



Molecular Markers for Allele Mining

Proceedings of a workshop, 22–26 August 2005, MS Swaminathan Research Foundation, Chennai, India

M. Carmen de Vicente, Jean-Christophe Glaszmann,
editors

International Plant Genetic Resources Institute, Generation Challenge Programme



**FUTURE
HARVEST**

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



Molecular Markers for Allele Mining

*Proceedings of a workshop, 22–26 August 2005, MS Swaminathan Research
Foundation, Chennai, India*

**M. Carmen de Vicente, Jean-Christophe Glaszmann,
*editors***

International Plant Genetic Resources Institute, Generation Challenge Programme

The International Plant Genetic Resources Institute (IPGRI) is an autonomous international scientific organization, supported by the Consultative Group on International Agricultural Research (CGIAR). IPGRI's mandate is to advance the conservation and use of genetic diversity for the well-being of present and future generations. IPGRI's headquarters is based in Maccarese, near Rome, Italy, with offices in another 19 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 2001, had been signed and ratified 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.

In 2000 financial support for the Research Agenda of IPGRI was provided by the Governments of Armenia, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, China, Croatia, Cyprus, Czech Republic, Denmark, Estonia, F.R. Yugoslavia (Serbia and Montenegro), Finland, France, Germany, Greece, Hungary, Iceland, India, Ireland, Israel, Italy, Japan, Republic of Korea, Latvia, Lithuania, Luxembourg, Macedonia (F.Y.R.), Malta, Mexico, the Netherlands, Norway, Peru, the Philippines, Poland, Portugal, Romania, Slovakia, Slovenia, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, Uganda, the UK and the USA and by the African Development Bank (AfDB), Asian Development Bank (ADB), Center for Development Research (ZEF), Center for Forestry Research (CIFOR), Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica (CATIE), Common Fund for Commodities (CFC), Technical Centre for Agricultural and Rural Cooperation (CTA), European Environmental Agency, European Union, Food and Agriculture Organization of the United Nations (FAO), Food and Fertilizer Technology Center for the Asia and Pacific Region (FFTC), Future Harvest, Global Forum on Agricultural Research (GFAR), Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología (COLCIENCIAS), Inter-American Drug Abuse Control Commission (CICAD), International Association for the Promotion of Cooperation with Scientists from the New Independent States of the former Soviet Union (INTAS), International Development Research Centre (IDRC), International Foundation for Science (IFS), International Fund for Agricultural Development (IFAD), International Service for National Agricultural Research (ISNAR), Japan International Research Centre for Agricultural Sciences (JIRCAS), National Geographic Society, Natural Resources Institute (NRI), Programme on Participatory Research and Gender Analysis for Technology Development and Institutional Innovation (PGRA), Regional Fund for Agricultural Technology (FONTAGRO), Rockefeller Foundation, Taiwan Banana Research Institute (TBRI), Technova, United Nations Development Programme (UNDP), UNDP Global Environment Facility (UNDP-GEF), United Nations Environment Programme (UNEP), UNEP Global Environment Facility (UNEP-GEF), United States Department of Agriculture (USDA), Vlaamse Vereniging voor Ontwikkelingssamenwerking en Technische Bijstand (VVOB) and the World Bank.

The Generation Challenge Programme (GCP) is a research and capacity building network that uses plant genetic diversity, advanced genomic science, and comparative biology to develop tools and technologies that enable plant breeders in the developing world to produce better crop varieties for resource-poor farmers.

The GCP is one of four Challenge Programmes established by the CGIAR to make high impacts in the short term through thematic approaches involving a multitude of research, development, health, and delivery organizations.

The Generation Challenge Programme brings together three sets of partners—the centres of the Consultative Group on International Agricultural Research (CGIAR), advanced research institutes (ARIs), and national agricultural research systems (NARS) in developing countries—to deliver the fruits of the Genomics Revolution to resource-poor farmers.

The Generation Challenge Programme has five subprogrammes that span the spectrum of research in germplasm, genomics, bioinformatics, and molecular breeding for agricultural development:

Genetic Diversity of Global Genetic Resources

Comparative Genomics for Gene Discovery

Trait Capture for Crop Improvement

Genetic Resources, Genomic, and Crop Information Systems

Capacity Building and Enabling Delivery

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, the Generation Challenge Programme 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.

Citation:

de Vicente, M.C., Glaszmann, J.C. (eds). 2006. Molecular Markers for Allele Mining. Proceedings of a workshop, 22-26 August 2005, MS Swaminathan Research Foundation, Chennai, India. International Plant Genetic Resources Institute, Rome, Italy.

Cover images: Banana plant (photo: Carlos A. Tovar), wheat heads (courtesy of CIMMYT), Andean climbing beans (photo: Mathew Blair), group of participants (courtesy of the Generation Challenge Programme). Design: Carlos Tovar, Frances Ferraiuolo.

ISBN-13: 978-92-9043-705-5

ISBN-10: 92-9043-705-7

IPGRI

Via dei Tre Denari 472/a

00057 Maccaresse

Rome, Italy

© International Plant Genetic Resources Institute, 2006

Acknowledgements

The authors wish to express their gratitude to Dr. Monkombu S. Swaminathan and the MS Swaminathan Research Foundation for hosting the workshop, to Carlos Andrés Tovar, Monique Costes and Adriana Sánchez for their support with the logistics, to Elizabeth McAdam for her collaboration in editing the document, and to Ana Luisa Triana for her help with the elaboration of these Proceedings.

Preface

M. Carmen de Vicente

Leader of SP5, GCP

IPGRI–Americas, c/o CIAT, Cali, Colombia

As the workshop progressed, participants expressed the feeling that both presentations and discussions were indeed informative and useful. While all year round many opportunities exist to attend scientific presentations, this workshop not only offered the chance to do so, but also allowed debate, giving rise to what was deemed as, altogether, a very special learning event. As a result, at the end of the workshop, participants decided to prepare a publication on session contents, agreeing to provide summaries of their presentations. The summarized proceedings will soon be available on the Generation Challenge Programme's Web site at www.generationcp.org

This publication closely follows the development of the sessions as organized in the workshop's agenda (Appendix 2). Themes include the progress made in crop-genotyping activities in SP1, description and achievements of other ongoing projects in SP1, and parallel scientific results of research projects outside the GCP.

The document also reflects the diversity of the participating groups who attended the workshop. While summaries of research activities at home institutions representing NARS are provided, three significant participating groups (SGRP, observers from developed country institutions, and scientists from developing country institutions) also share their group views on the usefulness of the research undertaken by SP1, together with recommendations for the subprogramme's future directions.

Perhaps not so evident, but not the least for it, is the appreciation that SP5 promoted this type of workshop. Scientists, from both the GCP and outside it, were given the opportunity to exchange experiences and knowledge within a unique situation that permitted the creation of bridges along the lines of the GCP's delivery philosophy: that all research products should be destined for and reach a user. Without doubt, this feature of the workshop was ground-breaking in the scientific ambience. From now on, we hope it will be GCP's signature in all its work as it takes up the various challenges of improving agriculture in developing countries.

Content

Acknowledgements	v
Preface	vii
<i>M. Carmen de Vicente</i>	<i>vii</i>
Introduction	5
The GCP's workshop on molecular markers for allele mining	5
<i>Jean-Christophe Glaszmann</i>	<i>5</i>
Session I: Review of results for tier 1 crops	13
Genotyping composite germplasm sets: wheat	13
<i>Susanne Dreisigacker (with the collaboration of Marilyn Warburton, Z. Xueyong, François Balfourier, and Nachit Miloudi)</i>	<i>13</i>
Cowpea genotyping	14
<i>Sarah Hearne, Morag Ferguson, and Visvanathan Mahalakshmi</i>	<i>14</i>
Assessing the global diversity of cassava (<i>Manihot esculenta</i> Crantz) genetic resources, according to simple sequence repeat markers	15
<i>Paula Hurtado, Jaime Marín, César Ospina, Charles Buitrago, Wilson Castelblanco, Ana María Correa, Edgar Barrera, Janneth Patricia Gutiérrez, Luis Guillermo Santos, Graciela Mafla, M. Carmen de Vicente, Sarah Hearne, Morag Ferguson, Alfredo Alves, Claudia Ferreria, and Martin Fregene</i>	<i>15</i>
Genotyping composite germplasm sets: maize	17
<i>Susanne Dreisigacker (with the collaboration of Marilyn Warburton, Alain Charcosset, Visvanathan Mahalakshmi, and Yu Li)</i>	<i>17</i>
Genotyping composite germplasm sets: rice	18
<i>Kenneth McNally (with the collaboration of Ruairaidh Sackville Hamilton, Claire Billot, B. Courtois, L. Benoit, C. Martínez, M. Lorieux, C. Brondani, L. Han, M.-N. Ndjiondjop, and S. McCouch)</i>	<i>18</i>
Sorghum genetic data analysis	20
<i>Claire Billot (with Tom Hash)</i>	<i>20</i>
Common bean genotyping	21
<i>Mathew Blair</i>	<i>21</i>
Diversity of barley genetic resource in China and progress in barley genotyping for the GCP	22
<i>Zhang Jing, Zhang Chihong, Michael Baum & Jan Valkoun</i>	<i>22</i>
Gene pool structure of cultivated potatoes assessed by SSR marker analyses	24
<i>Marc Ghislain, Jorge Núñez, Guillermo Trujillo, María del Rosario Herrera, Frank Guzmán, and David Spooner</i>	<i>24</i>
Mining the chickpea composite collection for allelic variation	25
<i>Hari Upadhyaya, M. Baum, H.K. Buhariwala, S.M. Udupa, B.J. Furman, S.L. Dwivedi, David H. Hoisington, P.M. Gaur, S. Chandra, C.L.L. Gowda, J.H. Crouch, and Jan Valkoun</i>	<i>25</i>
Genotyping composite germplasm sets: <i>Musa</i>	26
<i>Isabelle Hippolyte and Nicolas Roux (presented by Claire Billot)</i>	<i>26</i>
Session II: Selected Experiences	27
Characterizing and managing genetic diversity in the bread wheat collection held at INRA	27
<i>François Balfourier</i>	<i>27</i>
Barley research at IPK	29
<i>Andreas Graner</i>	<i>29</i>

Maize at Cornell University	31
<i>Gael Pressoir</i>	31
Session III: Review of other ongoing SP1 projects	33
Assessing Ecotilling as a methodology for targeted genotyping and SNP discovery: rice, sorghum, and <i>Musa</i>	33
<i>Kenneth McNally (with the collaboration of Ruairaidh Sackville Hamilton, Claire Billot, M. Deu, I. Hippolyte, F.-C. Baurens, and J.-F. Rami)</i>	33
Assessing linkage disequilibrium in sorghum	35
<i>Jean-Christophe Glaszmann, Monique Deu and Claire Billot</i>	35
Assessing linkage disequilibrium in coconut	36
<i>Luc Baudouin, Angélique Berger and Patricia Lebrun</i>	36
A unified mixed-model method for association mapping: accounting for multiple levels of relatedness	37
<i>Jianming Yu and Gael Pressoir (with the collaboration of William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, Brandon S. Gaut, James B. Holland, Dahlia Nielsen, Stephen Kresovich, and Edward S. Buckler)</i>	37
Characterizing the genetic diversity of maize populations: documenting the global migration of maize from its centre of origin	38
<i>Susanne Dreisigacker (with the collaboration of Marilyn Warburton, L. George, S. Taba, Visvanathan Mahalakshmi, A. Menkir, Alain Charcosset, Z. Muthamia, S.H. Zhang, B.M. Prasanna, Sutrisno, N. Sawon, Pichet Grudloyma, P.X. Hao, and A. Salazar)</i>	38
Session IV: Use of molecular data for subsampling	39
Developing decision-support systems for sampling germplasm	39
<i>Marco Bink, and Xavier Perrier</i>	39
Controlling population structure in association mapping	40
<i>Gael Pressoir</i>	40
Subsampling among 700 sorghum accessions	41
<i>Claire Billot</i>	41
Allele binning	42
<i>Dave Hoisington</i>	42
Session V: Guests's research activities: NARS.....	43
Coconut germplasm conservation and research with molecular markers at the Coconut Research Institute of Sri Lanka	43
<i>Lalith Perera</i>	43
University of Southern Mindanao: research activities	44
<i>Emma K. Sales</i>	44
AVRDC: The World Vegetable Center	45
<i>Liwayway M. Engle</i>	45
Genetic enhancement for the coastal ecosystem: research at the MSSRF	46
<i>Ajay Parida</i>	46
Developing and using molecular markers at the Plant Biotechnology Centre, IARI	47
<i>T. Mohapatra</i>	47
Using AFLP markers and passport data when sampling a core collection and estimating genetic diversity	49
<i>Kioumars Ghamkhar, and R. Snowball</i>	49
Agriculture and Agri-Food Canada	50
<i>Campbell G. Davidson</i>	50

