples are then taken from these groups. Using this process, core subsets can be identified.

Genetic studies in selected crops have shown that widespread and localized alleles occurring in the entire collection are usually contained in the core subset, with only rare localized alleles excluded. The core subset often provides an entry point to further study of biodiversity of the entire collection or to the use of these resources (Hodgkin et al. 1995; Johnson and Hodgkin 1999; van Hintum *et al*. 2000).

Conclusion

Genebanks can significantly contribute to the use of conserved wild and cultivated genetic resources through adequate management practices. Increased application of molecular tools will further facilitate the use of such germplasm in crop breeding efforts and add new value to the existing collections. In particular, the identification of specific traits in wild relatives of crop species and their transfer into genotypes with a desirable genetic background is a field in which genebanks can play an important role. Furthermore, the new technologies will allow genebanks to contribute to more cost-efficient conservation efforts and to more rational conservation approaches. The increased opportunities to transfer genes across unrelated species might well have an influence on the type of germplasm collections that genebanks want to establish in the future, e.g. trait-specific collections might be added to the traditional crop, species or genepool focused collections.

References

Cooper, H.D., C. Spillane and T. Hodgkin. 2001. Broadening the genetic base of crops: an overview. Pp. 1–23 *in* Broadening the Genetic Base of Crop Production (H.D. Cooper, C. Spillane and T. Hodgkin, eds.). CAB International, Wallingford, UK.

Engels, J.M.M. and L. Visser (eds.). 2003. A Guide for Effective Germplasm Collection Management. Handbooks for Genebanks, no. 6. IPGRI, Rome.

Engels, J.M.M. and D. Wood. 1999. Conservation of agrobiodiversity. Pp. 355–386 *in* Agrobiodiversity: Characterization, Utilisation and Management (D. Wood and J.M. Lenné, eds.). CAB International, Wallingford, UK.

Frison, E., M. Mitteau, S. Sharock and B. Visser. 2003. Sharing responsibilities. Pp. 107–121 *in* A Guide for Effective Germplasm Collection Management (J.M.M. Engels and L. Visser, eds.). Handbooks for Genebanks, no. 6. IPGRI, Rome.

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

Hodgkin, T., A.H.D. Brown, Th.J.L.van Hintum and E.A.V. Morales (eds.). 1995. Core Collections of Plant Genetic Resources, International Plant Genetic Resources Institute (IPGRI), Rome, Italy. Wiley-Sayce, Chichester.

IPGRI/ECP/GR/AVRDC. 2001. Descriptors for Allium (Allium spp.). In association with the European Cooperative Programme for Crop Genetic Resources Networks and the Asian Vegetable Research and Development Center. IPGRI, Rome.

Johnson, R. 1992. Past, present and future opportunities in breeding for disease resistance, with examples from wheat. Euphytica 63:3–22.

Johnson, R.C. and T. Hodgkin 1999. Core Collections for Today and Tomorrow. IPGRI, Rome.

Lenné, J.M. and D. Wood. 1991. Plant diseases and the use of wild germplasm. Annu. Rev. Phytopathol. 29:35–61.

Lund, B., R. Ortiz, I. Skovgaard, R. Waugh and S.B. Andersen. 2003. Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor. Appl. Genet. 106:1129–1138.

McIntosh, R.A. 1992. Pre-emptive breeding to control wheat rusts. Euphytica 63:103–113.

Mahalakshmi, V. and R. Ortiz, 2001. Plant genomics and agriculture: from model crops to other crops, the role of data mining for gene discovery. Electron. J. Biotechnol. 4(3):169–178. <http://ejb.ucv.cl/content/vol4/issue3/full/5/index.html>

Mahalakshmi, V., P. Aparana, S. Ramadevi and R. Ortiz. 2002. Genomic sequence derived simple sequence repeat markers: case study with *Medicago* spp. Electron. J. Biotechnol. 5(3):233–242. <http://www.ejbiotechnology.info/content/vol5/issue3/full/2/index.html>

Ortiz, R. 1998. Critical role of plant biotechnology for the genetic improvement of food crops: perspectives for the next millennium. Electron. J. Biotechnol. 1(3):1–8. <http://www.ejb.org/content/vol1/issue3/full/7/>

Ortiz, R. 1999. Genetic diversity of cultivated crops and in situ conservation of genetic resources. Bot. Lithuanica Suppl. 2:15–30.

Ortiz, R. 2002. Germplasm enhancement to sustain genetic gains in crop improvement. Pp. 275–290 *in* Managing Plant Genetic Diversity (J.M.M. Engels, V. Ramanatha Rao, A.H.D. Brown and M. Jackson, eds.). IPGRI and CAB International, Rome.

Ortiz, R. and J.H. Crouch. 2001. Applied biotechnology in the improvement of the tropical food species. In Proceedings REDBIO IV, Goiâna, Brazil, 3–8 June 2001. REDBIO—FAO/RLAC, Santiago, Chile.

Prescott-Allen, R. and C. Prescott-Allen. 1988. Genes from the Wild Using Wild Genetic Resources for Food and Raw Materials. Earthscan Publications, London.

Sackville-Hamilton, N.R., J.M.M. Engels, Th.J.L. van Hintum, B. Koo and M. Smale. 2002. Accession Management: Combining or Splitting Accessions as a Tool to Improve Germplasm Management Efficiency. Technical Bulletin no. 5. IPGRI, Rome.

Tanksley, S.D., N.D. Young, A.H. Paterson and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–264.

IV. Plant genetic resources: benefits and implications of using molecular markers

Andreas Graner, Klaus J. Dehmer, Thomas Thiel and Andreas Börner
Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

About 100 years ago, the rediscovery of Mendel's principles of heredity turned genetics from a mystery into a serious science. By 1980, the deployment of DNA marker technologies had ushered in a new era in the field of genome analysis, which has culminated in the determination of the complete sequence of complex organisms, including higher plants. The rapidly expanding knowledge of the structure and function of genomes will increase our understanding of the role of individual genes and their orchestrated interplay in a cell, tissue or organism. Molecular genetics will also open up new avenues for studying genetic diversity to understand the dynamics of evolution and for using the genetic diversity currently locked in genebanks to improve cultivars.

The *ex situ* conservation of about 6 million accessions of PGR represents an essential contribution to the conservation of both intra- and interspecific diversity of crops and their wild relatives. The establishment and management of *ex situ* collections are complex, relying largely on empirical procedures. Hence, only circumstantial evidence has been gathered on the comprehensiveness of individual collections, as well as on the redundancy within and between collections. Similarly, the genetic integrity of individual accessions and changes in their genetic make-up have escaped closer examination. Moreover, systematic approaches for using genetic resources require extensive phenotypic evaluations, which are time consuming and expensive.

Many of the above-mentioned issues can be addressed in more detail by using information derived from DNA markers. Based on the current state of DNA marker technology, the present paper aims to highlight its potential impact, as well as its limitations, on managing and using genetic resources.

Marker-based characterization of germplasm

Hammer (2001) estimates that 7000 cultivated plant species exist, including their wild relatives. To provide *ex situ* conservation for all of them clearly exceeds the current capacities of genebanks. Hence, the dilemma of almost any genebank lies in finding a compromise between the number of species to be conserved (biodiversity) and the number of accessions of a given species to be kept (genetic diversity). As a result, most conservation efforts have focused on agriculturally important species. About one third of all *ex situ* accessions represent just 5 species—wheat, barley, rice, maize and *Phaseolus* beans—and the remaining two thirds cover only 30 species.

The relative overrepresentation of these agriculturally important species does not necessarily mean that their genetic diversity has been fully covered. Certain geographical regions are still not well represented in collections. By complementing geographic and ecological information, molecular marker data may help determine the extent to which accessions from diverse regions represent distinct samples (e.g. Ordon *et al*. 1997).

This approach, however, requires that marker data are available for reference from all existing accessions. For the 'top 30' crops, a substantial financial investment will be needed to generate the corresponding data sets. For example, fingerprinting the estimated 370 000 barley accessions (*Hordeum* ssp.) with 28 genetic markers (i.e. 2 markers per chromosome arm) would require €5 million (at an estimated cost of €0.50 per data point). This is a prohibitively large amount of money. However, with decreasing costs per data point and the potential spin-off effects (some of which are described below), the investment for systematically fingerprinting complete collections may be justified over the medium term.

Similarly, DNA marker data may provide invaluable information for taxonomic issues. The taxonomic determination of PGR is essential, for both their conservation and their use. Currently, it is mainly based on morphological descriptors and requires extensive expertise, particularly for intraspecific

resolution. Figure IV.1 illustrates how amplified fragment length polymorphism (AFLP) marker data provided important clues on the taxonomic status of several hitherto undetermined entries of the difficult *Solanum nigrum* complex (nightshade) (Dehmer and Hammer 2004).

Basically, a DNA marker-based taxonomy can be developed for any given genus to at least the species level. In many cases, congruency was shown between the DNA-based and classic systems, whereas, in other cases, the taxonomy has had to be revised according to DNA marker and sequence data. In such a context, DNA sequence and marker data are of particular value for understanding the phylogeny of polyploid species, as recently shown for the genus Hordeum (F. Blattner pers. comm.; El-Rabey *et al*. 2002; Pedersen and Seberg 2003).