Molecular marker research on wheat germplasm at the Ch. Charan Singh University	52
<i>H.S. Balyan and P.K. Gupta</i>	52
Marker-assisted selection for drought tolerance in rice	53
<i>H.E. Shashidhar</i>	53
Current status and recent progress of conservation and use of PGR in the Republic of Korea	54
<i>Ma kyung-ho</i>	54
Genetic improvement of rice in Tamil Nadu, India	56
<i>Sabariappan Robin, S. Rajeswari, K. Mohana Sundaram, and T.S. Raveendran</i>	56
Using molecular marker technology for mapping quantitative trait loci, functional genomics, and genetic diversity analysis	57
<i>M. Maheswaran</i>	57
Session VI: The Long-Generation Crops Project: population structure, phenotypic information, and association studies	59
Population structure, phenotypic information, and association studies in long-generation crops: A project overview	59
<i>M. Carmen de Vicente</i>	59
Population structure, phenotypic information, and association studies for coconut in Vanuatu	60
<i>Luc Baudouin and Patricia Lebrun</i>	60
Yam in Vanuatu	61
<i>Jean-Louis Noyer and Vincent Lebot (presented by Jean-Christophe Glaszmann)</i>	61
Association analysis in the course of varietal improvement	62
<i>M. Bonierbale, Marc Ghislain, E. Mihovilovich, Jorge Nuñez, J. Landeo, and M. Gastelo</i>	62
Population structure, phenotypic information, and association studies in long-generation crops: cassava update	64
<i>Paula Hurtado, Janneth Patricia Gutiérrez, Isabel Moreno, Edgar Barrera, and Martin Fregene</i>	64
Assessing the genetic diversity of <i>Musa</i> in Congo Basin	65
<i>FC Baurens, K Tomekpe, JL Noyer</i>	65
Session VII: Perspectives	67
The need for allele mining: perspectives of the System-wide Genetic Resources Programme (SGRP)	67
<i>Ruaraidh Sackville Hamilton, Moussa Sie, Jan Valkoun, Hari Upadhyaya, and Visvanathan Mahalakshmi</i>	67
Observers' viewpoint	69
<i>Campbell G. Davidson, François Balfourier, Ajay Parida, Kioumars Ghamkar, Andreas Graner, and Gael Pressoir</i>	69
NARS' perspectives	70
<i>Emma K. Sales, Lalith Perera, Liwayway M. Engle, T. Mohapatra, H.S. Balyan, H.E. Sashidhar, M. Maheswaran, Robin Sabariappan, Kyung-Ho Ma, and Kioumars Ghamkhar</i>	70
SP5's activities	72
<i>M. Carmen de Vicente</i>	72
Appendix I. Acronyms and abbreviations used in the text.....	73
Appendix II: Programme	77
Appendix III. Participants.....	81

Introduction

The GCP's workshop on molecular markers for allele mining

Jean-Christophe Glaszmann

Leader of SP1, GCP

CIRAD, Montpellier, France

Background

Separately from all the resources held by national programmes, various partners of the GCP also hold and manage large germplasm collections. The CGIAR centre collections, held in trust under the auspices of FAO, are the largest and most representative. They gain by being complemented with materials from national programmes, for enhancing global diversity coverage, and for stimulating vitality and solidarity among all contributors.

The diversity of global genetic resources provides material for varietal improvement. The GCP undertakes in-depth studies to better understand the diversity of traits and modes of adaptation available in crop germplasm, and to access the corresponding genetic factors for use in developing new varieties. Subprogramme 1 (SP1), focused on genetic diversity, is organized for identifying novel, diverse, and superior variants of genes involved in targeted traits. It receives orientations from SP2 and aims at applications in SP3. In accordance with global priority, particular emphasis is given to drought tolerance. The basic strategy followed by SP1 recognizes several objectives:

- To optimize and apply molecular methods to monitor global and readily accessible germplasm and to enhance the description of global diversity. Attention is paid to promoting user-friendly markers that can be easily applied by national programmes to describe and compare their own materials. Attractiveness to national programme scientists is key to enhancing global cohesion and future sustainability. SP1 must therefore help develop and consolidate a global facility for the molecular description of germplasm, with specific attention to efficiency, throughput, flexibility, and accessibility.
- To improve the methodology for assessing drought tolerance, a particularly complex challenge. The diversity of crops, environments, and practices necessitates exchange of experiences and development of platforms with global access.
- To implement germplasm evaluation within an analytical framework that will yield information on the genetic factors (genes/alleles/haplotypes) that underlie the diversity of response to drought stress. This combines the comparative description of molecular polymorphisms and phenotypic variation, and the study of statistical associations. Various types of populations will yield different resolutions and will necessitate a matching genotyping intensity. Of primary importance in attracting the attention and enthusiasm of researchers in national programmes, some of these genetic analyses must be directly connected to the activities of breeders and possibly to farmers' practices.

As are the other subprogrammes, SP1 is made up of several contracted projects. Some were commissioned in 2003 and started in 2004; others started in 2005 after a competitive process, or were commissioned for priority research lines that were as yet uncovered. Most of the concrete results achieved so far derive from projects initiated in 2004, whereas those started in 2005 have essentially focused on organizational matters and producing specialized resources.

Workshop structure

The workshop's goal was to review the progress of various projects, explicating both scientific and managerial issues, provide the SP1 community with the opportunity for

internal coordination, and discuss a range of scientific issues and problems with three groups of scientists (Appendix III), who were:

- Colleagues from the System-wide Genetic Resources Program (SGRP), bringing together germplasm bank curators,
- Colleague experts in the fields covered by SP1 and coming from advanced laboratories outside the GCP, and
- Colleagues from various NARS, who are personally involved in SP1-related research fields.

The programme (Appendix II) was organized in various sessions that represented SP1's structure:

1. Review of results for tier 1 crops, with presentations on different crops.
2. Selected experiences (from outside the GCP), including wheat at INRA and barley at IPK.
3. Review of other ongoing SP1 projects on topics such as:
 - I. Phenotyping platforms and modelling,
 - II. Genotyping methodologies such as DArT and Ecotilling,
 - III. Linkage disequilibrium assessments, and
 - IV. Further genotyping in maize.
4. Using molecular data for subsampling.
5. Research activities of guests from NARS.
6. Population structure, phenotypic information, and association studies, a review of research by the Long-Generation Crops Project, covering various issues and crops.

The programme concluded with a general discussion on SP1 and SP5 perspectives, followed by presentations of the viewpoints of germplasm curators, GCP observers, NARS representatives, and the two subprogrammes.

Review of results for tier 1 crops: core samples and molecular diversity for describing global genetic resources

SP1 has undertaken systematic work on the crops with which the CGIAR has active breeding programmes, beginning in 2004 with 11 crops. In 2005, another 7 were incorporated, and pearl millet is to be included in 2006. For each crop, the first step consists of collating information on various existing collections (the 'composite set') to apply a simple rationale for extracting a representative sample (the 'core sample'). This is coordinated by the IARC (or IARCs) in charge of the crop and has been completed for all 18 crops. The size of the core samples depends heavily on the global amount of resources available in the collections, and may range from several hundred accessions to a maximum of 3000 for the most important crops. The size of some core samples may need to be increased to optimize coverage of existing diversity.

The second step consists of characterizing each core sample with molecular markers to reveal the structure of its diversity and to extract a reduced sample, called the 'reference sample', that will be made available for additional characterization and evaluation to reveal functional correlations. This is in progress for 17 crops (Table 1), of which 11 were reviewed in Session I of the Workshop and presented as summaries in this volume.

Table 1. Molecular characterization of core samples in 19 GCP crops.

Crop	Lead institution and partners	Year starting and ending	Core sample genotyping target (no. acc. × no. loci)	Genotyping, percentage by mid-2005
Reviewed in Workshop				
Rice	IRRI-CAAS-CIAT- WARDA-Agropolis	2004-05	3000 × 50	50
Maize	CIMMYT-CAAS-IITA- Agropolis	2004-05	1775 × 50	40
Wheat	CIMMYT-CAAS- ICARDA-Agropolis	2004-06	(2600 + 400) × 50	60
Sorghum	ICRISAT-Agropolis- CAAS	2004-06	(700 + 2300) × 50	30
Barley	ICARDA-CAAS	2004-06	(500 + 2500) × 50	30
Com'n bean	CIAT-EMBRAPA	2004-05	3000 × 50	60
Cowpea	IITA	2004-05	2000 × 50	–
Chick-pea	ICRISAT-ICARDA	2004-06	286 + 2714	50
Cassava	CIAT-EMBRAPA-IITA	2004-05	3000 × 36	60
Potato	CIP	2004-05	1000 × 50	80
<i>Musa</i>	IPGRI-IITA-Agropolis	2004-05	960 × 50	20
Others				
Ground-nut	ICRISAT-EMBRAPA	2005-06	1000 × 20	
Pigeon-pea	ICRISAT	2005-06	1000 × 20	
Lentil	ICARDA	2005-06	1000 × 30	
Yam	IITA	2005-06	350 × 20	
Coconut	Agropolis	2005-06	1000 × 22	80
Sweet potato	CIP	2005-06	500 × 50	
Finger millet	ICRISAT	2005-06	1000 × to be det'd	
Pearl millet	ICRISAT	2006	To be determined	

The markers used are primarily microsatellites (i.e. simple sequence repeats or SSRs), selected for their polymorphism and widespread distribution across the genome. Between 20 and 50 are being used, depending on the crop. Such molecular characterization is commonly shared by several partners, which permits exploring the value and constraints of decentralizing or centralizing this kind of work. The idea is to create a bank of simple methods and data that will serve as reference for all future surveys, so that any group in the world can compare any germplasm with the core sample collection of the species described by the GCP.

A temporary template for formatting data before incorporation into the GCP's bioinformatic framework was made available in May 2005 and results are being entered in the GCP database.

An initial target was set for each crop and widely distributed among partners. Eighteen months down the road, most of these targets will clearly not be met as planned, with completion rates expected to be between 50 and 100%. This is because of a range of diverse causes, linked not only to local difficulties of regular supplies and services but also to transversal difficulties in exchanging biological materials and comparing protocols and results. The lessons learned will help elaborate and promote a collective approach for optimizing high-throughput molecular characterization in international agricultural research.

Local difficulties in accessing supplies and services would justify concentrating increased capacity to a limited number of laboratories that have sufficient resources and better negotiating capacity.

One important factor for the initial delays and, ultimately, the incomplete data matrices was the restrictions on exchanging seeds, or even DNA, that constrained the NARS partners within the GCP. This constraint limits the creation of a wider circle including more NARS and, thus, requires both political action at the GCP level and beyond, and adaptation of practices, while keeping the GCP's founding spirit. More specific technical recommendations can be made:

- Time must be taken to obtain easy-to-use markers before embarking on massive genotyping.
- Effort must be made to ensure that the agreed-upon protocols are communicated to all partners and that the agreed-upon recommendations are being implemented by all partners.
- With regards to heterogeneity within accessions, the use of materials within the GCP needs to be differentiated from the standard management for keeping global diversity, including that within accessions. For an accession to become part of the reference sample, it must be homogeneous (as far as one can judge). It will then be given a specific germplasm identifier or GID and kept as genetic stock to be managed and distributed by the corresponding IARC.

Nevertheless, very large amounts of data have been produced and will be sufficient for refining smaller germplasm samples with enhanced representativeness. These reference samples will be made available and distributed to collaborators around the world.

Selected experiences

INRA's presentation on wheat illustrated a rigorous process for selecting a reference germplasm sample, using various types of data, including molecular data, as a basis for subsampling and a posteriori validation.

Likewise, IPK's presentation on barley highlighted a carefully integrated approach for applying a series of tools and methods involving both structural and functional genomics.

The presentation on maize illustrated an integrated and very ambitious approach for uncovering useful genes and alleles in maize through association mapping and, more recently, joint linkage analysis and association mapping.

All three speakers made a highly commendable effort in synthesizing a wealth of information and gave remarkable presentations, with complementary contributions that covered the whole SP1 and described the best current science.

Review of other ongoing SP1 projects

This allowed workshop participants to perceive all the domains covered by SP1 and obtain a sense of its global strategy and perspectives.

Optimizing assessment of drought tolerance

The backbone of this project consists of (1) optimizing access to phenotyping platforms that are efficient, coordinated, and multi-locational; (2) supporting evaluations by environmental descriptions and whole-plant modelling activities; and (3) drawing experience from advanced physiological characterization performed in crop-specific projects.

A Drought Tolerance Phenotyping Network, coordinated by EMBRAPA, is being put in place and reinforced for integrating materials of international origin for evaluation. The first months were dedicated to team organization and improving site-specific qualifications for the cereals and legumes phenotyping network. In the coming months, equipment will be

purchased or upgraded. Numerous trials are being conducted, and materials from the GCP reference samples can be integrated as they are identified and made available.

The groups involved in the EMBRAPA Phenotyping Network and Whole Plant Modelling (coordinated by CIRAD) met in May 2005 at Embrapa Arroz e Feijão (Embrapa Rice and Bean), Goiânia, Brazil, to initiate the collaborations.

The priority crops for the project will be upland rice, maize, and sorghum. The regions involved will be (1) in Brazil, covering the states of Piauí, Tocantins, Goiás, Minas Gerais, and, possibly, Mato Grosso—if data sets can be obtained; (2) in West Africa, covering the sub-Saharan region for sorghum and, possibly, upland rice crops; and (3) Central America for maize (to be confirmed).

A Target Population of Environments analysis will be carried out in all these regions and crops. An EMBRAPA scientist will be in charge and will collect the necessary meteorological and soil data with support from all institutions participating, according to the site and country where they are stored. The scientist will work at CIRAD, Montpellier, from September 2005 to August 2006, and spend 2 months at CSIRO, in Australia, in 2006 for training on the APSIM model and clustering tools.

Another EMBRAPA scientist will be trained at CIRAD on the SARRAH model in November 2005 and will build a specific model for maize. A scientist from Embrapa Meio-Norte (Embrapa Mid-North) will also be trained in November 2005 or 2006, to specifically adapt the model for cowpea and other legumes. At the end of the project, a general training course for EMBRAPA scientists may be organized by CIRAD, CSIRO, and EMBRAPA.

To carry out fine phenotyping, five accessions will be selected from the GCP reference samples. Different cases are identified:

- Upland rice: complementary experiments to be conducted in Brazil (Porangatu, Teresina, and Goiânia) during rainy and dry seasons.
- Maize: complementary experiments (to be confirmed) to be conducted, again in Brazil (Teresina, Sete Lagoas, and Janaúba).
- Sorghum: complementary experiments to be conducted in Janaúba, Sete Lagoas, Goiânia, and/or Porangatu during rainy and dry seasons. Existing data for West Africa will be exploited.

Assessing molecular characterization methods

To complement standard molecular characterization in elucidating germplasm structure, other technologies would be very useful if proven efficient. Single-nucleotide polymorphisms (SNPs) are the markers of choice for those crops where massive sequence data are available, such as ESTs from diverse germplasm. The quick evolution of SNP-detection technologies and the present relative efficiency of re-sequencing versus setting up SNP detection suggest that it is wise to remain as ‘observer’ for the time being. The present best option is to outsource or build alliances with partners for detecting SNPs in large numbers. When the appropriate links are established, whether commercial or partnership, then exploring updated SNP identification methods for those crops where sequence information is limited would be worthwhile. However, for such crops, alternative technologies may exist. For example, SP1 is assessing Ecotilling and diversity arrays technology (DART).

Ecotilling is used to identify SNPs in diverse germplasm pools without needing to sequence many genotypes. A pilot study is being run by IRRI, which has advanced experience on rice, and adaptation work is being done at Agropolis for sorghum and banana (triploid). A workshop on *SNP Discovery through EcoTilling* has been held for training colleagues from ARIs and 15 NARES (Bangladesh, China, India, Indonesia, Korea, Philippines, Taiwan, Thailand, and Vietnam). Primers were designed for a set of 16 tentative candidate genes given in the proposal. Unlabelled primers were obtained, and amplification

efficiency was confirmed for 12 of them. Labelled primers for use on the LI-COR® DNA Analysis System were obtained and verified for six genes. These primers will be tested for their efficiency for amplification on sorghum and *Musa*. Ecotilling data have been generated for 48 to 1000 rice accessions on 10 genes.

DArT represents a potential platform for whole genome profiling in orphan crops. The work has several components, to be completed by the end of 2005. For each of sorghum, rice, and wheat, 100 accessions have been selected, and existing DArT arrays will be hybridized within the next three months. This will allow comparing the diversity revealed by DArT with that revealed by SSRs.

In cassava, DArT technology has been proven efficient in revealing numerous polymorphic markers, using cultivars and wild relatives. A scientist from Thailand has started a 9-month stay at DArT Pty Ltd in Canberra, Australia, to prepare new libraries and extend the analysis to many cultivars, focusing on accessions with large variation for dry matter content. DNA was extracted from 122 cultivated and 19 wild accessions. Targets are being prepared for the first array hybridization.

The potential of the method for coconut and banana is also being explored. The group at CIRAD extracted DNA samples from 96 representative coconut accessions and sent it to DArT Pty Ltd for array development. A scientist from Sri Lanka has arrived in Canberra for a 6-month stay, and brought 20 Sri Lankan accessions. Several complexity reduction methods were tested and one proved promising (*PstI/BstNI*). The first library has been developed and the preliminary results, based on 130 polymorphisms among 31 genotypes, are also promising. Work on banana will start in early September 2005 when a scientist from France joins DArT for a 3-month stay.

Assessing linkage disequilibrium (LD)

The extent of LD in a crop is essential for determining which populations can be used to map or reveal favourable genes and alleles through association studies. It is being monitored at the species level in two contrasting species, namely the annual autogamous sorghum and the perennial allogamous coconut, as well as in specific populations of rice in Indonesia.

In sorghum, the study of 12 RFLP marker loci in a single region of about 5 cM has revealed significant LD, spanning 2 to 3 cM. It also exemplified how important it is to refine the reference sample before undertaking association studies.

In coconut, a method for assessing the level of LD in this highly heterozygous species has been refined: two closely linked SSR markers (about 1 cM) display strong LD in populations (namely Mozambique, Panama, Vanuatu, and Brazil), while unlinked markers displayed little or no LD. Data are being produced for another eight couples of SSR loci linked at 0 to 7 cM.

In rice, the study focuses on three regions bearing disease resistance genes. A researcher from Indonesia went to Cornell for 4 months' training (April–July 2005) and designed 250 PCR primers. Of these, 60 were in the *Xa7* region on chromosome 6, 60 around *Xa13* on chromosome 8, and 130 around the cluster *Xa4/Xa22/Xa26* on chromosome 11. Sequences from 8 diverse Indonesian rice accessions have been generated for 215 primer pairs. Twenty-five of the sequences from the 8 rice accessions have been aligned and SNPs have been called for 12 of them.

Population structure and association studies (LD mapping) in long-generation crops

Association analyses, in a broad sense, involves the comparison of molecular with phenotypic diversity. In an attempt to integrate such analyses with ongoing characterization and breeding activities, similar exercises of LD assessment have been undertaken for other cases where accurate phenotypic data are available and the materials amenable to LD mapping. Such cases include the breeding programmes for cassava at CIAT, potato at CIP,

banana at CARBAP (Cameroon), and coconut at VARTC (Vanuatu), and the germplasm evaluation programme for yam (*Dioscorea alata* L.) at VARTC (i.e. the Vanuatu Agricultural Research and Technical Center).

In the first three species, the possibilities of choosing the best materials have been analyzed. The preference will be for materials derived from breeding activities that represent a large, well-characterized, phenotypic diversity, preferably with known pedigrees and representing related materials that have had few meioses since their main foundation events.

Various options still exist; the assessment of LD levels will be of primary importance. The case is most clearly defined in coconut: the materials consist of 200 trees representing four generations of breeding materials. A total of 219 trees will be sampled at VARTC and DNA extracted at CIRAD, Montpellier, France. They will then be analyzed by a scientist from Sri Lanka during a training session in Montpellier in September–November 2005. A total of 31 loci will be surveyed, including 13 international reference markers and nine couples of linked markers. In addition, a coconut breeder from Vanuatu will visit CIRAD in October for a 3-week training session on DNA extraction, principles, and applications of molecular breeding and data management. For yam, the idea is that its insular history in Vanuatu may have involved bottlenecks that have established strong LD in its well-characterized populations. A mapping exercise is being undertaken to identify linked markers among the already used AFLPs, for quick assessment of LD.

Session I: Review of results for tier 1 crops

Genotyping composite germplasm sets: wheat

*Susanne Dreisigacker*¹ (with the collaboration of *Marilyn Warburton*¹, *Z. Xueyong*², *François Balfourier*³, and *Nachit Miloudi*⁴)

¹CIMMYT, Mexico City, Mexico

²CAAS, Beijing, China

³INRA, Clermont-Ferrand, France

⁴ICARDA, Aleppo, Syria

The GCP's composite germplasm set of wheat comprises 2012 *Triticum aestivum* L. accessions, 260 *T. durum* Desf. accessions, and 234 ancient wheats. *T. aestivum* accessions include 1312 wheat cultivars and breeding lines, 500 wheat landraces, and 179 synthetic wheats and synthetic derivatives. Twelve species of ancient wheats are included: among others, *T. dicocoides* Körn. ex Aschers, *T. dicoccum* Schrank ex Schuebl., *T. monococcu* L., and *Aegilops tauschii* Cosson. All the partners' laboratories are collaborating to fingerprint the entire composite set, which task is almost finished. The alleles are currently being recorded for data analyses.

All data will be placed in the GCP's central repository. In addition, an access database with all data, plus all data run by CIMMYT or AMBIONET to date, is being printed for distribution. The database contains many helpful tools for data manipulation, including tools for variable input and output formats, combining different studies, error calculations, and search functions.

To date, one data set, consisting of 18 SSR markers run on 854 *T. aestivum* accessions, is available for preliminary analysis. A total of 261 alleles, with an average of 14.5 alleles per locus, was determined. The proportion of unique alleles was 8%, and of rare alleles 22%. The average polymorphism information content (PIC) value was 0.48, and the average gene diversity 0.506. Because of the very large number of accessions in this data set, their genetic structure could not be clearly visualized. A reference sample or germplasm subsamples must be chosen if further analyses are to be carried out.

Cowpea genotyping

Sarah Hearne¹, Morag Ferguson¹, and Visvanathan Mahalakshmi²

¹IITA, c/o ILRI, Nairobi, Kenya

²IITA, Ibadan, Nigeria

This study had aimed to identify as many as 50 SSR markers suitable for diversity assessment; evaluate intra-accession variation in cowpea, and assess global cowpea genetic diversity. To identify markers, we screened all 121 SSR markers developed for cowpea by Li and Scoles at the University of Saskatchewan, Canada. However, many did not amplify or had too much non-specific amplification that could not be reduced by adjusting PCR conditions. We requested the original sequence data so we could re-design the primers, but were unsuccessful in our request.

So far, we have 15 SSR markers derived from the 121 cowpea SSRs and two from bean-derived SSRs (courtesy of Matt Blair, CIAT). These give clean PCR products and are polymorphic across a set of 24 diverse accessions. We are still working to develop more markers from EST-derived SNPs and SSRs (as part of an ongoing GCP SP2 project with TIGR) and identify and characterize other genomic library-derived SSRs (as part of IITA's own work). As these markers become available, we will test them for polymorphism across the diverse set, and then use them for screening. This work will be conducted in 2006.