In addition to taxonomic studies, the effect of plant breeding on the formation of genepools can be evaluated and quantified, as has been done for barley, resulting in European barley cultivars forming distinct groups of spring and winter types. The latter are further subdivided into two-rowed and six-rowed barleys (Thiel *et al*. 2003), a population structure that resulted from cross-breeding activities. In genetic diversity studies, major emphasis is given to the quality and quantity of DNA marker data, because insufficient marker numbers result in uneven genome coverage, which may yield unsatisfactory results. Finding congruencies between molecular marker data and classical

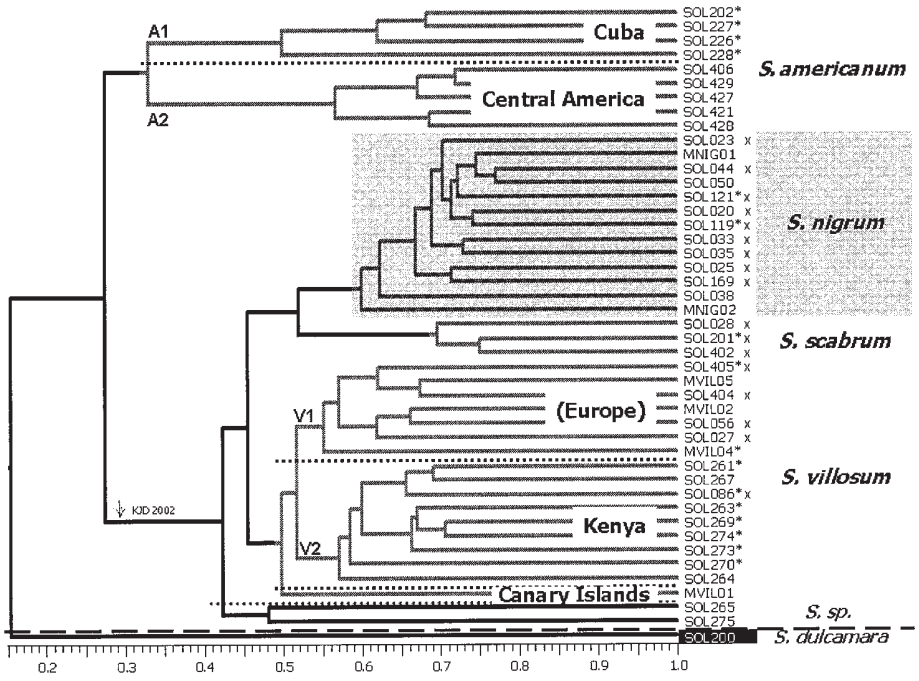


Figure IV.1. Marker-assisted taxonomy: a phenogram of 44 accessions of the *Solanum nigrum* (nightshade) complex shows four species clusters (right), and indicates cluster-specific areas of origin (left), based on data from two AFLP reactions and 523 fragment size classes. Outgroup SOL200 is separated by the broken line; dotted lines = intraspecific divisions; * = accessions taxonomically (re)classified; x = material of unknown geographic origin according to passport data.

taxonomy will also become difficult if several incompatible schools are in use, as is frequently the case.

Ex situ management of germplasm

The management of genetic resources includes all activities ranging from seed storage (or conservation in a vegetative state), through multiplication of seed to provision of genebank accessions on request. The genebank at IPK dispatches an annual average of 17 000 seed, plant and tuber samples, leading to a need for subsequent multiplication. Hence, about 5% of the seed collection is multiplied every year, translating into 7200 accessions that must be planted and monitored in the field or greenhouse. Meticulous precautions are undertaken to prevent contamination of accessions during multiplication, whether by use of particular agricultural practices, permanent control during the vegetative period, or establishment of herbarium collections. This last may serve by offering reference samples to check the authenticity of individual accessions.

Measures to check the authenticity of an accession are based on morphological characters, such as the descriptor traits that have been defined for many agriculturally important genera. Obviously, the descriptors to be recorded must be limited to a manageable number for each species. Also, the inheritance of many descriptor traits follows a monogenic inheritance (e.g. two rows vs six rows, long awns vs short awns or flower pigmentation), critically limiting genome coverage. Molecular markers, as tools for probing additional loci in a genome, can thus check for possible changes during multiplication that may otherwise, because of a lack of a visible phenotype, have gone undetected by morphological inspection in the greenhouse or field.

In a pilot study to check the quality of collection management of an inbreeding species, several accessions of the wheat (*Triticum aestivum*) collection at the IPK were fingerprinted with a set of simple sequence repeat (SSR, microsatellite) markers (Börner *et al.* 2000). No changes were detected in the accessions examined, which had been multiplied from 2 to 24 times over 50 years. The results underscored the efficiency of the precautions taken by the IPK genebank to preserve the genetic integrity of inbreeding collections.

However, SSR fingerprinting of a set of barley cultivars revealed unexpected differences between different accessions of identical cultivars (Figure IV.2). Because of the inbreeding nature of barley, identical genotypes are to be expected under the same cultivar name. Most of the 'duplicated'

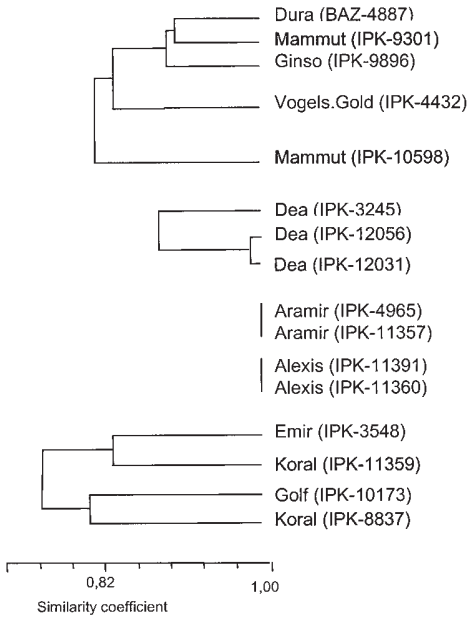


Figure IV.2. Discordance between cultivar designations and SSR-fragment patterns. A partial dendrogram has been extracted from a survey of 50 European barley cultivars and includes several duplicate genebank accessions, which initially were obtained from different donors. As expected for an inbreeding species such as barley, duplicate accessions of the cultivars 'Aramir' and 'Alexis' show 100% similarity. In contrast, duplicate accessions of the cultivars 'Mammut', 'Dea' and 'Koral' reveal discordant genotypes, indicating inconsistencies. Accession numbers are given in parentheses.

samples had been obtained from different donors, and seed lots may have been confused or wrongly named before the genebank received the samples. Although DNA markers help unveil such cases, identification within a set of homonymous accessions of the most original sample may remain difficult, although Lund *et al*. (2003) recently proposed a statistical approach.

The propagation of inbreeding crops is straightforward, but outbreeding species, both wind- and insect-pollinated, must be propagated as populations. An outbreeding population can be described in terms of its allelic frequencies, and the major objective of any conservation effort is to keep the genetic make-up of a population unaltered. To this end, insect-pollinated plants are grown in isolation chambers and wind-pollinated species in pollen-proof growth chambers or field plots that are sufficiently isolated from other accessions of the same species. Populations must be sufficiently large to prevent genetic drift. Environmental effects need to be eliminated to prevent selection.

Marker analysis of rye accessions, regenerated 2 to 13 times under standard conditions, revealed extensive shifts in allelic frequency (Chebotar *et al*. 2003, Figure IV.3). With some markers, a decrease, even loss, of alleles was observed, whereas with other markers, even new alleles were recorded, indicating pollen introgression from other populations. Principally, the extent of observed changes seemed to be a function of the number of multiplication cycles. Thus, molecular marker data provided clear hints for the need to revisit the conservation management of outbreeding species.

In some cases, the multiplication of a species is impaired by its lack of adaptation to the environmental conditions prevailing at the location of its genebank: for example, soils, occurrence of specific pathogens or pests, suboptimal temperatures, or inappropriate photoperiod. These problems can be alleviated by subcontracting seed increase to a collaborator or commercial partner with access to a more appropriate site. To monitor the subcontracting, DNA fingerprints of the samples first dispatched and those received after the increase is completed will provide the necessary documentation for authenticating the samples. DNA fingerprints may also

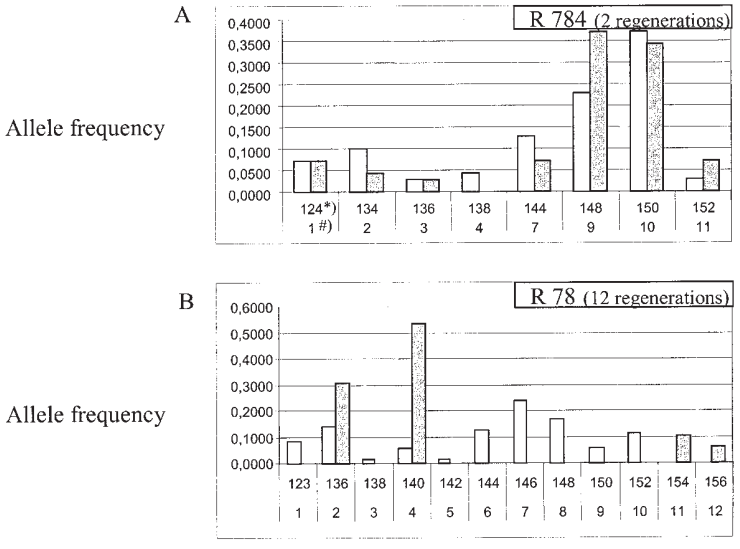


Figure IV.3. Genetic integrity of two rye accessions (R 784 and R 78) after 2 and 12 regeneration cycles, respectively, based on the allele frequencies of the rye microsatellite marker RMS18 (according to Chebotar *et al*. 2003).
* = allele size, #=number.

constitute a tool for monitoring multiplication within a genebank, which becomes a crucial issue once ISO certification or similar standards are to be achieved.

Despite the accuracy of DNA marker technology, two major questions must be solved:

- What percentage of bands should be identical before two accessions are identified as duplicates?
- What changes in allele frequency are acceptable in outbreeding populations?

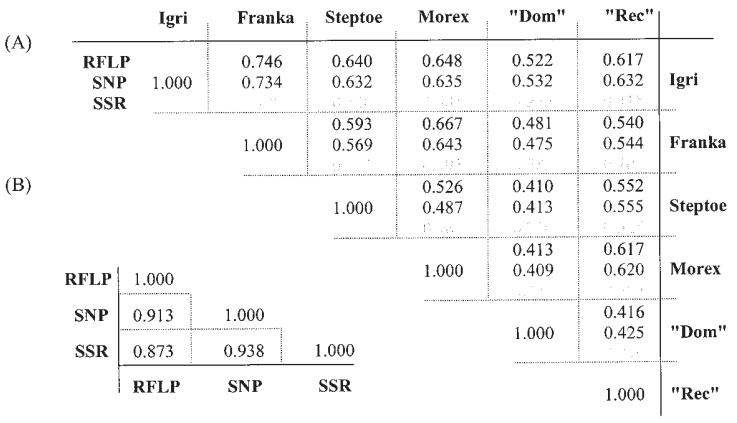
To find answers to these questions, pilot studies need to be performed to generate a database that will adequately allow the establishment of meaningful threshold values.

Technical aspects

Numerous molecular marker technologies are used for DNA fingerprinting in plants. To discuss the merits and demerits of any individual marker system is beyond the scope of this paper, but microsatellite and AFLP markers have been preferred in many diversity studies. Attempts have been made to fingerprint comprehensive collections of as many as 1000 accessions: for example, Huang *et al*. (2002) for wheat. However, because few attempts have been made to establish a reference marker set for use in parallel studies, the results of most marker studies for a given species cannot be readily compared and integrated.

Despite worldwide activity in DNA fingerprinting of *ex situ* germplasm, the integration of data in corresponding genebank documentation systems is so far insignificant. Nevertheless, efforts are being made to develop databases and software tools to visualize and analyse DNA fingerprinting data. In this context, fingerprinting data based on DNA fragments (e.g. AFLPs or SSRs) can be documented either as gel pictures or as tables containing the length of individual DNA fragments. For the latter, mandatory and extensive internal controls would be needed to obtain accurate estimates of fragment sizes for cross-referencing between experiments and laboratories.