To determine intra-accession variation, we aimed to assess variation in 100 cowpea accessions, using 10 plants per accession and 17 SSR markers (1000 individuals). This initial screening was conducted with the 17 polymorphic markers we currently have, and completed in mid-September. We are now assessing 1938 cowpea accessions held at IITA and another 200 accessions from CAAS, using all 17 markers. The origins of IITA's accessions ranged over 86 countries, with one third coming from Nigeria, Niger, USA, and India.

In attempting to accomplish our two goals, we suffered technical difficulties in several areas, preventing us from providing results for this workshop. These were:

- The lack of SSR markers that can amplify and produce clean clear products (as discussed above). This problem encourages us to recommend that all markers be proven before detailing proposals and work plans.
- The quality of the DNA we worked with. Unfortunately, the spectrophotometer at ILRI was not well calibrated. We are now using quantification gels.
- Through Biosciences eastern and central Africa (BecA), we received the use of new equipment, specifically the ABI3100 and ABI3730 capillary systems. These arrived 5 months late and experienced mechanical 'teething' troubles on installation, significantly delaying the initiation of work. Another problem lay in differences in polymer chemistry that meant that results of the two systems were not interchangeable. The 3100 is now being upgraded to use the same polymer as the 3730 to ensure that work does not stop when one machine is not operational. New projects should therefore be carried out on already tested equipment, enabling improved planning and scheduling.
- Kenya has recently implemented a new customs policy at all ports and airports, resulting in unforeseen delays in importing primers. The 'turn around' period of all orders has increased such that, to avoid delays, we now order when we have 50% of stocks.

Assessing the global diversity of cassava (*Manihot esculenta* Crantz) genetic resources, according to simple sequence repeat markers

Paula Hurtado¹, Jaime Marín¹, César Ospina¹, Charles Buitrago¹, Wilson Castelblanco¹, Ana María Correa¹, Edgar Barrera¹, Janneth Patricia Gutiérrez¹, Luis Guillermo Santos¹, Graciela Mafla¹, M. Carmen de Vicente², Sarah Hearne³, Morag Ferguson³, Alfredo Alves⁴, Claudia Ferreria⁴, and Martin Fregene¹

¹CIAT, Cali, Colombia

²Leader of SP5, GCP; IPGRI–Americas, c/o CIAT

³IITA, c/o ILRI, Nairobi, Kenya

⁴EMBRAPA–CNPMPF, Cruz das Almas, BA, Brazil

Cassava is native to the New World where it has been known for about 10,000 years and continues to be cultivated extensively from Mesoamerica to Colombia, Brazil, Paraguay, and Argentina. It was introduced into sub-Saharan Africa in the early 1500s from Brazil. In Africa, because of its excellent adaptability to erratic rainfall and low-fertility soils, it spread along trade routes into the interior of the continent. Cassava arrived in tropical Asia in the 18th and 19th centuries and is now a significant cash and food crop in Thailand, Indonesia, Vietnam, India, the Philippines, and China.

Cassava is an allogamous crop that is vegetatively propagated. This, together with the traditional slash-and-burn farming systems, has led to the development of numerous varieties being grown by small farmers in the Neotropics, Asia, and Africa.

Several germplasm collections, with hundreds to thousands of accessions, exist in international and national programmes. The collection held at CIAT is the largest, with over 5000 accessions, followed by the collections held by EMBRAPA (about 3000) and IITA in Nigeria (about 2000).

Under the auspices of the GCP's SP1, the genetic structure of a large and representative sample of cassava accessions held at the germplasm banks of CIAT, EMBRAPA, and IITA was studied, using simple sequence repeat markers.

A subset of 3000 cassava accessions (1500 from CIAT, 1000 from IITA, and 500 from EMBRAPA) was selected according to criteria of site and key agronomic traits such as drought tolerance, resistance to major pests and diseases, and adaptation to different ecologies. Passport data of all 3000 cassava accessions were collated and sent to SP4 for processing and storage. DNA samples were isolated from the 3000 accessions at the respective centres (except for accessions from EMBRAPA, which were extracted at CIAT from duplicate copies in the CIAT germplasm bank), assembled at CIAT, and aliquots of each sample sent to genotyping teams at IITA (Nairobi) and CIAT.

SSR marker analysis was carried out, using 36 markers (22 at CIAT and 14 at IITA), and the data collated at CIAT for analysis. Data analysis was based on SSR marker data from 30 loci (22 from CIAT and eight from IITA). The genetic structure was assessed, using principal coordinate analysis (PCoA) and multiple correspondence analysis (MCA), and the genetic diversity and allelic richness estimated.

Preliminary results revealed a separation of African accessions from the rest of the world, confirming findings from several other studies that showed that the global diversity of cassava germplasm is structured by region, with those from Africa and the Neotropics showing the highest differentiation.

Sources of such genetic differentiation may be selection for adaptation to agro-ecologies and their constraints, particularly disease, found in Africa; mutation; and even biased sampling. Data analysis is expected to be completed by the end of the year. In addition, SP4 has already received molecular marker data files from the diversity study, including

molecular weight information by locus and genotype; alleles by locus and genotype; and binary data by genotype.

Genotyping composite germplasm sets: maize

Susanne Dreisigacker¹ (with the collaboration of Marilyn Warburton¹, Alain Charcosset², Visvanathan Mahalakshmi³, and Yu Li⁴)

¹CIMMYT, Mexico City, Mexico

²INRA, Montpellier, France

³IITA, Ibadan, Nigeria

⁴CAAS, Beijing, China

The GCP's composite germplasm set of maize is divided into 987 inbred lines and 467 populations. The inbred lines are being fingerprinted by CIMMYT and CAAS, and the populations by CIMMYT and INRA. The fingerprinting is almost finished and allele records are now complete for data analyses.

Three data sets became available for preliminary analyses: (1) 40 SSR markers run on 956 inbred lines; (2) 30 SSR markers run on 100 populations; and (3) 15 SSR markers run on 400 populations.

A total of 481 alleles, with an average of 12 alleles per locus, was observed for the first data set. The proportion of unique alleles was 20% and of rare alleles (i.e. present in less than 5% of inbreds) was 55%. The average polymorphism information content (PIC) was 0.56, and the average genetic distance (using the Cavalli-Sforza and Edwards distance) was 0.724. Five main clusters were determined after applying the UPGMA clustering method adapted to the Cavalli-Sforza and Edwards distances. On the basis of these clusters, three "reference samples" containing 50, 75, and 100 inbred lines, respectively, were selected, using the Franco et al. D-sampling strategy. Each reference sample expressed equal measures on several diversity indices, compared with the entire composite germplasm set.

The second data set was composed of 100 landraces from Latin America. An average of 8.2 alleles per locus was observed. The proportion of unique alleles was 12%, and of rare alleles 29%. The average PIC was 0.59. The UPGMA cluster analyses clearly separated the populations according to their geographic origin into six main clusters.

The third data set included mainly landraces that were geographically distributed across America and Europe. Although 22 groups were observed, the data set was insufficiently complete for a clear structure of the germplasm to be revealed.

Genotyping composite germplasm sets: rice

Kenneth McNally¹ (with the collaboration of Ruaraidh Sackville Hamilton¹, Claire Billot², B. Courtois², L. Benoit², C. Martínez³, M. Lorieux³, C. Brondani⁴, L. Han⁵, M.-N. Ndjiondjop⁶, and S. McCouch⁷)

¹IRRI, Manila, Philippines

²CIRAD, Montpellier, France

³CIAT, Cali, Colombia

⁴EMBRAPA, Brasília, Brazil

⁵CAAS, Beijing, China

⁶WARDA, Cotonou, Benin

⁷Cornell University, Ithaca, NY, USA

We chose a panel of 51 SSRs that were suitable for fluorescence-based genotyping, according to McCouch and co-workers. This panel was split among our partners, with IRRI taking 24; CIRAD, CAAS, and CIAT seven each; and WARDA six. CIAT is genotyping four of its SSRs while EMBRAPA is running the other three.

The GCP's composite set of 3000 rice entries was designed by partitioning the set into gene pools and types, consisting of:

1. 2150 accessions of traditional *Oryza sativa* L. selected from five geographical regions (East Asia, South-East Asia, South Asia, West Asia, and Oceania) and two groups of regions (Europe together with North America, Oceania, and Africa with Central and South America);
2. 500 accessions of improved *O. sativa* selected from varieties, breeders' materials, genetic stocks, and other reference materials;
3. 300 accessions of the African cultivated rice *O. glaberrima* Steud.; and
4. 50 wild, inbreeding, AA genome accessions from *O. barthii* A. Chev., *O. glumaepatula* Steud., *O. meridionalis* Ng, *O. nivara* Sharma et Shastry, and *O. rufipogon* Griffith.
5. Because 1536 accessions had already been chosen from the IRRI *O. sativa* core collection, the 1464 remaining accessions were selected so that the target number was achieved whenever possible. In early October 2004, these accessions were selected according to N.R. Sackville Hamilton's MaxiMin approach, using GenStat.

Accessions were chosen from the nominated materials outside of IRRI as follows: 231 from CAAS (31 improved and 200 traditional), 106 accessions from CIAT (97 improved, 7 traditional, and 2 *O. glaberrima*), 61 accessions from CIRAD (47 improved, 10 traditional, and 4 *O. glaberrima*), and 98 accessions from WARDA (93 improved and 5 *O. glaberrima*). If the selected nominations were available at IRRI (e.g. those selected from CIRAD's nominations), the IRRI source was used.

DNA production was centralized at IRRI, using a modified CTAB method with gel densitometry quantification. Lyophilized DNA samples (500 ng) in 96-well plates were shipped.

Seed from the 106 CIAT entries were obtained in 2005, and DNA has been prepared from these. No materials have yet been obtained from WARDA or CAAS. WARDA's recent situation has made obtaining the materials difficult. The process for their importation into the Philippines is currently under way. For those lines from CAAS, even though DNA from these has already been prepared, the Ministry of Agriculture has yet to give permission for their shipment to IRRI and distribution among the partners.

The initial set of 1536 entries was shipped in two lots during 2004, with the first being in May–June for 48 samples to CIRAD, CAAS, CIAT, and EMBRAPA; and the second of 1488 samples to CIRAD, CIAT, and CAAS in October and to EMBRAPA in December (because of importation requirements). For WARDA, the 1536 samples were shipped to Cornell in March

2005 in time for a WARDA technician to arrive at Cornell to help with the genotyping. Of the remaining 1464 samples needed, DNA has been shipped (August 2005) for 1221 entries. Some of these samples replace some of those that were to have been obtained from outside of IRRI, most of which were nominations from CAAS.

For the first set of 1536 entries that were almost entirely *O. sativa*, IRRI has finished 6 SSRs on a MJ Research BaseStation 51, with another 6 almost completed. The remaining SSRs will be genotyped either on the BaseStation or on a LI-COR® 4300. CIRAD has finished seven SSRs, using LI-COR genotypers. CIAT has genotyped two markers on an ABI, with EMBRAPA genotyping their three, also on an ABI. CAAS has encountered technical difficulties with their ABI and has not yet completed any genotyping. WARDA has genotyped six SSRs at Cornell, using an ABI 3730.

Technical difficulties have slowed the process: equipment (IRRI) and primer failures (CIAT, CAAS), shipment of supplies (IRRI, CAAS, CIAT), software issues (IRRI, CIRAD), and IPR issues (IRRI, CAAS, EMBRAPA). Genotyping the remaining SSRs and entries is under way at all sites with completion expected by December 2005. Meanwhile, preliminary data analysis indicates a star-like population structure for *O. sativa* entries.

Sorghum genetic data analysis

*Claire Billot*¹ (with *Tom Hash*²)

¹CIRAD, Montpellier, France

²ICRISAT, Patancheru, Andhra Pradesh, India

Sorghum is the fifth most important cereal cultivated in the world. *Sorghum bicolor* (L.) Moench ssp. *bicolor* includes five races and ten intermediates, found throughout Africa and Asia and reflecting patterns of domestication and introgression. The main objective of the GCP–Tier 1 Genotyping Project was to assess the genetic diversity of 700 accessions, using 30 SSR markers. Work was shared between three Consortium members: ICRISAT (the Principal Investigator for Tier 2 work), CAAS, and Agropolis–CIRAD (PI for Tier 1 work). The work was defined in terms of three tasks:

- *Task 1: choosing 50 SSRs.* We chose 104 SSR markers from among published and unpublished markers, using criteria such as previously known genetic diversity, previous experience of each laboratory, and known map position. Each linkage group was separated into 5 bins and, where possible, 1 SSR marker was chosen from each bin. In all, 104 SSR markers were chosen and tested on 48 accessions in each laboratory.

The results from the different laboratories were compared for the level of heterozygosity detected and congruency between allele sizes. We could, although with great difficulty, choose 34 markers, which were then shared among the laboratories, according to linkage group position, emphasizing mandatory use of the same controls between laboratories so that an allelic ladder could be used.

- *Task 2: analyzing the genetic data from the 700 accessions.* Of the different accessions available in worldwide collections, 700 were chosen (SP1, Cluster 1 activity) and genotyped in each laboratory. Only 20 loci remain for analysis, showing considerable genetic diversity that is roughly structured according to racial and geographical origins.

In the set of 700 accessions, a subset of 200 accessions was also studied with RFLP markers. The comparison of the two analyses highlighted some homology (i.e. organization of accessions into groups) and also some differences (i.e. separation of accessions of the same race into different groups and according to geographical origins).

- *Task 3: developing new markers.* New EST-derived SSR markers, based on rice EST sequences (ICRISAT) and genomic SSRs (Agropolis–CIRAD), have been developed, from which the remaining 20 markers will be chosen.

A further genotyping of 2300 accessions will be performed towards the end of 2005 and a finer analysis will be done at that time.

Common bean genotyping

Mathew Blair

CIAT, Cali, Colombia

The bean genotyping project has evaluated a minicore collection of 44 genotypes with 150 markers to choose the best SSR markers for use in full genotyping experiments. Heterozygosity was determined for 57 gene-based and 72 genomic microsatellites. Fluorescent microsatellite panels were also designed for 75 markers and implemented for 40 markers.

A decision was made to establish race structure with an additional minicore of 120 genotypes representing all the morphological races of common bean, using 30 markers. Analysis showed that the race structure broadly agreed with morphological classification. However, new patterns of diversity were uncovered. Microsatellite analysis was very useful for distinguishing Nueva Granada from Peru races and the Guatemala race from other Mesoamerican races. Genotypes misclassified by morphological analysis or varietal name could be placed into their correct race, using microsatellite genotyping.

Mass genotyping was then carried out with 30 or more SSR markers on accessions from partner institutions EMBRAPA (Brazil; 560) and CAAS (China; 230), and from the CIAT germplasm collection for primary centres of diversity in Bolivia (200) and Colombia (200). A secondary centre of diversity was distinguished in Cuba (210) and other parts of the Caribbean (310). Accessions included landraces, modern varieties, and some breeding lines chosen from each national and FAO-designated collection. Genotyping of the core collection (200 Andean and 200 Mesoamerican) was begun.

Several bean researchers from Bolivia (2), Brazil (1), China (1), Colombia (4), and Cuba (1) were trained to evaluate genetic diversity in the species.

Diversity of barley genetic resource in China and progress in barley genotyping for the GCP

Zhang Jing¹, Zhang Chihong¹, Michael Baum² & Jan Valkoun²

¹Institute of Crop Science, CAAS, Beijing, China

²International Center for Agricultural Research in the Dry Areas (ICARDA)

Barley, a major world crop, has the longest history of cultivation. Not only is it a staple crop in most uplands, providing food for most of the world's poor people, but it is also the main raw material used by beer and animal feed industries in wealthy countries. With global warming, however, natural calamities such as drought, salinity, strong winds, and flooding are more frequent than before, greatly limiting the productivity and sustainability of barley production. Exploring and using related new genes from the barley gene pool will help resolve these problems, but the crop's genetic diversity must be understood.

China's Gene Bank holds 19,015 accessions of barley, of which 12,489 are native and 6526 exotic. The genus *Hordeum* has 29 species, 23 of which are represented by 176 accessions. The remaining 18,839 accessions belong to *H. vulgare* L., of which 2585 are wild barley, including 695 that are ssp. *spontaneum* and 1890 that are ssp. *agriocrithon*, collected mostly as weeds from Tibetan barley fields. Another 16,254 accessions are cultivated barley, with 3891 being ssp. *distichon*, 12,359 ssp. *hexastichon*, and four ssp. *intermedium*. All 18,839 accessions have been classified into 544 varieties, based on the morphological characters of their spikes and caryopses. All 19,015 accessions have been characterized, and many evaluated for agronomic traits.

When we used the Shannon–Weiner formula, we found that, within the 18,839 *H. vulgare* accessions, the phenotypic diversity of our native barley was higher than that of the exotic accessions. Wild barley was the most diverse, followed by landraces, and then bred cultivars, which had low phenotypic diversity. No significant relationship was found between diversity and geographic region (provinces), although diversity tended to concentrate in three regions: West and South-west plateaux, the Yellow River Valley, and the central-lower valley of the Yangtze River.

The Institute of Crop Science and ICARDA collaborated to genotype 500 barley accessions with 49 SSRs, each institution providing 250 accessions. We extracted DNA samples from fresh leaf tissues, and performed PCRs with 49 pairs of SSR primers evenly spread in the seven barley chromosomes. The data in peak values were transformed into a 0/1 format and analyzed on a preliminary basis. Alleles at each locus were counted and differences compared simply among chromosomes, countries and provinces, and between wild and native barley.

We found 2307 alleles at 49 loci in total. On average, 47 alleles were found at each locus. Locus Scind1691 on chromosome 5H had the highest number with 78 alleles, followed by Bamc0063 with 71 alleles on chromosome 1 H. Scsr02306 had the lowest number with only 19 alleles.

We found similar trends in changes of allele numbers for the seven chromosomes between world and Chinese barley. Chromosomes 3H, 5H, and 6H had more alleles than the others, 1 H and 7H had the lowest allele number. China had the most, followed by Iran, Afghanistan, Ethiopia, Jordan, Syria, and Turkmenistan. However, the number of alleles may be related to sample number.

For China, barley from Tibet had the highest allele number, followed by Yunnan barley. Barley from Zhejiang, Jiangsu, Qinghai, Shan'xi, and Heilongjiang also had high allele numbers.

Overall, allelic diversity correlated with phenotypic diversity, with bred barley having the lowest allele number, similar to that of wild barley collected in Tibet, Qinghai, Si-chuan,

and Yunnan. Landraces held significantly more alleles than did the wild or bred barley, demonstrating that, in China, alleles mainly exist in landraces, not in wild barley.

Gene pool structure of cultivated potatoes assessed by SSR marker analyses

Marc Ghislain¹, Jorge Núñez¹, Guillermo Trujillo¹, María del Rosario Herrera¹, Frank Guzmán¹, and David Spooner²

¹CIP, Lima, Peru

²ARS-USDA, Beltsville, MD, USA

The high-throughput genotyping facility established in 2004 has produced a large data set of microsatellite (SSR) markers for cultivated potato. The objectives of this research were to analyze the gene pool structure of the cultivated potato and increase the number of SSR markers for potato germplasm analyses. Potato landraces from each cultivar groups have been included in the potato composite genotyping set. So far, 716 genotypes by 53 SSR markers constitute the potato SSR data set. The analysis of this large data set was performed by cluster analyses, using DARwin software 4.0 kindly provided us by CIRAD. The results revealed both expected and unexpected grouping of the cultivated potato germplasm:

- The tetraploid cultivated potatoes grouped together into a larger cluster distinct from the diploid cultivar groups even though almost all of the SSR alleles found in diploid landraces were also found in tetraploid landraces.
- Within the tetraploid potatoes, the Chilean *Chilotanum* Group forms a clearly separated cluster from *Andigenum* potatoes. This finding agrees with the hypothesis that the *Chilotanum* Group derives from the *Andigenum* Group through hybridization with possibly a wild potato species.
- Within the diploid potatoes, the *Phureja* Group forms a distinct cluster, which was rather unexpected based on previous studies at CIP that had used RAPD markers.
- The *Stenotomum* Group forms a wide and highly diverse group, with one major cluster composed of cultivars of the *Goniocalyx* and *Stenotomum* Groups. This finding was expected as the cultivars of the *Goniocalyx* Group are recognized by potato taxonomists to be a sub-group the *Stenotomum* Group.
- The triploid *Chaucha* Group does not form distinct clusters; instead, its landraces are found intermixed with all the other groups which is probably an evidence of its hybrid origin between diploid and tetraploid potatoes.
- Finally, the diploid *Ajanhuiri* Group, the triploid *Juzepczukii* Group, and the pentaploid *Curtilobum* Group grouped into three distinct clusters markedly separated from all other cultivar groups. This may indicate independent domestication from other wild ancestors. Many of these potatoes are indeed referred to as 'bitter potatoes' and are usually found in the high Andes.

In conclusion, the SSR markers are revealing interesting genetic structure of the South American potato landraces. Although very few private alleles were found, their occurrence might be only due to limited sample size, groups of genotypes with certain SSR allele combinations appear to exist. These are forming various well-separated clusters which match remarkably well the traditional cultivar group taxonomy with the exception of the triploid cultivars of the *Chaucha* Group. Hence, SSR markers are uncovering higher degree of population structure than conventional descriptors of genetic diversity.

Mining the chickpea composite collection for allelic variation

Hari Upadhyaya¹, M. Baum², H.K. Buhariwala^{1,3}, S.M. Udupa², B.J. Furman², S.L. Dwivedi¹, David H. Hoisington¹, P.M. Gaur¹, S. Chandra¹, C.L.L. Gowda¹, J.H. Crouch³, and Jan Valkoun²

¹ICRISAT, Patancheru, Andhra Pradesh, India

²ICARDA, Aleppo, Syria

³Present address: CIMMYT, Mexico City, Mexico

Chickpea, *Cicer arietinum* L., is believed to have originated in south-east Turkey. However, at present, the major chickpea-growing countries are India, Pakistan, Iran, Turkey, Australia, Ethiopia, and Mexico. Chickpea is a leguminous food crop, self-pollinating, and diploid. Its gene pool consists of 43 species: one annual cultivated (i.e. chickpea), eight annual wild, and 34 perennial wild species. Two types of chickpea are known: desi types with coloured flowers, and angular-shaped and dark-coloured seeds, primarily grown in South Asia and Africa; and kabuli types with white flowers, owl's head-shaped and beige-coloured seeds, and grown mostly in Mediterranean countries. To study the allelic richness and diversity associated with beneficial traits, a composite set of 3000 chickpea germplasm accessions was constituted. This set included the chickpea core collection, old and new cultivars and trait-specific germplasm accessions from ICRISAT and accessions representing the ICARDA collection.

Some progress on genotyping chickpea accessions has already been achieved. In 2004, a set of 288 chickpea accessions that included 211 minicore subset accessions (75% desi type), 57 accessions of kabuli type, and 20 accessions of wild *Cicer* species were genotyped, using 35 SSR markers at ICRISAT and 15 at ICARDA. Preliminary analysis revealed a broad allelic diversity in this set, detecting 873 alleles, with an average of 25 alleles per locus.

A smaller number of alleles (averaging 11 alleles) was detected for dinucleotide motifs than for trinucleotide motifs (averaging 27 alleles). Similarly, gene diversity was lower for the dinucleotide motif (average 0.723) than for the trinucleotide motif (0.898). The mean gene diversity of all the SSR markers was 0.873. The dendrogram constructed as per shared allele distance, using the unweighted pair-group mean average (UPGMA) method, indicated two main groups: one consisting mainly of accessions from the Indian subcontinent, and the other from the Mediterranean Region, Middle East, and Ethiopia. The wild species (*C. reticulatum* Ladiz. and *C. echinospermum* P. H. Davis) were split into two groups, flanking two ends of the group of chickpea accessions.