The increasing availability of sequence information for various plant species enables direct analysis of the point mutations that give rise to single nucleotide polymorphisms (SNPs). Point mutations are the most frequent type of intraspecific DNA variation (polymorphism), and can be detected in more or less unlimited quantity. Together with insertions and/or deletions, other marker systems also detect point mutations but, because of technical limitations, only a tiny subset of the SNPs present between two genotypes can be detected with these systems.



(A)

	Igri	Franka	Steptoe	Morex	"Dom"	"Rec"	
RFLP		0.746	0.640	0.648	0.522	0.617	
SNP	1.000	0.734	0.632	0.635	0.532	0.632	Igri
SSR							
			0.593	0.667	0.481	0.540	
		1.000	0.569	0.643	0.475	0.544	Franka
				0.526	0.410	0.552	
			1.000	0.487	0.413	0.555	Steptoe
					0.413	0.617	
				1.000	0.409	0.620	Morex
						0.416	
					1.000	0.425	"Dom"
						1.000	"Rec"

(B)

	RFLP	SNP	SSR
RFLP	1.000		
SNP	0.913	1.000	
SSR	0.873	0.938	1.000

Figure IV.4. Equivalency of different, EST-derived marker types. (A) Pairwise genetic similarities (Dice) of 6 unrelated barley cultivars obtained from analysis with RFLP (253), SSR (632) and SNP (508) markers. (B) Correlation coefficients (Mantel's *r*) of the corresponding distance matrices are significant in all cases ($p > 0.01$).

To study the principal equivalency of SNPs identified in several barley genes with gene-derived SSR or RFLP markers, a set of six barley lines was analysed in parallel with all three types of markers (Figure IV.4). The results revealed their principal equivalence, because the correlation of the corresponding similarity matrices ranged from 0.87 to 0.93, being significant in all cases. Because SNPs can be described in an alphanumeric manner according to the four nucleotides, their documentation is simple and straightforward.

Given that the same SNP loci are being studied in different laboratories, the corresponding results form compatible data sets that can be combined and analysed in a decentralized process. The analysis of SNPs currently requires expensive detection platforms that are not available to all laboratories and may result in considerable costs per data point (Kota *et al*. 2001; Jenkins and Gibson 2002). However, the development of bioinformatic tools to facilitate exploitation of available DNA sequence databases is bringing down the costs of identifying SNPs. The technology will become more widespread. In addition, many of the SNPs that were mapped in the barley genome affect the recognition site of restriction enzymes and thus can be assayed as simple CAPS (cleaved amplified polymorphic sequence) markers, requiring a minimum of technical equipment.

With the present state of knowledge, SNP markers seem best for meeting the requirements for marker-assisted management of genetic resources. In the short term, these markers will be developed in amounts sufficient only for major crop species. Thus, the marker-assisted management of species that are less important in developed countries needs to be studied, using conventional marker systems. In these cases, however, intra-genebank management issues may be more important than inter-genebank issues: for example, intra-laboratory performance of a given marker system may be more important than inter-laboratory standardization and compatibility.

From genome diversity to gene diversity: a shift in paradigm?

Systematic fingerprinting of germplasm collections will provide additional knowledge in terms of molecular diversity and genetic relationships, both within and between genepools and at both genome and gene levels. Although DNA fingerprinting is usually performed at the genome level, gene-based strategies may help analyse collections for the presence of specific alleles. For example, barley collections may be searched for the presence of distinct alleles of the enzyme β-amylase, which differ in terms of thermostability (Malysheva *et al.* 2004).

In the future, collections may be screened for the presence of new alleles at a given locus. These alleles could later be assayed for their functional value. This approach would require the prediction of a gene's phenotype from its DNA sequence, a capacity that is still to be reached. However, recent advances in the analysis of linkage disequilibrium may help identify genes underlying traits of interest by association mapping (Rafalski 2002). This approach obviates the requirement for experimental populations, and genetic studies could be performed directly on the plant material available at a genebank. The time span from identifying a target gene to its deployment in a breeding programme might be reduced, thus further increasing the value of germplasm collections.

As a result, implementing analytical tools based on molecular markers may cause a shift in paradigm on the use of genetic resources: in the past genetic resources were used in terms of knowledge of the phenotype but, in the future, genebank collections will be increasingly searched for specific genotypes or even structural features of a specific gene. Genebanks may extend their services from mainly providing seed samples to providing DNA samples. Currently, however, implementation of marker-assisted germplasm management and use is on a minor scale. For successful large-scale implementation, major

interaction between the areas of bioinformatics and genebank documentation is needed to generate the required infrastructure to handle and deconvolute the large amount of data.

Despite the exciting potential of molecular marker technologies, several issues on technological impact assessment need to be considered:

- What are the social issues if marker technologies become available mainly for agriculturally important species?
- Does the additional funding required for deploying marker technology increase the gap between rich and poor genebanks?
- Who will take over the costs for improved management: donors or clients?

Notwithstanding these questions, DNA markers are about to make their way into genebank laboratories. The outcome, however, will depend mainly on the ability of individual laboratories to generate compatible data sets. This is both a chance and a challenge, which requires international coordination on the issues of a standardized marker system, standardized laboratory protocols and quality checks, and standardized data management. The adoption of appropriate guidelines will ensure that the added value of marker-assisted management of genetic resources will materialize, to the benefit of all.

References

Böner, A., S. Chebotar and V. Korzun. 2000. Molecular characterization of the genetic integrity of wheat (*Triticum aestivum* L.) germplasm after long-term maintenance. Theor. Appl. Genet. 100:494 497.

Chebotar, S., M.S. Röer, V. Korzun, B. Saal, W.E. Weber and A. Böner. 2003. Molecular studies on genetic integrity of open pollinating species rye (*Secale cereale* L.) after long term genebank maintenance. Theor. Appl. Genet. 107: 1469-1476.

Dehmer, K.J. and K. Hammer. 2004. Taxonomic status and geographic provenance of germplasm accessions in the *Solanum nigrum* L. complex: AFLP data. Genet. Res. Crop. Evol. 51: 551-558.

El-Rabey, H.A., A. Badr, R. Schäfer-Pregl, W. Martin and F. Salamini. 2002. Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. Plant Biol. 4: 567 575.

Hammer, K. 2001. Agrobiodiversity, plant genetic resources and ecological performance [in German]. Pp. 1 13 *in* Schriften zu genetischen Ressourcen (K. Hammer and T. Gladis, eds.). Zentrale für Agrardokumentation und Information (ZADI), Bonn.

Huang, X.Q., A. Börner, M.S. Röder and M.W. Ganal. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699–707.

Jenkins, S. and N. Gibson. 2002. High-throughput SNP genotyping. Comp. Funct. Genom. 3:57–66.

Kota, R., M. Wolf, W. Michalek and A. Graner. 2001. Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (*Hordeum vulgare* L.) Genome 44:523–528.

Lund, B., R. Ortiz, I.M. Skovgaard, R. Waugh and S.B. Andersen. 2003. Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor. Appl. Genet. 106:1129–1138.

Malysheva, L., M.W. Ganal and M.S. Röder. 2004. Evaluation of cultivated barley (*Hordeum vulgare* L.) germplasm for the presence of thermostable alleles of β-amylase. Plant Breeding 123: 128-131.

Ordon, F., A. Schiemann and W. Friedt. 1997. Assessment of the genetic relatedness of barley accessions (*Hordeum vulgare* s L.) resistant to soilborne mosaic-inducing viruses (BaMMV, BaYMV, BaYMV-2) using RAPDs. Theor. Appl. Genet. 94: 325 330.

Pedersen, G. and O. Seberg, 2003. Phylogenetic analysis of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. Syst. Bot. 28:293–306.

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

Thiel, T., W. Michalek, R.K. Varshney and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) Theor. Appl. Genet. 106:411–422.

V. Connecting plant germplasm collections and genomic centres: how to better link curators, molecular biologists and geneticists?

Serge Hamon[†], Emile Frison[‡] and Luis Navarro[§]

[†]UMR 1097, IRD-Centre de Montpellier, France, [‡]INIBAP, Parc Scientifique Agropolis, Montpellier, France and [§]IMA, Department of Plant Protection and Biotechnology, Mondaca, Valencia, Spain

The position of the curator: central but isolated

Most current germplasm holdings began in the 1960s as a result of breeders' demands and the development of genetic resources networks in western Europe, the USA, Australia, New Zealand and the USSR. During the 1970s, regional approaches, based on Vavilov's concept of centres of diversity, were adopted. An increasing concern in the international community about the need to conserve, for future generations, genetic resources that were in danger of disappearing also helped to establish several germplasm banks. A huge number of collecting missions were planned and constitutions of genebanks were established (Frankel 1974; Brown 1989).

In the early 1990s, there were technical and political revolutions. The discovery of the PCR was the starting point for rapid development of new molecular techniques in the genomics area. Political change began with the signing of the Convention for Biological Diversity (CBD) and the introduction of the 'Sovereign Rights' concept. Since 2000, the distance between the needs of end users and basic science programmes has progressively increased.

Currently, plant molecular tools are usually regarded as being able to improve the efficiency of breeding programmes but, more critically, form a link between different scientific and technical approaches. On one hand, many curators have little awareness of the possibilities offered by molecular genetics and genomics, and, on the other hand, these scientists are not aware that end users exist and need accessible data from them. This information exchange, between basic scientists and end users, is far from being a direct one-step process. Consequently, the collection curator, who is still in a central position, finds it difficult to link the two worlds.

The world of plant genetic resources (PGR) is criss-crossed with more or less efficient networks, developed during recent decades (Pistorius 1997). The genomic world has recently and rapidly set up genomic initiatives, although only in a few crops, and frequently focused on one particular aspect (Plant Physiology 2003).

With respect to plant germplasm management and use, there is no connection between basic scientists and end users such as farmers (Figure V.1). The curator, managing base and core collections, seems to be strategically placed, being located at the top of a hypothetical triangle with putative relationships with the other two extremes. In practice, however, the curator relates, more or less directly, with breeders and molecular geneticists but is distant from both farmers and basic scientists. Likewise, geneticists establish contacts in both genomics and genetic resources, but rarely relate with basic, or fundamental, scientists. Breeders, in their turn, tend to relate only with geneticists, curators and end users.