Discriminant function analysis (DFA) was used to determine the extent to which the genotypic data supported the 28 clusters from which the chickpea minicore was selected based on phenotypic data. For DFA, 40 SSR markers data on 210 minicore accessions were used. Overall, most individuals were assigned to the original phenotypic clusters. Only 27% of the individuals were re-assigned into new clusters according to data from markers, identified mainly within clusters 4, 6, and 7 of the minicore.

Genotyping composite germplasm sets: *Musa**Isabelle Hippolyte¹ and Nicolas Roux² (presented by Claire Billot¹)*¹CIRAD, Montpellier, France²INIBAP, Montpellier, France

The *Musa* plant is the world's largest herb, growing to as high as 15 meters. This monocotyledonous perennial presents a pseudostem, or false stem, which develops as its enormous leaves wrap around each other before unfurling their large green fronds. It bears its inflorescences of female, hermaphrodite, and male flowers on the same peduncle. The flowers' long ovaries develop into fruit known as bananas or plantains. Although reproduction can be sexual, it is largely vegetative, by means of suckers.

The *Musa* species—*M. acuminata* Colla (AA), *M. balbisiana* Colla (BB and BBB), and *M. schizocarpa* N.W. Simmonds—involved in the GCP belong to the Eumusa section, from which most edible varieties are derived. *Musa* species are diploid, triploid, or even tetraploid, with their chromosome sets coming from two different genomes—A and B. Edible varieties derive from crosses of the two genomes, interspecific hybrids, or selection of phenotypic variants. Typically, these varieties are triploids, that is, with $2n=3X=33$ chromosomes, and can be grouped into three major types: AAA (Cavendish or dessert bananas), AAB (plantain), and ABB (cooking or dessert bananas). The diploid, type AA, is also common.

Many collections of *Musa* are found throughout the tropics. The largest—an *in vitro* reference collection of 1144 accessions—however, is held in Belgium by the INIBAP Transit Centre (ITC).

A diversity analysis was conducted with the collaboration of IITA, University of Leicester, CARBAP, CIRAD/Agropolis, and University of Georgia. Regrouping all information, and taking into account those accessions for which a sample was held at the ITC, a collection of 567 accessions was created. Composed of three sets, it includes materials chosen from major collections held by CIRAD–Guadeloupe, IITA, and CARBAP. To evaluate the materials, 50 microsatellite markers, half provided by IITA and the other half by CIRAD, were selected.

For the first set (48 accessions), some of the experiments need to be repeated for verification, and, for the third set, analyses are yet to be done. The second set, from IITA and for which analysis is furthest advanced, 90 accessions were eventually chosen for analysis with 26 SSRs from CIRAD.

Ninety genotypes were recognized, with 13 SSR loci defined for each genome, involving 300 alleles, ranging from 5 to 22 per genotype, with a mean average of 11. They were then genotyped and statistically analyzed, using the Jaccard index, factorial analysis, and bootstrap clustering. The scattergrams from the factorial analyses and the cluster analysis confirmed that the markers successfully differentiated the accessions into the major *Musa* groups: Mutika/Lujugira (AAA), Cavendish/Gros Michel/Red (AAA and AA), plantain types derived from hybrids (AAB, ABB, and AB), and BB (derived from *M. balbisiana*).

Session II: Selected Experiences

Characterizing and managing genetic diversity in the bread wheat collection held at INRA

François Balfourier

INRA, Clermont-Ferrand, France

INRA executed, in 2000 to 2004, a large project to describe its bread wheat genetic resources collections. Specific goals were to (1) understand how the crop's genetic diversity is organized, (2) document the diversity of bread wheat within historical and selective scenarios, and (3) establish a core collection suitable for advanced characterization.

INRA, at his station in Clermont-Ferrand, holds about 10,000 accessions of bread wheat, most of which are landraces, but also include old varieties released during the last two centuries and elite lines developed only recently. Although these accessions originate from more than 80 countries, one third is from France, another third from other European countries, and the rest from other parts of the world.

The entire collection was first evaluated for about 10 agromorphological traits, and its diversity described with near-infrared spectroscopy. From these preliminary evaluations, we selected 3946 accessions, covering more than 40 different geographical origins and 7 consecutive registration periods from mid-19th to late 20th centuries. This large sample was then evaluated at the molecular level, using a set of 42 SSR loci that had been selected according to their location on the chromosome arm, readability, and reproducibility. Where possible, each locus was located on a different chromosome arm to cover the whole genome of bread wheat. Diversity was evaluated at different levels:

- *Within-country diversity* was observed for the French accessions, with AMOVA indicating that both temporal (period of release) and breeder effects (breeder origin) were significant. Landraces clustered separately, while the 7 temporal groups clustered into two sets, one before and one after the 1960s.
- *The between-country diversity* was also evaluated, considering only European accessions. Both temporal and geographical effects appeared significant. The accessions clustered according to time in exactly the same way as observed for the French cultivars. One hypothesis to explain such a temporal drift is a bottleneck effect linked with the 'Green Revolution'. The combination of this bottleneck effect and the introduction of completely new germplasm in some European countries during the 1970s and 1980s (i.e. *Rht* genes, introgressions from *Aegilops* sp.) is probably responsible for the accessions clustering into two main groups.
- *Analysis for geographical structure* showed accessions separating into groups from either north or south of the arc formed by the Alps and Carpathian Mountains. This could be explained by the wheat germplasm's initial adaptation to different climatic and environmental conditions. For example, north-western countries present large agro-ecological areas that are relatively homogenous for soil and climatic characteristics. In contrast, the Mediterranean Region possesses relatively small agro-ecological areas with highly diversified climatic conditions and soil characteristics.

The separation of accessions into northern or southern groups for Europe may also reflect the oldest migration paths of the initial wheat germplasm brought by the first farmers from the Middle East into Western Europe, when the Alps and Carpathian Mountains may have formed a natural barrier to both human and gene flows.

1. *The overall worldwide diversity structure* was examined for the entire sample of 3946 accessions, which had clustered into 40 groups according to geographical origins, of which three main ones were observed: (1) European wheat accessions, (2) Asian wheat

gene pool, and (3) accessions from countries strongly affected by the 'Green Revolution' (South America, Middle East, and North Africa).

Overall, analysis at the molecular level showed that wheat germplasm diversity is not randomly distributed, but structured temporally and geographically as a result of the combined effects of both the historical processes of wheat germplasm evolution (distinction between Asian and European germplasm) and recent breeding practices throughout the world ('Green Revolution').

The above-mentioned results were used to construct a core collection (372 accessions) from the base collection (3946 accessions). The M strategy, which maximizes diversity, was applied not only to the SSR allele number, but also to the number of geographical origins and number of different registration periods of accessions.

To make the DNA analysis easy, controlled analysis was conducted on a high-throughput platform, using a full, deep, 384-well plate. So far, results indicate that the core collection captures more than 98% of the total allelic diversity measured at 38 SSR loci in the base collection. All the different geographical origins and registration periods are also represented. Once the accessions have undergone seed multiplication, the core will be available for further genomic analyses.

Barley research at IPK

Andreas Graner

IPK, Gatersleben, Germany

The IPK's Gene Bank develops genomics-based strategies to improve the use of plant genetic resources for plant breeding. Most of these efforts focus on barley (*Hordeum vulgare* L.) because the 7 chromosomes of this species represent the base complement of all Triticeae genomes and thus serve as a model to extrapolate results into related cereals such as wheat.

To generate a basis for systematic genome analysis, a series of resources have been developed in recent years. These include:

- A comprehensive collection of 183,000 expressed sequence tags (ESTs), covering more than 15 different tissues and developmental stages of the barley plant. These ESTs formed the basis for the development and genetic mapping of gene-derived SSR, SNP, and RFLP markers. The resulting transcript map of the barley genome currently comprises more than 1000 genes.
- A barley cDNA array has been developed for studies in functional genomics. It comprises 10,000 genes and is being used to perform transcriptional profiling in the context of resistance to biotic and abiotic stresses and seed quality.
- As additional resources, a 'targeting induced local lesions in genomes' (TILLING) population is being developed and a transformation facility has been established.

With these resources, candidate genes for selected agronomic traits are being identified. These, in turn, form the prerequisite for developing strategies aimed at an allele-based use of PGRs. Two case studies on isolating candidate genes are described below:

1. *For the complex trait 'malting quality', we hypothesize that differences are, at least partly, based on differences in gene expression. A functional association approach was developed to identify candidate genes. To this end, a cDNA array was used to analyze gene expression in germinating seeds of 10 unrelated barley cultivars. By correlating the transcript levels with the phenotypic data available for various malting quality components, a set of candidate genes could be identified, whose variation in expression correlated with the variation in the corresponding traits. In a further verification experiment, the expression through a QTL of one of those candidate genes (serine carboxypeptidase or *Cxp1*) was shown to correlate with a QTL for the trait 'diastatic power'. As the gene was shown to be regulated in cis, intragenic SNPs could then be used to screen a germplasm collection for genotypes that show high and low expression levels of this gene. Thus, marker-assisted selection could be performed at the expression level of this gene.*
2. *The rym4 locus complex confers resistance to the barley yellow mosaic virus. The complex was isolated by map-based cloning. Resistance is conferred by recessive mutations in the coding region of the gene *eIF4E*, which represents a component of the translation machinery of the plant cell. Several of the mutations, which are confined to a defined area of the protein—which area is the supposed site of interaction with the viral protein—are diagnostic for alleles that lead to altered resistance specificities. This knowledge is now being exploited to search a core collection of 2000 barley accessions for the presence of novel, functional alleles. The sequence analysis of a set of 220 accessions led to the identification of several novel alleles, which now are further studied on the phenotypic level. To speed up the process of allele mining, an Ecotilling strategy is being established.*

To conclude, the availability of gene-specific sequence information, together with recent advances in bioinformatics, opens up new opportunities for using genetic resources. Traditionally, use was based on the phenotype and was thus mainly restricted to traits that

were easy to measure such as disease resistance or photoperiod response. The availability of the candidate genes that underlie agronomic traits will facilitate a gene-based use of genetic resources, which, at least in theory, can be extended to any given trait.

Although many hurdles still need to be overcome, the application of allele-mining strategies is expected to result in a quantum leap in the use of PGRs. They will not only allow a deliberate selection of the most appropriate germplasm but also will provide access to those collections, which, because of their large size, are hard to handle by conventional means.

Maize at Cornell University

Gael Pressoir

IGD, Cornell University, Ithaca, NY, USA

We use functional genomic approaches to dissect complex traits in plants, specifically maize; exploit the natural diversity of these plant genomes; and identify the individual nucleotides responsible for quantitative variation. This information can be found at www.maizegenetics.net.

For maize diversity-based genomics, we are developing a platform to rapidly dissect complex traits in maize by using both association and linkage-based approaches. To conduct these analyses, we must develop linkage and association populations that capture much of the natural variation inherent in the maize genome. Extensive phenotyping and surveys of tens of thousands of candidate gene sequences will then be employed. The development and adaptation of novel statistical genetic approaches are also required to study these diverse mapping populations. Such activities should allow the rapid dissection of complex traits down to the gene level.

In trait dissection, a full range of genomic and field genetic approaches is being used to identify alleles involved in such traits as improved nitrogen efficiency, plant and inflorescence architecture, and kernel quality.

Making the connection between genomics and plant breeding remains a formidable challenge for current bioinformatics tools. We are developing improved bioinformatics tools that will integrate public databases with genomic diversity data and agronomic data.

Session III: Review of other ongoing SP1 projects

Assessing Ecotilling as a methodology for targeted genotyping and SNP discovery: rice, sorghum, and *Musa*

Kenneth McNally¹ (with the collaboration of Ruairadh Sackville Hamilton¹, Claire Billot², M. Deu², I. Hippolyte², F.-C. Baurens², and J.-F. Rami²)

¹IRRI, Manila, Philippines

²CIRAD, Montpellier, France

A set of tentative candidate genes has been identified in rice. These are protein phosphatase 2a-4 (on CHromosome 10), DREB2 (CH1, annotated by TIGR as DREB1), AP2 domain TF similar to the *Arabidopsis thaliana* (L.) Heynh. EREBP4 gene (CH1), trehalose-6-phosphatase (TPP, two loci on CH2 and one on CH7), viviparous-14 or NCED (vp14 on CH12), 14-3-3 protein (CH2), MAPK (CH7), extensin (CH10), TRAB1 ABA responsive TF (CH10), Ca-dependent protein kinase or CDPK (CH2), Fiery1 for inositol polyphosphate 1-phosphatase (CH3), sucrose synthase (CH7), BZIP protein (CH1), and actin-depolymerizing factor (ADF, two loci on CH2 and CH7).

Primers were designed for these genes and checked by e-PCR, using BLAT to the TIGR assembly version 2 for rice. Unlabelled primers were obtained for all the genes and the amplification efficiency was checked. Good amplification products with the expected size were obtained for most target primer pairs, except those for one of the TPP loci on CH2, the ADF locus on CH7, TRAB1, and Fiery1. New primers for these loci are being designed. Labelled primers for use on the LI-COR system have been obtained and verified for DREB2, EREBP4, AP2 domain TF, one of the TPP loci on CH2 and the one on CH7, MAPK, vp14, and the two ADF loci on CH2.

These primers need to be tested for their efficiency for amplification on sorghum and *Musa*. Furthermore, the sequence homology of these loci to known sorghum and *Musa* sequences needs to be defined. Meanwhile, tests have been performed with the waxy gene on sorghum and simulations of polyploidy were performed with bulks of genetically diverse DNA.

A modified Ecotilling procedure has been adopted at IRRI that allows detection of cleavage products on agarose gels. This effort was based on reports by several authors who suggested that Cel1 cleavage products could be detected on denaturing PAGE with SYBR gold detection for tomato and on agarose gels for mouse mutants and human mitochondrial mutants.

For this modified technique, we compared the efficiency of detection on agarose for putative SNPs detected on the LI-COR system, using the GCP microcore panel of 48 accessions for one of the TPP loci on CH2, the ADF locus on CH7, and the 14-3-3 protein on CH2. The results were comparable. This modification will allow quicker and less costly screening than on the LI-COR platform, as unlabelled primers are used and shorter gel runs are possible. This technique makes SNP discovery and genotyping possible for NARES because only agarose gels and ethidium bromide staining for visualization are needed.

Another modified procedure has been developed at CIRAD with the use of SP6 and T7 tailed primers. It significantly reduces the cost of labelled primers, especially for small numbers of individuals. Tests comparing the two methods have shown no significant differences and, hence, the modified method is used extensively for sorghum at CIRAD.

Ecotilling on the LI-COR system has been accomplished for 48 (for all available labelled primers) to 1000 (for the putative AP2 domain TF) GCP rice accessions at IRRI. We have accomplished Ecotilling on agarose for 48 GCP accessions for BZIP, CDPK1, sucrose

synthase, the 14-3-3 protein, one of the TPP loci on CH2, and MAPK. We have sequence data on selected, representative haplotypes for the putative EREBP4 homolog and DREB2.

A total of 31 polymorphic sites occur in the aligned sequences for EREBP4 across 1.8 kb of upstream, UTR, and coding regions. These sequences group into nine haplotypes. DNA parsimony analysis with bootstrapping reveals three main groups corresponding to *indica*, *aus/boro*, and *japonica*, plus *aromatic* variety groups. Phenotyping data on 1536 entries of the GCP's composite rice set is being generated for upland and lowland drought tolerance (data on 600 entries so far), blast resistance, and grain quality. Association tests with the EREBP4 haplotypes and upland drought tolerance and yield component traits revealed positive association with lower yield decline and other traits in one of the *indica* haplotypes.

Assessing linkage disequilibrium in sorghum

Jean-Christophe Glaszmann, Monique Deu and Claire Billot

CIRAD, Montpellier, France

Linkage disequilibrium (LD) is defined as the association between alleles at distinct chromosomal loci in a population. It usually connects genetically linked loci, but can also reveal associations between unlinked loci when the population has an internal structure. Usually, LD in a crop relates to founder effects and admixtures involved in domestication and breeding. It is eroded by recombination (and occasionally by gene conversion) and thus remains essentially between linked loci. If LD extends sufficient distances (e.g. in the centiMorgan range), genome-wide LD mapping in population samples becomes feasible.

LD extension contrasts among grass crops, depending on the reproduction system and population history. For example, for maize (allogamous), association distances among landraces range from 0.4 to 7 kb, whereas for sugar-cane cultivars (also allogamous but of very recent origin), distances can reach 20 to 30 cM. Among autogamous cereals, rice landraces display LD over about 100 kb, whereas modern barley and wheat varieties display LD of up to 10–20 cM.

Sorghum bicolor L. (Moench), an autogamous annual crop, is described as having five main races that show considerable variation for panicle and seed shapes, sizes, and colours. Monique Deu and colleagues have developed a minicollection of 205 sorghum accessions with accurate neutral marker characterization. We examined two sets of data: 74 RFLP markers scattered across the genome, and 12 RFLP markers that were located in the distal part of linkage group D. Markers presenting strong LD were 7 SSRs specifically developed from an SSR-enriched library (French GènoPlante Project: “Development of genomic tools for sorghum” and BAC clones localized in a distal 4-cM region on linkage group D.

Using the 74 dispersed loci, clusters formed according to geographical origin—South Africa, Asia, and West Africa—and to a guinea-race sorghum (*S. margaritifera* Stapf) (Dice dissimilarity index, NJ tree). The 12 linked RFLP loci not only displayed numerous associations among themselves, but also with loci in other chromosomal regions.

To examine the LD further, we used approaches developed by Xavier Perrier (SP4 project) to derive subsamples with less structure than the initial 205 accessions. Using the Dice dissimilarity index expressed on an NJ tree, we pruned the tree of the most redundant accessions, which gave us a star-like distribution of 100 genotypes (E = ‘extracted sample’). From the original 205, we also selected a branch (B sample) of 71 genotypes. The number of structure-related associations between unlinked loci drastically decreased, whereas associations within the small target region decreased only slightly. Almost all the alleles were retained in the E sample (average number of alleles per locus was 2.84 instead of 2.94), whereas it decreased significantly in the B sample (2.46).

Based on our target region, LD in sorghum can span regions of 300 kb or 2 cM making association studies of coarse resolution. However, the global structure of germplasm collections is a major constraint for association studies and needs to be optimized.

Among the SSR loci, those that presented strong LD with close RFLP markers generally displayed multimodal allele size distributions, each mode being associated with RFLP alleles nearby. Considering these results, we suggest that, for LD studies, SSR alleles should be grouped within size modes if they exist.

Assessing linkage disequilibrium in coconut

Luc Baudouin, Angélique Berger and Patricia Lebrun
CIRAD, Montpellier, France

Coconut oil represents 3% of the world's oil market. Although this proportion is slowly declining, its economic importance to producing countries is still significant: two thirds of total production is used locally and 96% of it is produced by low-income farmers. For those farmers, it represents both an essential staple food and the main source of cash. But there is more to coconut than just oil: every part of the plant is exploited for edible or inedible uses.

Conventional breeding schemes, by developing hybrids between self-pollinating 'Dwarf' coconut and cross-pollinating 'Tall' coconuts, have succeeded in producing high-yielding and early bearing cultivars. However, they are rarely used because they are expensive and not always suitable for the various domestic uses.

Another major challenge for breeding is to improve resistance to several lethal diseases, including lethal yellowing, cadang-cadang disease, the Vanuatu coconut foliar decay, and phytophthora bud rot. For most of these diseases, artificial inoculation is not feasible and one must rely on resistance trials, the results of which may come years after field establishment and are often difficult to interpret.

Molecular breeding through association is expected to help plant breeders select coconuts on multiple criteria and save time by identifying resistance factors in areas where epidemic diseases are present. Inexpensive high-throughput genome-wide markers like DArTs are making this approach feasible.

In our project, we selected nine closely linked locus pairs and compared linkage disequilibrium (LD) at matched and unmatched locus pairs. We also collected a breeding population in Vanuatu. Genotyping was started for more than 2000 individuals, representing global diversity. So far, 84% of this task is completed. A Bayesian method for LD analysis was designed and the preliminary version of the program was used to study four populations with large numbers (Brazil, Mozambique, Vanuatu, and Panama).

On average, LD is stronger in matched locus pairs than in unmatched ones. However, there were instances of significant but moderate LD in unmatched pairs (possibly caused by subdivision) and of absence of LD in matched pairs. The Bayesian method appeared to work satisfactorily. It is relatively less demanding in calculation time, as resampling is not required. Furthermore, the model seems amenable to refinement to accommodate inbreeding and subdivision.

A unified mixed-model method for association mapping: accounting for multiple levels of relatedness

Jianming Yu¹ and Gael Pressoir¹ (with the collaboration of William H. Briggs², Irie Vroh Bi¹, Masanori Yamasaki³, John F. Doebley², Michael D. McMullen^{3,4}, Brandon S. Gaut⁵, James B. Holland^{4,6}, Dahlia Nielsen⁷, Stephen Kresovich^{1,8}, and Edward S. Buckler^{1,4,8})

¹IGD, Cornell University, Ithaca, NY, USA

²Department of Genetics, University of Wisconsin, Madison, WI, USA

³Dep't of Agronomy, University of Missouri, Columbia, MO, USA

⁴ARS-USDA, Beltsville, MD, USA

⁵Dep't of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

⁶Dep't of Crop Science, North Carolina State University, Raleigh, NC, USA

⁷Dep't of Statistics, North Carolina State University

⁸Dep't of Plant Breeding and Genetics, Cornell University

This presentation was published as 'Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38, 203 – 208 (2005).'

Because population structure can result in spurious associations, it has constrained the use of association studies in human and plant genetics. Even so, association mapping holds great promise if true signals of functional association can be separated from the vast number of false signals generated by population structure.

We have developed a unified mixed-model approach to simultaneously account for multiple levels of relatedness detected by random genetic markers. We applied this new approach to two samples: a family-based sample of 14 human families for quantitative gene expression dissection, and a sample of 277 diverse inbred maize lines with complex familial relationships and a population structure for quantitative trait dissection.

Our method showed improved control over both Type I and Type II error rates, compared with other methods. As this new method crosses the boundary between family-based and mixed association samples, it provides a powerful complement to current methods of association mapping.

Characterizing the genetic diversity of maize populations: documenting the global migration of maize from its centre of origin

Susanne Dreisigacker¹ (with the collaboration of Marilyn Warburton¹, L. George¹, S. Taba¹, Visvanathan Mahalakshmi², A. Menkir², Alain Charcosset³, Z. Muthamia⁴, S.H. Zhang⁵, B.M. Prasanna⁶, Sutrisno⁷, N. Sawon⁸, Pichet Grudloyma⁸, P.X. Hao⁹, and A. Salazar¹⁰)

¹CIMMYT, Mexico City, Mexico

²IITA, Ibadan, Nigeria

³INRA, Montpellier, France

⁴KARI, Nairobi, Kenya

⁵CAAS, Beijing, China

⁶IARI, New Delhi, India

⁷Ministry of Agriculture, Jakarta, Indonesia

⁸FCRI, Bangkok, Thailand

⁹National Maize Research Institute, Dan Phuong, Ha Tay, Vietnam

¹⁰Department of Agriculture, Quezon City, Philippines

The project aims to characterize the structural and functional diversity of maize populations from Asia and Africa and compare it with a previous study on diversity in Europe and the Americas. The organizational meeting for the partners in this project was held on 4–6 April 2005 in Nairobi, Kenya, at which the following was accomplished:

- Distributed protocols for SSR amplification and electrophoresis, and data analysis
- Sent contracts to all partners, and clarifications made.
- Discussed the list of accessions to be genotyped. Numbers were chosen for each country or region, and the institutes assigned for DNA extraction, as follows: For countries, India and China with 25 populations each; and Indonesia, Vietnam, Thailand, and the Philippines with 22 populations each. For African regions, IITA with 60 populations from western and central Africa; KARI, 60 populations from eastern and southern Africa. For CIMMYT, 15 populations of teosinte and 15 populations of controls
- Because not all the money budgeted for the first meeting was spent, each NARS was invited to submit a US\$1500 capacity-building grant. Two countries (China and Thailand) are sending people to CIMMYT for training, one (Kenya) will buy computers, but other countries are yet to submit grants
- Each institution created work plans, including detailed activities and schedules
- A virtual work space was also created, and a CD of relevant materials distributed to workshop participants. The CD included germplasm data collection sheets, reporting procedures, and expectations for this project; laboratory protocols; data analysis programs; database programs; and the project's full proposal.