In reality, the information flow between molecular geneticists and traditional breeders is highly inefficient. This is not so much because of how the system is organized but because of the highly specialized nature of the different parts, and their different centres of interests and languages. The curator cannot acquire knowledge and

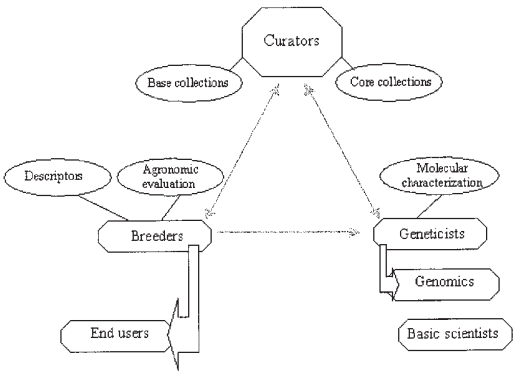


Figure V.1. The putative central position of the curator between basic science and end users. Arrows show tendencies of actual relationships.

understanding of either fast-moving area without appropriate help.

The traditional perception of the curator working only with plants and phenotypic evaluation is changing, or it should be. Firstly, more than one scientist may work in a germplasm bank, permitting the required specialization on new areas. Secondly, many germplasm banks are located in research institutions where scientists specializing in the new areas may be working and could collaborate with curators. Curators thus have, as a significant part of their job, the task of establishing close links with scientists of other disciplines to improve the maintenance and use of germplasm resources. Obviously, such activities are much easier for germplasm banks located in developed countries and, to a certain extent, in international agricultural research centres (IARCs). The objectives of germplasm banks should also change to integrate new technologies into their management.

Consequently, the current challenge facing the use of genetic resources remains focused on improving the management of genetic resources and their efficient use. But the curator must understand and integrate new data and technologies coming from molecular geneticists and biologists.

In this paper, we first summarize how PGR networks, genomic programmes and biological databases are organized. We then suggest improvements in organization, using the current examples of *Vitis* and *Musa* as illustrations.

Plant genetic resources networks

PGR networks vary in complexity from simple bilateral agreements to complex international networks that include different levels such as national networks, subregional programmes and regional networks (Frison *et al.* 2002). At the national level, PGR activities involve a wide range of people: policy-makers, scientists, universities, agronomy schools, breeders, rural communities and, in developing countries, NGOs.

In some countries PGR activities are supervised by a specific institution such as the Institute of Biodiversity Conservation and Research (IBCR) in Ethiopia. In other countries, the system is catalysed by a small group of scientists such as those located in the French Bureau des Ressources Génétiques (see the BRG Web site).* The BRG organizes discussions at the national level on the genetic resources of animals, plants and microorganisms. Recently, a project, developed with the participation of stakeholders, was written into the National Charter for the Management of Genetic Resources (BRG 1999). The BRG also organizes seminars and thematic conferences to promote scientific, socioeconomic and legal research in the field of genetic resources, and to facilitate the transfer of knowledge. Other countries, such as Spain, have national programmes to conserve PGR, involving all the country's research institutions.

At the subregional or regional levels, countries often share ecogeographic similarities and may have many crops in common. The clear benefits of collaboration between these countries include sharing conservation facilities and common objectives. For example, the Nordic Genebank (NGB Web site) has the mandate to conserve and document the genetic variation in Nordic plant material useful to agriculture. The NGB also aims to rationalize cooperation between Nordic countries wanting to use PGR for breeding and research. Stored material is made available for breeding, research and, in other countries, field use. At present, only some of the NGB's databases are directly searchable over the Internet, but effort is being made to provide all available information through an interactive interface.

* For details on this and other databases, see the respective entries in Web Site Addresses of Databases on page 43.

The NGB material is divided among three databases (NGB Web site):

- *Taxon Database*, which describes taxa within the NGB's mandate and species that are threatened or protected in Nordic countries
- *Culton Database*, which is an inventory database of commercial and primitive cultivars of taxa within the mandate
- *Accession Database*, which contains information about accessions in the NGB seed store.

Member states of the Southern African Development Community (SADC) established a similar system, known as the Plant Genetic Resources Centre (SPGRC Web site), as a non-profit intergovernmental institution. Its headquarters are near Lusaka, Zambia.

Another approach used in sub-Saharan Africa seeks to assist the region's countries to build up their capacities for research, conservation and use of both crop and forest germplasm through viable national programmes, subregional networks and selected crop networks such as those for coffee, coconut, *Musa* and yam. The project (details at the Wisard Web site) provides support for training and documentation of PGR activities at national and regional levels, and helps raise public policy awareness on the region's PGR issues. The strategy, implemented during the 1980s in West and central Africa, needs consolidating, considering that most countries lack the capacity to develop independent, fully functional systems.

At the moment, the most successful example of regional collaboration is the European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR Web site) in which 35 countries and 10 networks participate. Some networks are plant specific such as the Cereals Network, Forages Network and Fruit Network, whereas others are mostly thematic such as the Documentation and Information Network, *In situ* and On-farm Conservation Network, and Inter-regional Cooperation Network.

The ECP/GR networks are concerned with long-term *in situ* and *ex situ* conservation of PGR, and increasing the use of PGR in Europe. Data for five types of descriptors are collected: passport, management, environment and site, characterization and evaluation. Recommendations are made to produce information that closely follows the descriptor lists in terms of ordering and numbering descriptors, and using specified descriptors and recommended descriptor states. These descriptors lists, however, were established without reference to molecular data. None of the programme's eight missions focuses on integrating and using molecular and/or genomic data. Consequently, the networks, even though they are very well organized and efficient, do not directly relate to genomic data.

International crop networks are certainly the best way to bring together specialists from different fields. The European Barley Database (EBDB Web site) contains 92 000 accessions from 36 institutions in 29 countries. One of the first world core collections, the International Barley Core Collection (BCC), was constituted from data available in the 1990s (van Hintum 1995).

The importance of PGR centres, networks and databases as essential elements in the collaboration and effective use of genetic resources is evident. However, they pay little attention to the new areas of molecular markers and genomics. One reason for this is that these networks were mostly set up during the 'golden' period of genetic resources—the 1960 to the 1980s—before the molecular sciences took off.

Another reason is that, during the 1990s, in terms of the broad objective of improving tools for analysing genetic diversity, the technical race was fast and seemingly endless. For example, PCR-derived techniques evolved extremely quickly, from RFLPs in the early 1990s to single sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) in the late 1990s. During that period, other

markers such as random amplified polymorphic DNA fragments (RAPDs) and amplified fragment length polymorphisms (AFLPs) came and went, inducing a degree of complexity for users and uneasiness among curators and the scientists involved in genetic resources studies.

During the 1980s and 1990s, lists of morphological and agronomic descriptors were published without reference to common sets of molecular markers. Base collections or subsets of collections were evaluated with highly pertinent descriptors, and databases were built up. In contrast, the samples studied with different types of markers for purposes such as phylogeny, population structure and plant reproductive behaviour were tiny and inadequately standardized, and the data poorly stored. In addition, some types of markers (i.e. RAPDs and AFLPs) were not easily transferred from one laboratory to another or were used only on small sets of genotypes.

Genomic programme organization

The potential usefulness of genomics arose through the idea that genetic diversity could be used beyond specific boundaries. In practice, genomic programmes could be divided into at least six categories (Figure V.2): (a) developing efficient markers and high-density maps, (b) assessing the relationships between genetic and chromosome maps, (c) elaborating physical maps, (d) studying gene expression, (e) conducting functional analyses of genes and (f) bioinformatics. Until recently, only category (a) was taken into account. Now, genetic diversity can and must be observed, not only at the DNA level, but also at the expressed level. Categories (a), (d) and (e) are therefore of major interest to curators, who could broaden their knowledge through networking.

Many molecular markers have been produced to analyse genetic diversity. Their characteristics and availability have dramatically evolved during the last decade. Currently, codominant markers are preferred, with microsatellites being of particular interest (Goldstein and Schlötterer 1999). Such markers are easy to use, can be extrapolated to closely related species or genera, and, through sequencers, can be readily used automatically.

Another family of markers of increasing interest is expressed sequence tags (ESTs). EST programmes permit identifying families of genes expressed under different conditions, such as drought resistance, cold resistance or pathogen resistance, and help explain the differential behaviour of different genotypes at a higher integrative level. Developing these markers is more expensive than for the others mentioned, necessitating sharing and exchange through networks and consortiums to reduce redundancy and optimize budget use.

High-density maps are constructed on the basis of recombination frequencies of chromosomal markers, either genes or DNA markers (e.g. RFLPs and SSRs). In applied genetics, the breeder may benefit by knowing links with neutral or adaptive characters. Knowing recombination frequencies may facilitate planning of population sizes for selection. A reference population must be prepared (i.e. crossing between distinct genotypes) and enough progeny must be available (a minimum of 200 individuals). Where possible, such

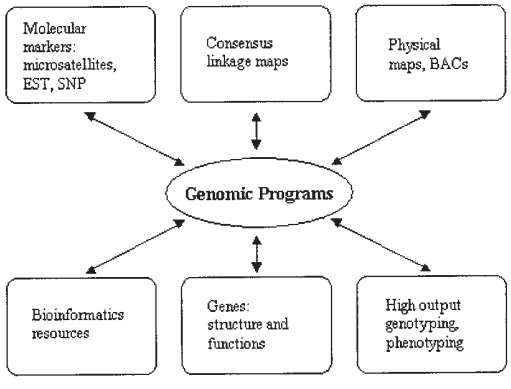


Figure V.2. Organization of genomic programmes.

populations must be duplicated and their references kept in a unique place. These populations could be sent to any user, either as plants or as DNA. The developed markers would be SSRs or SNPs.

For any one crop, the minimum reference needed is a saturated map where all linkage groups are identified and 200–300 markers are available. Combinations of desirable genes are most helpful for introgressing characters from distant species. Syntenic groups of genes are often observed. For example, comparing genetic maps within *Poaceae* was greatly facilitated by conserving, in order, markers and genes along chromosomes (Devos and Gale 2000).

Curators are increasingly interested in having information on the location of markers in linkage groups. The main problem is being able to compare different maps. Hence, establishing consensus maps is frequently the top priority of networks. When a group finally obtains a consensus map and the different marker types, the next difficulty is to build up a database that is accessible to every member of the consortium, preferably through the Web, using access names and passwords. Finally, to develop markers, coordination of activities is a must.

New programmes tend to search for relationships between physical and recombination maps. Distribution of genes along chromosomes is not uniform, with no direct correlation between recombination and physical maps. In such a context, identifying a genotype of particular importance or having specific attributes is essential, that is, researchers and teams must identify and develop particular genotypes (e.g. dwarf architecture or short seeding cycle).

Such reference maps must each be elaborated in only one place. The bacterial artificial chromosome (BAC) library varies, according to the precision required, from 5 to 20 copies of the genome equivalent. The higher the number of copies, the more precision generated, but also the more expensive and time consuming. If the goal is to prepare genome sequencing or anchor specific genes or ESTs, the number of copies required can be changed. Integrated collaboration between teams is essential for success.

To summarize, a standard genomic programme has a minimum of six activities:

- developing efficient markers and high-density maps
- assessing the relationships between genetic and chromosome maps
- elaborating physical maps
- studying gene expression,
- conducting functional analyses of genes
- bioinformatics.

Without data harmonization and intergroup collaboration, using the products is nearly impossible. Major crops now have several thousands of EST sequences coming from different plant organs or developmental stages. If these sequences are to be used, they must be structured into efficient and available databases. Over the last two decades, databases have become essential resources for biologists worldwide.

Molecular biological databases

Molecular biological databases have emerged, but mostly unrelated to the world of PGR, largely because this revolution was initiated through the Human Genome Initiative. This initiative produced new databases such as the human gene prediction database maintained by the European Bioinformatics Institute (EBI Web site). This non-profit academic organization forms part of the European Molecular Biology Laboratory (EMBL Web site), and is a centre for research and services in bioinformatics. The EBI manages databases of biological data, including nucleic acid and protein sequences and macromolecular structures. Its mission is to ensure that the growing body of information from molecular biology and genome research is placed in the public domain and is freely accessible to all sectors of the scientific community, thus promoting scientific progress.

EBI's goals are to build, maintain and provide biological databases and information services to support data deposition and exploitation. The best known is the EMBL Nucleotide Database, Europe's primary collection of nucleotide sequences. EBI collaborates with GenBank of the National Center for Biotechnology Information, Bethesda, MD (NCBI Web site), and the DNA Data Bank of Japan (DDBJ Web site) to contribute to ENSEMBL (ENSEMBL Web site). This database provides up-to-date data on completed genomes and the best possible automatic annotation.

Other EBI databases include Swiss-Prot, which provides complete annotated protein sequences; the Macromolecular Structure Database, a European project to manage and distribute data on macromolecular structures; and ArrayExpress for gene expression data.

Other initiatives were set up for the sequencing of microorganisms and animal models such as *Drosophila melanogaster* or *Caenorhabditis elegans*, leading to, respectively, the FlyBase Consortium and the WormBase Web sites.

Despite their different functions, all these databases consist of three tiers of software: a database management system, database access software and Web server, and the Web browser. Even so, this is not sufficient to have an integrated system. Some databases do not recognize orthological relationships (i.e. homologous genes derived from speciation by vertical descent). Others do, but do not integrate map positions. The diverse databases reflect the expertise and interests of the groups maintaining them.

Among the problems reported is the clash of concepts that users come up against as they move from one database to another. This can be important: for example, the allele concept varies in meaning according to scientific community; it can mean 'any genomic variant, including parts located outside genes' or 'the variant that changes only genes'. Another severe limitation is the continual change that occurs as new data types are added, and fields and nomenclature are altered. Recently, considerable effort was made to create the Gene Ontology Database (see Web site), which currently holds three ontologies—'structured, controlled' vocabularies that aim at precision and consistency in definitions of gene functions, processes and terms. Such vocabularies help facilitate integration.

To overcome problems of integration, several approaches have been suggested, including that of 'knuckles and nodes' (KN) suggested by Stein (2003). In this system, the source databases, as we have now, form the nodes, each of which uses a distinct and independent data model. These nodes are richly detailed. In contrast, the 'knuckles' are carefully maintained curator services that provide the information needed to relate data from one node to another. They are restricted to a single task and constrained to use standard interfaces. For example, one 'knuckle' may service only orthological relationships. It will also ensure that the symbols used for a given species are translated into those used for another.

Although awareness and concern for post-genomic possibilities are increasing, plant genomics remains underdeveloped. Only genomic programmes on model plants have developed applications (Cronk 2001). For example, The Arabidopsis Information Resource (TAIR Web site) is a database for *Arabidopsis thaliana* (Brassicaceae), which structures information into four categories: advanced search; analytical tools; external links; and *Arabidopsis* information. Each category encompasses several smaller databases (TAIR Web site).

For other crops, such as tomato, potato, barley and maize, The Institute for Genomic Research (TIGR Web site) centralizes gene indexes and offers database searches according to nucleotide or protein sequence, identifier, gene product name and EST annotations. A plant family integrated project is exemplified

by the Gramene Initiative (Web site). Its objective is to provide comparative mapping resources for monocotyledons. Database searches include the rice genome browser; rice blast; maps; markers, proteins; phenotypes; ontology; literature, including a capacity to download physical maps; genetic maps; *in silico* data; microsatellites; phenotype data; and protein databases.

How can curators be helped to use the new information? An application of the KN approach for connecting PGR is suggested in Figure V.3. The right arm of the figure shows IGRCs being largely managed by the biological database system under bioinformatics supervision. The aim is to link international genomic resources centres to national plant genomic programmes and non-plant genomic research (e.g. microorganisms, worms or flies). The figure's left arm shows management being developed by the international PGR system under IPGRI's supervision. It is necessary to develop specific concept and tools to better link those two sides.

From theory to practice: examples of *Vitis* and banana initiatives

Although an idealized situation could be proposed for any given plant (Figure V.4), no clear example of such development currently exists, with the closest being the *Vitis* and *Musa* initiatives.

The *Vitis* initiative

The French *Vitis* collection, located at the Institut National de la Recherche Agronomique (INRA Web site) in Vassal, southern France, is internationally significant for this genus, and is very well documented. The collection contains about 7500 accessions, corresponding to 210 species originating from 35 countries. The collection includes 4850 *Vitis vinifera* corresponding to 3000 *cépages* (or vine types) and 1300 hybrids. Each accession is represented by five plants. Each *cépage* is carefully described according to a descriptor list, and its genetic history is elucidated, using molecular markers (Bowers *et al*. 1999). The team is now developing an automatic system for a high-output genotyping system.

The genome size of *Vitis vinifera*—about 480 Mbp—is similar to that of the rice genome. To develop *Vitis* genomic resources, the international scientific community has decided to create the International Grape Genome Program (IGGP Web site) to optimize technical and

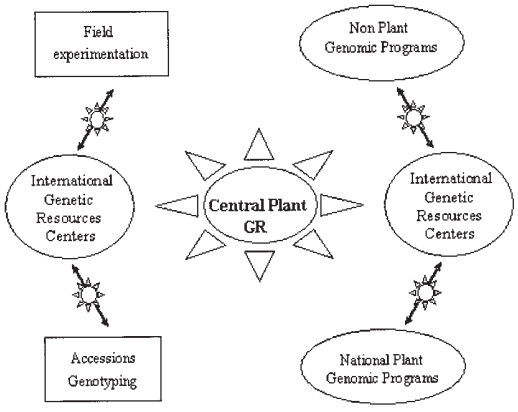


Figure V. 3. Connections between existing structures, using the 'knuckles and nodes' approach. PGR = plant genetic resources.

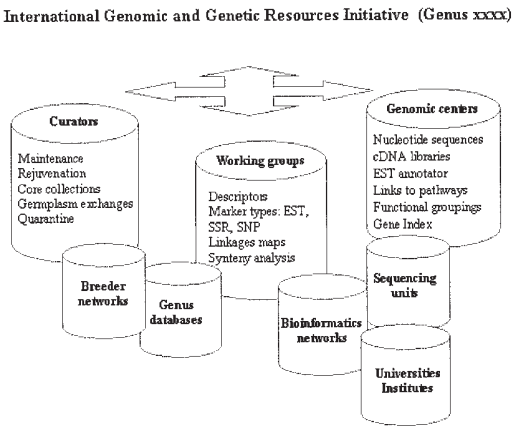


Figure V. 4. Proposed network for the International Genomic and Genetic Initiative. Large cylinders refer to activities that are more or less directly connected to plant genetic resources; small cylinders refer to activities with external groups and/or facilities.

financial resources. Coordinated by an Australian scientist, the programme, as defined by IGGP, is organized into five main activities:

- development of a high-density genetic map (Italy)
- development of a physical map (France)
- study of expressed genes (Australia)
- analysis of functional genes (Germany)
- development of bioinformatic resources (USA)

In Europe, the IGGP is strongly connected with genetic resources collections, and the research teams, involving 19 participants, are working on this plant through the European *Vitis* Database (Web site).

The high-genetic density map will be developed on a reference hybrid population (Cabernet S × Riesling). So far, 200 descendents are available, and the number is increasing. A reference population was created and is kept in the USA. Duplicates will be sent to genetic resources centres, and DNA samples are available for all teams wanting to participate in increasing the number of molecular markers. For the first round, microsatellites were selected because they are codominants and easily manageable with automatic genotypers. Currently, 200 microsatellites have been located on the map (Riaz *et al.* 2003). Through the Genoplante Initiative (Web site), INRA has sequenced a new set of 200 microsatellites.

Three teams are now involved: the Istituto Agrario San Michele (ISMAA), Italy; the Institute for Grapevine Breeding (IGB), Geilweilerhof, Germany; and INRA—Vigne (Web site). They are also developing other sets from ESTs. The first development towards the physical map was selecting for a 13X Cabernet BAC library using three restriction enzymes. The first markers and/or genes will then be anchored onto BAC clones to connect the genetic and physical maps. The BAC clones will then be fingerprinted.

For expressed genes, different teams started years ago to sequence ESTs. Currently, about 30 000 *Vitis* sequences are available in the genebank. Discussions have started on preparing 10 000 gene microarrays, using data from different international groups: the University of California—Davis, University of Nevada—Reno (UNR) (both in the USA), INRA, ISMAA and IGB. Collections of full-length cDNA, the foundation of functional gene analysis, were initiated with two projects, one at UNR and the other at the French Genopole System (Web site).

Workers at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia have selected a dwarf genotype of *Vitis* that is homozygous and has a short biological cycle. This grape will be used by the network. INRA—Colmar has also produced near-isogenic genotypes derived from a black Pinot. At the UNR and INRA—Montpellier, proteomic and metabolomic platforms are being developed. In Australia and France, research programmes are under way to analyse natural mutants found in collections. To regroup all this information, a specific Web page, covering recent and available molecular data, has been created by TIGR (TIGR Vitis Web site). To harmonize data, a bioinformatic seminar was organized in 2003 at UNR (UNR Web site).

Networking in *Musa* genomics

An initiative to apply genomics technologies to the sustainable improvement of banana (*Musa* spp.) was launched in July 2001. Researchers from the world over came together to form a consortium. Their aim was to develop freely accessible resources for *Musa* genomics and use new knowledge and tools to help improve the crop through both targeted conventional breeding and transgenic strategies. The end result will be new high-yielding *Musa* varieties that will respond to local needs and ever-changing environmental challenges.

Deciphering the banana genome is an enormous task that requires the full participation and collaboration of many scientists. The Global *Musa* Genomics Consortium (Web site) functions under the guidance of a management committee within the framework of

the Global Programme for *Musa* Improvement (PRO*MUSA*). The International Network for the Improvement of Banana and Plantain (INIBAP Web site) functions as secretariat, and assumes the responsibility for external communications, including the development of an Internet portal that makes current results and information available to the wider world.

The *Musa* Consortium brings together expertise from 28 publicly-funded institutions from 15 countries. As well as providing close collaboration, all members agree to share materials and resources, including sequence data and enabling technologies. The sequences produced by the Consortium are placed in the public domain and any new varieties are made freely available to smallholders.

INIBAP set up an e-mail list server for the *Musa* Consortium to encourage members to exchange information. A *Musa* Genomic Resources Centre was created, and resources such as BAC and cDNA libraries are being made readily available to members of the Consortium. The protocol for producing transformed plants is well established and is now applied for promoter tagging. The Consortium is also developing a strong bioinformatics component. The genomic data assembled and analysed will also be made available in a user-friendly fashion through a Web-based integrated *Musa* information portal, together with other relevant public knowledge on *Musa*.

The Consortium's overall strategy is to adopt a stepwise approach, focusing on comparative genomics and targeting early gene discovery. As a monocotyledon that is taxonomically distantly related to rice, *Musa* is ideal for studying synteny between distantly related species. Indeed, *Musa* is now being recognized as a powerful model for studying fundamental aspects of plant genomes. The Global *Musa* Genomic Consortium is attracting new partners who, through their more upstream research, will provide considerable information and results that will greatly strengthen the Consortium's more applied-oriented research goals.

Conclusions and recommendations

Genomic and molecular markers projects are quickly developing new tools that have the potential to greatly improve the management and use of PGR. However, in practice, applying the new developments for routine use on a large scale in germplasm banks is often very difficult for the following reasons:

- Availability of permanent scientific staff is often restricted to the curator, who usually specializes in phenotype characterization of accessions—still a priority task for germplasm banks.
- Difficulty in using, or even being aware, of new genomic technologies. Molecular markers come and go very quickly, making the selection of marker types for application to all accessions of a given crop very difficult. A task lasting several years may finish with data considered obsolete by the rest of the scientific community.
- Molecular markers are expensive to use and often beyond the resources of most germplasm banks.

To solve these problems is not easy, and will require a multi-pronged approach. Curators are still in a strategic position to link research activities between geneticists and breeders, thus approaching basic science to end users. However, their traditional role of working only with plants and phenotypic evaluation will have to change to include activities oriented towards establishing close ties of cooperation with scientists of other disciplines. Good examples are the Grape Genomics Consortium and the Citrus Genomic Functional Project in Spain (Genomica Web site) where the germplasm banks have been playing an important role since the project's beginning, thus guaranteeing that results will have direct application to the better management and use of genetic resources.

While getting geneticists involved in developing and using molecular markers for

small sets of accessions for different research purposes is relatively easy, getting them involved in the routine application of markers to characterize the entire germplasm bank is very difficult. Hiring permanent scientific staff specialized in these new areas would help solve this problem, but such an action requires a significant increase in the budget, which only rich germplasm banks can afford.

An alternative approach would be to establish regional networks, and even worldwide networks, with the specific objective to characterize the germplasm of specific crops through molecular markers. Such networks would facilitate the use of the most appropriate tools, reduce costs for individual parties and improve strategies for conservation and use. Obviously, intellectual property rights will have to be taken into account in these possible networks. International agricultural research centres such as IPGRI are in the best position to take these types of initiatives.

References

Arabidopsis: plant system biology. 2003. Plant Physiology, special issue 132(2).

Bowers, J.E., Boursiquot, J.M., This, P., Chu, K., Johannson, H. and Meredith, C.P. 1999. Historical genetics: the parentage of Chardonnay, Gamay and other wine grapes of north-eastern France. Science 185:1652-1565.

BRG (Bureau des Ressources Génétiques). 1999. Charte Nationale pour la Gestion des Ressources Génétiques [*National Charter for the Management of Genetic Resources*]. Paris.

Brown, A.H.D. 1989. The Use of Plant Genetic Resources. Cambridge University Press, UK.

Cronk, Q.C.B. 2001. Plant evolution and development in a post-genomic context. Nature Rev. Genet. 2:607-619.

Devos, K.M. and Gale, M.D. 2000. Genome relationships: the grass model in current research. Plant Cell 12:637-646.

Frankel, O.H. 1974. Genetic conservation: an evolutionary responsibility. Genetics 78:53-65.

Frison, E.A., Mitteau, M. and Sharrock, S. 2002. Sharing responsibilities for *ex situ* germplasm management. Plant Genet. Resour. Newsl. 131:7-15.

Goldstein, D.B. and Schlötterer, C. 1999. Microsatellites: Evolution and Applications. Oxford University Press, UK.

van Hintum, Th.J.L. 1995. Hierarchical approach to the analysis of genetic diversity of crop plants. Pp. 23-34 *in* Core Collections of Plant Genetic Resources, 1st ed. (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). IPGRI, Rome.

Pistorius, R. 1997. Scientists, Plants and Politics: A History of the Plant Genetic Resources Movement. IPGRI, Rome.

Riaz, S., Dangl, G.S., Edwards, K.J. and Meredith C.P. 2004. A microsatellite marker based framework linkage map of *Vitis vinifera* L. Theor. Applied Genet. 108(5):864–872.

Stein, L.D. 2003. Integrating biological databases. Nature Rev. Genet. 4:337-345.

Web site addresses of databases

Bureau des Ressources Génétiques (BRG)	http://www.brg.prd.fr
Cornell University	http://www.aem.cornell.edu/
African Food Security and Natural Resources Management Doctoral Training Program	special_programs/AFSNRM/RF/
DNA Data Bank of Japan (DDBJ)	http://www.ddbj.nig.ac.jp/
European Barley Database (EBDB)	http://www.barley.ipk-gatersleben.de/ebdb
European Bioinformatics Institute (EBI)	http://www.ebi.ac.uk/
ArrayExpress	http://www.ebi.ac.uk/
EMBL Nucleotide Database	http://www.ebi.ac.uk/embl/index.html
Macromolecular Structure Database	http://www.ebi.ac.uk/msd/index.html
Swiss-Prot	http://www.ebi.ac.uk/swissprot/index.html
ENSEMBL	http://www.ensemb.org/
European Cooperative Programme for Crop Genetic Resources Networks (ECP/GR)	http://www.ecpgr.cgiar.org
includes Cereals Network, Forages Network, Fruit Network, Documentation and Information Network, *In situ* and On-farm Conservation Network, Inter-regional Cooperation Network	http://www.ecpgr.cgiar.org/Networks/
European Molecular Biology Laboratory (EMBL)	http:// www.embl.org/
European *Vitis* Database	http://www.Genres.de/CF/eccdb/vitis
FlyBase Consortium	http://flybase.bio.indiana.edu/
French Genopole System	http://www.genopole.org/
Gene Ontology Database	http://www.geneontology.org/
Genomica	http://www.genomica.ibmcp.upv.es/
Genoplante Initiative	www.genoplante-info.fr
Global *Musa* Genomics Consortium	http://www.promusa.org/genomics/model.htm
Gramene Initiative	http://www.gramene.org/
International Grape Genome Program (IGGP)	http://www.vitaceae.org

International Network for the Improvement of Banana & Plantain (INIBAP)	http://www.inibap.org/
IPGRI	http://www.ipgri.cgiar.org/
Institut National de la Recherche Agronomique (INRA)	http://www.inra.fr/
includes Diversité et Génomes des Plantes Cultivées Mixed Research Unit (DGPC) and Institut de Recherche pour le Développement (IRD)	www.dgpc.org/ressourceshumaines/liste_personnel_this.html
INRA Vigne	http://www.inra.fr/gap/departement/espèces/vigne.html
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/Genbank/
GenBank	http://www.ncbi.nlm.nih.gov/Genbank/
Nordic Genebank (NGB)	http://www.ngb.se/
includes Taxon Database, Culton Database, Accession Database	http://www.ngb.se/Databases/
Plant Genetic Resources Centre (SPGRC)	http://www.ngb.se/sadc/spgrc.html
The *Arabidopsis* Information Resource (TAIR)	http://www.arabidopsis.org/
includes:	
Advanced search:	
Genes, proteins, markers, germplasm, ecotype, polymorphism/allele, people/labs, publications	http://www.arabidopsis.org/servlets
Sequences, GO Annotations, locus history, microarray	http://www.arabidopsis.org/tools/bulk/
Analytical tools:	
SeqViewer, MapViewer	http://www.arabidopsis.org/servlets
AraCyc pathways	http://www.arabidopsis.org/tools/aracyc/
BLAST	http://www.arabidopsis.org/blast/
WU-BLAST2	http://www.arabidopsis.org/wublast/index2.html
FASTA	http://www.arabidopsis.org/cgi-bin/fasta/nph-TAIRfasta.pl
Chromosome map tool	http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp
External links: Stock centres, insertion, knockout and other mutations, nomenclature, sequence analysis, genome databases, proteome resources, microarrays	http://www.arabidopsis.org/links/

Arabidopsis information: about *Arabidopsis* genome initiative, functional genomics, gene expression, education & outreach, gene symbol list, ontologies, data submission, protocols and lab manuals	http://www.arabidopsis.org/info/
Monsanto SNPs and Ler	http://www.arabidopsis.org/Cereon/index.html
The Institute for Genomic Research (TIGR)	http://www.tigr.org/
TIGR *Vitis*	http://www.tigr.org/tdb/tgi/vvgi/
University of Nevada—Reno bioinformatics seminar (UNR)	http://www.ag.unr.edu/GBC/default.htm
Wisard—African network	http://www.wisard.org
WormBase	http://www.wormbase.org/

VI. Capacity-building and training

Theresa Fulton and Stephen Kresovich
Institute for Genomic Diversity, Cornell University, Ithaca, New York, USA

The changing role and increasing potential of genebanks in the 'genomics age'

Recent advancements, particularly in the fields of molecular and population genetics, biotechnology and genomics, have greatly expanded the potential use, quality and impact of genebanks. Since the inception of genebanks, with their aim to formally acquire germplasm resources, their main goal was to conserve and store these resources for current and future use. Curators of these genebanks did not necessarily require a high level of scientific research knowledge, and the user community consisted mainly of farmers and plant breeders wanting various accessions. This situation began to change, however, with the advent of molecular techniques during the 1980s and even more so with the new genomics tools during the late 1990s.

Among the many benefits that these techniques brought was the comparative ease with which scientists could incorporate genes from wild relatives into new cultivated varieties. This led to greatly increased demand for accessions that had previously been seen as undesirable or even useless (Tanksley and McCouch 1997). Such increased demand for germplasm resources has been exacerbated by the global food security situation and attempts to alleviate hunger through crop improvement.

Concurrently, the user community became more diverse, ranging from plant breeders wanting to screen material to molecular geneticists looking for alleles for introgression and crop improvement, and now, even to genomics scientists wanting to sequence additional alleles of genes of interest. Although not always cost-effective, new DNA technologies have also made possible the examination and comparison of accessions in a genebank at the molecular level, identifying precise genetic differences or redundancies among accessions. DNA sequencing could, one day, make knowing the entire sequence of each individual in a collection possible, thus establishing a true 'genebank'.

Thus, the plant germplasm resource community is expecting more and is asking questions on handling the increasing demand for germplasm resources. The community is also querying whether genebank curators should be taking advantage of the new techniques to improve or streamline the collections. Answering these questions, however, is compounded by decreases in funding for many genebanks, the difficulties curators face in learning or even keeping abreast of the quantity of new technologies, and the complexities of selecting those technologies most likely to answer specific biological questions.

The widening technological divide

These new technologies often come with a high initial price tag, primarily associated with acquisition of equipment. In addition, the fast pace of new developments related to these technologies requires frequent retraining of personnel and upgrading of equipment and software. Both factors make the technologies inaccessible to most developing countries, leading to the unfortunate effect of *increasing* the technological divide—the gap between the technical capabilities of developing and developed nations—at a time when much effort has been invested to *decrease* it.

Louise Fresco, Assistant Director General of the Food and Agriculture Organization of the United Nations (FAO), warns of an increasing 'molecular divide' between between developed and developing nations, meaning that the promise and potential of new technologies are not being shared equally (Fresco 2003; Northoff 2003). Developing countries are not able to take advantage of the full range of biotechnology tools to harness the value of their genetic resources. Fresco fears that biotechnology could actually aggravate current global inequalities unless something is done to bridge this gap (Fresco 2003).

While eliminating or at least minimizing the technological or molecular divide is laudable, attempts to do so have been, to date, largely unsuccessful. A major reason for this

is the lack of comprehensive training. Isolated training of a few individuals who are then sent back to their home laboratories to put their newly learned techniques into practice, alone and unsupported, is ineffective. Data are generated and, unfortunately, not analysed. Even more disappointingly, little of this information is used to improve the quality or use of collections.

Another challenge in bridging the technological divide is another apparent 'divide' in the perception of how biotechnology should be used in this task. On one side of the perception divide are many researchers who feel that cutting-edge technology is an absolute must for conducting top scientific research. On the other side are those that feel that biotechnology is simply the latest 'toy', unaffordable and inaccessible to all but the wealthy few. In fact, biotechnology is neither—rather, it is one of the many tools available to those working in the field of conserving plant genetic resources. Biotechnology should never be thought of as an end in itself, but as something that can be efficiently targeted to solve real curatorial or user needs.

The need for comprehensive global training

Comprehensive training is the foundation on which to bridge the technological divide *and* effectively use biotechnology in germplasm resource management. Although necessary in some instances, buying new equipment and installing new facilities in developing countries is not enough. Also essential are researchers who can think critically and independently about the objectives of their research programmes and the biological questions being addressed. The goal of training programmes is to produce such researchers.

Researchers and curators must be able to make the best choice of strategy and technology for each particular biological question. They must be able to interpret the data that they generate and understand how best to use the knowledge gained. Thus, training cannot be limited to instructions on how to use new equipment and follow new protocols.

For long-term effectiveness, training must emphasize basic scientific concepts in biology, genetics, genetic resources management, experimental design, data analysis, statistics and genomic sciences. This training should begin at, but not be limited to, the graduate student level. However, given the speed at which technology and biological sciences are moving and the pace that data are being generated, training cannot stop on the receipt of an advanced degree. New scientists must receive continuing education and ongoing support, especially those in developing countries who may not have access to adequate local support networks.

Continuing education can be made available through many mechanisms. An example is the training materials available online from IPGRI's Web site and through such publications as Karp *et al.* (1997). IPGRI and the Institute for Genomic Diversity (Cornell University) have collaborated to produce a new training module on molecular markers, soon to be available online, which will be continually updated to include new technologies and information on, for example, differences in cost between the many molecular marker techniques (de Vicente and Fulton 2003).

The Internet is a good medium for people who cannot afford to travel and thus cannot access institutions that host molecular science training workshops and other learning experiences. However, many studies have shown that most deep learning, that is, learning that includes true understanding and not mere memorization of facts and protocols, occurs only in hands-on sessions (Deboer 1991; Lederman 1992; AAAS 1993). Thus, where practical, training must be conducted through hands-on laboratory workshops rather than through books, lectures, or the Internet.

Ongoing support must also be available to help scientists implement the concepts and

techniques that they have learned. Feedback, including from responses to survey questions, indicates that a key barrier to effective impact of many training programmes comprises the problems encountered when trying to put newly learned concepts and protocols into real practice at the home laboratory. One way of overcoming this barrier is to encourage the development of 'cohorts', people with common backgrounds, training and interests who can be called upon when questions and problems arise.

A Rockefeller-funded programme at Cornell University is establishing the African Food Security and Natural Resources Management Doctoral Training Program (see their Web site). The goals of this project include training interdisciplinary teams to conduct research, providing education on topics relating to agricultural productivity in Africa (especially soil degradation problems) and examining how to encourage African scientists to continue these activities once they return home (AfricaGrant 2001). So far, the programme has gone very well: six of the eight students have already passed their preliminary exams (at the time of writing, the other two were scheduled for examinations) and one student won a Heinz award. These students will soon be returning home to Kenya to do their field work, where they will receive some follow-up supervision and encouragement to network among themselves (Alice Pell, pers. comm.).

For germplasm curators, a community of other curators with whom to discuss current issues and coordinate efforts is especially important. This should be done not only by offering workshops and training, but also by encouraging frequent contact through conferences, reciprocal visits and online resources such as list-servers and bulletin boards.

Potential impact of CGIAR networking

CGIAR networking has played and will continue to play an important role in promoting comprehensive training towards bridging the technological divide in several ways. First, the CGIAR should promote regional collaboration not only across borders, but also and particularly among institutions *within* developing countries. Germplasm resource centres in resource-poor nations should be encouraged to help and support each other, to make the best use of their finite resources, both material and human. Systems should be set up to facilitate sharing of expensive equipment and expertise.

Second, a special strength of CGIAR networking is and should be that it allows the development of excellence in a particular niche. No centre can be expert on all topics and do everything well. Instead, each centre should be recognized as having a particular area of expertise. Centres should be ready and willing to share their complementary expertise and collaborate on projects to form a united front in the challenge of keeping up with cutting-edge science.

Most importantly, the CGIAR centres should coordinate information sharing. In this age of an ever-increasing pace in new developments in research, many CGIAR constituents and national programmes remain aware of advances only with difficulty or not at all. Yet to make appropriate choices on new technologies, researchers must keep informed. Of particular importance to those at genetic resources repositories are current issues related to access, benefit sharing and intellectual property rights, which continually change. The CGIAR centres must take responsibility for holding appropriate workshops to discuss and update constituents about these issues.

Researchers in developing countries, especially those working on 'orphan crops', must be aware of current research in related crops so they can take advantage of this information from a comparative genomics perspective. Genebank scientists should stay informed on current use so they can be flexible and adapt to their changing role as the user community sees it. The CGIAR centres should also form a

coordinated network where equipment and expertise is available for shared use.

For many research laboratories, new technologies such as DNA sequencers are expensive and technically difficult to maintain, and usually not used at a high enough rate to make their purchase worthwhile. A much more efficient approach would be to buy and maintain this type of equipment at a few centralized locations, thus leaving the laboratories with more financial resources for other purposes, more flexibility and increased currency in an ever-progressing field.

Outreach: public awareness and attracting the next generation of scientists

Outreach and public awareness should also continue to be important functions of CGIAR networking. The general public must realize the importance of conserving and understanding the earth's genetic resources. Not only will this ensure that funding the centres remains a priority but it will also encourage young people to consider the field as a viable career option.

Attracting young students to careers in plant genetic resources conservation and plant breeding is increasingly difficult, probably because these fields are seen as unglamorous or obsolete compared with biotechnology, or as too difficult because of the broad knowledge base needed to be successful in these fields. The future of plant genetic resources conservation depends on young students continuing to see this field as significant, as well as providing rewarding career options. Thus, it is in the best interests of the CGIAR to promote the importance of plant genetic resources conservation and foster a supportive and encouraging environment for both new researchers in the field and those already in the system.

References

AAAS (American Association for the Advancement of Science). 1993. Benchmarks for Science Literacy. Oxford University Press, New York.

AfricaGrant. 2001. News item at <http://www.news.cornell.edu/releases/Jan01/AfricaGrant.bpf.html>

Deboer, G.E. 1991. A History of Ideas in Science Education: Implications for Practice. Teachers College Press, New York.

Fresco, L.O. 2003. Which road do we take? Harnessing genetic resources and making use of life sciences, a new contract for sustainable agriculture. Paper presented at the EU Discussion Forum 'Towards Sustainable Agriculture for Developing Countries: Options from Life Sciences and Biotechnologies', Brussels, 30–31 January.

Karp, A., S. Kresovich, K.V. Bhat, W.G. Ayad and T. Hodgkin. 1997. Molecular Tools in Plant Genetic Resources Conservation: A Guide to the Technologies. IPGRI, Rome.

Lederman, N. 1992. Students' and teachers' conceptions of the nature of science: a review of the research. J. Res. Sci. Teaching 29:331–359.

Northoff, E. 2003. FAO warns of 'molecular divide' between North and South. FAO, Rome. <http://www.fao.org/english/newsroom/news/2003/13960-en.html>

Tanksley, S.D. and S.R. McCouch. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066.

de Vicente, M.C. and T.M. Fulton 2003. Using Molecular Market Technology Effectively in Plant Genetic Diversity Studies. IPGRI and Cornell University Institute for Genomic Diversity, Rome.

ACRONYMS AND ABBREVIATIONS

AAAS	American Association for the Advancement of Science
AFLPs	amplified fragment length polymorphisms
ARS	Agricultural Research Service (of USDA)
AVRDC	Asian Vegetable Research and Development Center, Taiwan
BAC library	bacterial artificial chromosome insert library
BCC	barley core collection
BRG	Bureau des Ressources Génétiques, France
CAAS	Chinese Academy of Agricultural Sciences, P. R. China
CAPS	cleaved amplified polymorphic sequences
CBD	Convention on Biological Diversity (of UNEP)
cDNA	complementary DNA
CENARGEN	Centro Nacional de Pesquisa de Recursos Genéticos e Biotecnologia (of EMBRAPA)
CGIAR	Consultative Group on International Agricultural Research
CIMMYT	Centro Internacional para Mejoramiento de Maíz y Trigo, Mexico
CIP	Centro Internacional de la Papa, Peru
CSIC	Consejo Superior de Investigaciones Científicos, Spain
CSIRO	Commonwealth Scientific and Industrial Research Organisation, Australia
DDBJ	DNA Data Bank of Japan
DGPC	Diversité et Génomes des Plantes Cultivées, France
EBDB	European Barley Database
EBI	European Bioinformatics Institute
ECP/GR	European Cooperative Programme for Crop Genetic Resources Networks
EMBL	European Molecular Biology Laboratory
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária, Brazil
ESTs	expressed sequence tags
FAO	Food and Agriculture Organization of the United Nations
FECYT	Fundación Española para la Ciencia y la Tecnología (of INIA)
GIS	geographic information systems
GRIN	Germplasm Resources Information Network (of USDA)
GRST	Genetic Resources Science and Technology (of IPGRI)
IARC	international agricultural research centre
IBCR	Institute of Biodiversity Conservation and Research
IBPGR	International Board for Plant Genetic Resources (now IPGRI)
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics, India
IGB	Institute for Grapevine Breeding, Germany
IGD	Institute for Genomic Diversity (of Cornell University, NY, USA)
IGGP	International Grape Genome Program, CA, USA
IITA	International Institute of Tropical Agriculture, Nigeria
INIA	Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Spain

INIBAP	International Network for the Improvement of Banana and Plantain, France
INRA	Institut National de la Recherche Agronomique, France
IPK	Institute of Plant Genetics and Crop Plant Research, Germany
IRD	Institut de Recherche pour le Développement, France
IRRI	International Rice Research Institute, Philippines
ISMAA	Istituto Agrario San Michele, Italy
ISO	International Organization for Standardization
IVIA	Instituto Valenciano de Investigaciones Agrarias, Spain
KN	'knuckles and nodes' (approach to database integration)
L	logarithmic (sampling strategy)
NBPGR	National Bureau of Plant Genetic Resources, India
NCBI	National Center for Biotechnology Information, MD, USA
NGB	Nordic Genebank, Sweden
NGOs	nongovernmental organizations
NSSL	National Seed Storage Laboratory (of USDA-ARS)
PCR	polymerase chain reaction
PCS	principal component score (sampling strategy)
PGR	plant genetic resources
PROMUSA	Global Programme for *Musa* Improvement
QTLs	quantitative trait loci
RAPD	random amplified polymorphic DNA
RFLPs	restriction fragment length polymorphisms
SADC	Southern African Development Community
SGRP	System-wide Genetic Resources Programme (of CGIAR)
SINGER	System-wide Information Network for Genetic Resources
SNP	single nucleotide polymorphism
SPGRC	SADC Plant Genetic Resources Centre
SSAGR	sub-Saharan African genetic resources project
SSR	simple sequence repeat
T	taxonomic (sampling strategy)
TAIR	The *Arabidopsis* Information Resource
TIGR	The Institute for Genomic Research
UMR	Unité Mixte de Recherche, France
UNEP	United Nations Environment Programme
UNR	University of Nevada-Reno, NV, USA
USAID	United States Agency for International Development
USDA	United States Department of Agriculture
USSR	Union of Soviet Socialist Republics (now various countries)
VIR	N. I. Vavilov All-Russian Scientific Research Institute of Plant Industry, Russia
ZADI	Centre for Documentation and Information in Agriculture, Germany

PARTICIPANTS AT THE 2002 SGRP–IPGRI EXPERT CONSULTATION MEETING, LEÓN, SPAIN

All participants with an asterisk against their names have contributed to the papers in this volume.

Cubero, José Ignacio
Professor, Genetics and Plant Breeding
ETS de Ingenieros Agrónomos y de Montes
Apartado 3048
Cordoba, Spain
Tel.: 34-95-7218504
Fax: 34-95-7218503
ge1cusaj@uco.es

Dulloo, Ehsan
Germplasm Conservation
IPGRI
Via dei Tre Denari, 472/a
00057 Rome, Italy
Tel.: 39-06-6118206
Fax: 39-06-61979661
e.dulloo@cgiar.org

Engels, Jan*
Group Director, GRST
IPGRI
Via dei Tre Denari, 472/a
00057 Rome, Italy
Tel.: 39-06-6118222
Fax: 39-06-61979661
j.engels@cgiar.org

Ferreira, Marcio Elias
Plant Genetics and Molecular Marker Analysis
CENARGEN
Parque Estação Biológica
Final Av. W/5 Norte
Caixa Postal 02372 CEP 70770-900
Brasília, Brazil
Tel.: 55-61-4484600
Fax: 55-61-3403624
ferreira@cenargen.embrapa.br

Freymark, Peter
Research Technology Service
PIONEER
7300 N.W. 62nd Avenue
PO Box 1004
Johnston, Iowa, USA
Tel.: 1-515-2532478
peter.freymark@pioneer.com

Fulton, Theresa*
Director of Outreach, IGD
Biotechnology Center, Cornell University
135 Biotechnology Building
Ithaca, NY 14853-2703, USA
Tel.: 1-607-2554323
Fax: 1-607-2556249
tf12@cornell.edu

Graner, Andreas*
Head, Dept of Genebank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstraße 3
D 06466 Gatersleben, Germany
Tel.: 49-0-39482 5155
graner@ipk-gatersleben.de

Hamon, Serge*
UMR-DGPC
c/o Centre IRD de Montpellier
911, Avenue Agropolis,
B.P. 64501
34394 Montpellier Cedex 05, France
Tel.: 33-4-67416194
Fax: 33-4-67416222
serge.hamon@mpl.ird.fr

Hoogendoorn, Coosje
Deputy Director General, Programmes
IPGRI
Via dei Tre Denari, 472/a
00057 Rome, Italy
Tel.: 39-06-6118200
Fax: 39-06-61979661
c.hoogendoorn@cgiar.org

Karihaloo, Jawahir Lal
Acting Project Director
National Research Centre on DNA Fingerprinting
NBPGR
Pusa Campus,
New Delhi, 110012
India
Fax: 91-581-5789203-337
jlk@nbpgr.delhi.nic.in

Kresovich, Stephen*
Professor and Director, IGD
Biotechnology Center, Cornell University
158 Biotechnology Building
Ithaca, NY 14853-2703, USA
Tel.: 1-607-2551492
Fax: 1-607-2556249
Sk20@cornell.edu

Martínez Zapater, José Miguel
Deputy Director of Research and Technology
Instituto Nacional de Investigación y Tecnología Agraría y Alimentaria (INIA)
Crta. de la Coruña, Km. 7,5
28040 Madrid, Spain
Tel.: 34-91-5854500
Fax: 34-91-5854506
zapater@inia.es

Navarro, Luis*
Director, Dept of Plant Protection and Biotechnology
IVIA
Apartado Oficial 46113
Moncada, Valencia, Spain
Tel.: 34-96-1391000
Fax: 34-96-1390240
lnavarro@ivia.es

Ortiz, Rodomiro*
Director, Crop Improvement Division
International Institute of Tropical Agriculture (IITA)
c/o Lambourn (UK) Limited
Carolyn House, 26 Dingwall Road
Croydon CR9 3EE, UK
Tel.: 234-2-2412626
Fax: 234-2-2412221
r.ortiz@cgiar.org

de Ponti, Orlando
Director, Nunhems Zaden B.V.
PO Box 4005
Haelen, Netherlands
Tel.: 31-475-599222
Fax: 31-475-599223
o.deponti@nunhems.com

Richards, Christopher*
Population Geneticist
Plant Germplasm Preservation Research Unit
USDA-ARS
National Seed Storage Laboratory
1111 South Mason Street
Fort Collins, CO 80521-4500, USA
Tel.: 1-970-4953200
Fax: 1-970-2211427
crichard@lamar.colostate.edu

Roca, William
Plant Cell Physiologist
CIP
PO Box 1558
Lima 12, Peru
Tel.: 51-1-3496017
Fax: 51-1-3495638
w.roca@cgiar.org

Sackville Hamilton, Ruaraidh
Head, Genetic Resources Center
International Rice Research Institute (IRRI)
DAPO Box 7777
Metro Manila, Philippines
Tel.: 63-2-5805600 ext 2809
Fax: 63-2-8450606
rhamilton@cgiar.org

Skovmand, Bent
Formerly Head, Wheat Germplasm Bank, CIMMYT
As of 1 January 2004, Director
Nordic Genebank
PO Box 41
S-230 53 Alnarp, Sweden
Tel.: 46-40-536640
Fax: 46-40-536650

Villalobos, Victor
Coordinator, Asuntos Internacionales
Ministerio de Agricultura
Insurgentes Sur # 476
Colonia Roma Sur, Delegación Cuauhtemoc
México City, DF, Mexico
Tel.: 52-5-557228323 /591831055
Fax: 52-5-55843485
vvilla@sagarpa.gob.mx

de Vicente, M Carmen*
Plant Molecular Genetics
IPGRI
A.A. 6713
Cali, Colombia
Tel.: 57-2-4450029
Fax: 57-2-4450096
c.devicente@cgiar.org

Zhang, Xueyong
Institute of Crop Germplasm Resources
Chinese Academy of Agricultural Sciences (CAAS)
Key Laboratory of Crop Germplasm and Biotechnology
Chinese Agricultural Ministry
12 Zhongguancun Nandajie
Beijing 100081, P. R. China
Tel.: 86-10-62186630
Fax: 86-10-62186629
xueyongz@mail.caas.net.cn

Observers

Bolivar Salcedo, José Miguel
Subdirección General, Prospectiva y Coordinación de Programas
INIA
José Abascal, 56 28071
Madrid, Spain
Tel.: 34-1-3473949
Fax: 34-1-4423587
bolivar@inia.es

Hormaza, Iñaki
Scientist
Consejo Superior de Investigaciones Científicas (CSIC)
Estación Experimental La Mayora
29750 Agarrobo-Costa (Málaga), Spain
Tel.: 34-952-552656
Fax: 34-952-552677
ihormaza@ulm.csic.es

ISBN 92-9043-623-9