Each laboratory is now growing plants and extracting DNA, except for Vietnam and Kenya, who must first collect germplasm from farmers' fields. Their DNA will be available by the end of the year.

The second workshop for this project has been scheduled for 5–9 December 2005 in Beijing, China. At this workshop, participants will give updates on progress and problems, exchange DNA if this has not yet been done, and learn about the data analysis programs that will be used to analyze the bulked maize populations for this project.

Session IV: Use of molecular data for subsampling

Developing decision-support systems for sampling germplasm

Marco Bink¹, and Xavier Perrier²

¹*Plant Research International b.v., Wageningen University and Research Centre, Netherlands*

²*CIRAD, Montpellier, France*

The project aims to identify and quantify the relationships between markers and genes underlying agronomic traits, using QTL detection and association mapping. Activities include (1) maximizing allelic diversity, using large germplasm collections, and (2) to reduce costs and delays, concentrating phenotyping efforts on an 'optimal' subsample.

Strategies for optimal subsampling in a structured population include (1) representative subsampling, and (2) subsampling to reduce linkage disequilibrium related to known or unknown population structure, that is, to reduce 'structural disequilibrium' or SD. These strategies are expected to result in user-friendly software that will be made available to the GCP community.

Currently, the project is extending the Dissimilarity Analysis and Representation for Windows (DARwin) software so it will meet the project's goals.

Comments were received from the workshop in terms of how different sources of information can be valued to calculate genetic distances (and the resulting trees), and how confident users can be with the resulting dissimilarity trees. Statistical algorithms are being implemented to allow multiple sources and different types of information to calculate genetic differences. The DARwin software includes bootstrapping procedures to calculate confidence on inferred dissimilarity trees.

Controlling population structure in association mapping

Gael Pressoir

IGD, Cornell University, Ithaca, NY, USA

Current association mapping methodologies do not account effectively and simultaneously for complex pedigrees, families, founding effects, and population structure. As population structure can result in spurious associations, it has constrained the use of association studies. Association mapping, however, holds great promise if true signals of functional association can be separated from the vast number of false signals generated by population structure. In this introductory talk, we review how population and individual plant history may create significant linkage disequilibrium between a trait of interest and a locus unlinked to any of the trait's QTLs.

Subsampling among 700 sorghum accessions

Claire Billot

CIRAD, Montpellier, France

This presentation illustrates that of Marco Bink's on sorghum, which discussed the two strategies of subsampling: maximizing the length of the branch (which could be compared to maximizing genetic diversity) and minimizing the linkage disequilibrium related to structure (structural disequilibrium or SD).

To test SD, the high number of alleles provided by SSR markers must be reduced to help eliminate spurious linkage disequilibrium related to small numbers of comparisons. The first approach explored was based on grouping the alleles around the allelic distribution modes. However, the results of comparing the two trees obtained were not satisfactory. Overall SD was high (0.68). After sampling to minimize SD in the homogeneous group of Chinese sorghum accessions, the branch length was maximized, which gave the group a star-like organization. Depending on the studies, one or the other strategy can be used. We still need to choose a method for grouping the alleles.

Allele binning

Dave Hoisington

ICRISAT, Patancheru, Andhra Pradesh, India

The various genotyping projects are producing massive amounts of allelic data. Through the use of Genescan/Genotyper/Genemapper software, it is possible to perform semi-automated sizing of alleles; however, results so far have indicated a large number of alleles if only rounding-off to the nearest integer is used to determine the "called allele size" from the raw ABI output. We have investigated the use of "binning" to better predict the true allele size and to provide a quality estimate of the ABI dataset.

While several methods are available, we have implemented the algorithm in the paper by Idury and Cardon (Genome Research 7:1104-1109) as a C program (AlleloBin). The program analyzes how well the raw data fits the expected allele size based on the repeat unit of the microsatellite, as well as possible allelic drift. The output of the program is a frequency table of the various allele classes, ranges, etc. and the normalized standard deviation as an indicator of the accuracy of binning. The program will also produce a new allele dataset with the "called allele sizes" indicated for each entry.

Whether to use the rounded or called allele sizes depends on what the user knows about the data in terms of quality, etc. The current GCP format allows the user to provide both the original, un-binned allele size and, if it was binned, the called allele size plus what method was used.

Please contact me if interested in obtaining a copy of the program.

Session V: Guests' research activities: NARS

Coconut germplasm conservation and research with molecular markers at the Coconut Research Institute of Sri Lanka

Lalith Perera

Genetics and Plant Breeding Division, CRISL, Lunuwila, Sri Lanka

The mandate of the CRISL's Genetics and Plant Breeding Division is to develop coconut cultivars and/or hybrids that are improved in terms of high yield, early flowering and drought, pest, and disease resistance, using both conventional and biotechnology applications. The Division's activities comprise several research programmes that involve collecting and conserving coconut germplasm, developing and applying molecular markers to coconut, characterizing coconut germplasm through both phenotypic and molecular markers, and developing segregating populations for genome mapping towards marker assisted selection (MSA).

The locally available gene pool has already been collected and conserved in *ex situ* gene banks. Three varieties are recognized, i.e Typical (tall coconuts), Nana (dwarf coconuts), and Aurantiaca (intermediate between tall and dwarf). Among them, 15 forms can be determined, based on pollination behaviour and morphological features. In addition to these varieties, several other distinct coconut forms were identified in 2004 in the Southern coastal area.

Recently, coconut was extensively subjected to genetic erosion on the island. Systematic collection and conservation of germplasm was initiated in 1984 and is still continuing. To date, about 100 accessions have been collected and conserved in *ex situ* gene banks. The collections are undergoing in-depth morphological and molecular characterization.

The CRISL is conducting several research activities in collaboration with various partners, for example:

- With SCRI, Scotland, developing 39 SSR primers for coconut, 20 of which have been optimized and are being used.
- Developing DArT markers for coconut, an activity supported by the GCP.
- Developing markers for variety identification and hybrid testing.
- Verifying, through molecular markers, the use of petiole colour as a phenotypic marker for culling illegitimates in hybrid nurseries. This activity is supported by IFS, Sweden.
- Completing pollination between Sri Lanka red dwarf and Sri Lanka tall to develop a mapping population. The nuts are now being harvested. For this mapping population, progeny is expected to number about 250 individuals, which will soon be genotyped and scored for early vegetative growth characters and physiological parameters.

University of Southern Mindanao: research activities

Emma K. Sales

University of Southern Mindanao, Kabacan, Cotabato, Philippines

The University of Southern Mindanao Agricultural Research Center is also the National Research Center for maize, sorghum, fruits, and plantation crops (rubber, cacao, oil palm, and coffee). It carries out the following biotechnology projects:

- *Project 1:* Marker-assisted selection for high-quality maize improvement in the Philippines:
 - a. 153 inbred USM maize lines were fingerprinted and analyzed for diversity, using 40 SSR primers.
 - b. Polymorphism survey of 19 QPM lines, using 37 SSR markers.
 - c. Field and molecular evaluation of inbred maize lines; 54 yellow QPM, 17 white QPM, and 33 white QPM were single-crossed.
 - d. Marker-assisted selection of QPM materials is continuing: 11 white BC₁F₂ populations were generated, while a BC₁F₁ population is being generated from 11 yellow BC₀F₂ populations.
 - e. Four yellow hybrids were developed and submitted as test entries to National Corn Testing (NCT) for a wet-season trial.
 - f. Four promising yellow hybrids are being developed for submission to the NCT for a dry-season trial.
- *Project 2:* Improving the productivity of banana farms in the Philippines by providing a technical support system (banana indexing) and continuously analyzing for the BBMV, BBTV, and BMV viruses in tissue culture-derived planting materials.
- *Project 3:* Using molecular marker techniques to select induced mutation of bananas with improved post-harvest qualities, for which the following has been done:
 - a. Germplasm collection and field establishment of 25 *Musa* cultivars (AA, BB/BBB, and AAB).
 - b. Assessing and documenting existing post-harvest handling of *Musa*.
 - c. Benchmark survey of post-harvest qualities and maturity indices.
 - d. Preliminary evaluation of some post-harvest qualities of four popular *Musa* cultivars ('Lakatan', 'Latundan', 'Cardaba', and 'Señorita').
 - e. Confirmation of ploidy level, using cytometry; and of genome constitution, using SSR marker analysis.
 - f. Radio-sensitivity tests for mutation induction (at 10, 20, 30, 40, 50, and 60 Gy).
- *Project 4:* Marker-assisted hybridization and selection for high timber latex yield in rubber. Activities were:
 - a. A total of 33 rubber clones were morphologically evaluated.
 - b. The protocols for RAPD and SSR analyses of rubber clones were optimized.

AVRDC: The World Vegetable Center

Liwayway M. Engle

Genetic Resources and Seed Unit, AVRDC, Shanhua, Tainan, Taiwan

The AVRDC houses more than 53,000 accessions of numerous vegetables and their wild relatives, making it one of the biggest of such collections in the world. The collection includes *Glycine* (e.g. soybean), *Capsicum* (e.g. bell pepper), *Lycopersicon* (e.g. tomato), *Solanum* (e.g. eggplant), *Brassica* (e.g. Chinese cabbage), the bulb *Allium* (e.g. onion), *Vigna* (including *Vigna radiata* (L.) R. Wilczek or mung bean) and various other beans and legumes, and cucurbits. In all, 153 genera, involving 434 species, from 151 countries are held at the AVRDC.

The Center's Genetic Resources and Seed Unit serves as an *ex situ* conservation centre for these vegetables, many of which are globally, regionally, and locally important. Species indigenous to South and South-East Asia, and Africa are also part of the collection.

Accessions are characterized morphologically, using standard sets of descriptors. Evaluation for quality traits and resistance to pests and diseases are done by units specializing on the trait (e.g. nutrition laboratory, virology, or entomology).

The Unit also conducts research on vegetable genetic resources, and their conservation and use. Themes include germination; drying; diversity analysis, using cluster analysis based on RAPD, AFLP, and seed protein data; heterogeneity within accessions; building core collections of different crops, including for *Capsicum*, *Lycopersicon*, *Solanum*, and *Vigna*; and genotyping prioritized crops from these genera.

To deal with heterogeneity within accessions, the Unit first determines whether variation results from mixtures of distinct populations or from segregation of heterozygous materials. Mixtures are separated into sub-accessions, whereas segregating materials are bulked and designated as such.

These research activities are carried out in collaboration with national agricultural research systems (NARS) and other institutions such as the Japan International Research Center for Agricultural Sciences (JIRCAS) and the National Plant Genetic Resources Center (NPGRC) of Taiwan, and with other AVRDC units such as the Molecular Biology Unit.

AVRDC is a member of CGIAR's System-wide Information Network for Plant Genetic Resources (SINGER), and is currently contemplating entering the GCP.

Genetic enhancement for the coastal ecosystem: research at the MSSRF

Ajay Parida

MSSRF, Chennai, India

Salinity is a significant limiting factor to agricultural productivity, affecting about 900 million hectares worldwide. About one third of all irrigated land is affected by salt from secondary salinization and, by 2050, half of all arable lands are expected to be salinized. The problem of salinity is most acute in coastal regions. Improving or maintaining yield potential of crops under increased salinization will be more significant in the future.

To identify and isolate novel genetic combinations offering resistance to coastal salinity, the MSSRF has initiated work on mangrove species. Mangroves are salt-tolerant plant communities that occupy the coastal estuarine regions of the tropics. They serve as a vital link between terrestrial and aquatic ecosystems and provide livelihoods and ecological security for coastal communities.

The MSSRF is the first institution to have undertaken modern molecular marker-based analysis of mangroves. These studies have provided substantial information for developing unambiguous identification systems for individual species, elucidating the nature and extent of genetic diversity at intra- and interpopulation levels, depicting interspecific relationships, and establishing phylogenetic trends in mangroves.

In addition, studies on the genetic indexing of wild relatives and landraces of cultivated legumes, millets, and rice have contributed immensely to the understanding of the genetic structures of these species and/or populations. These studies have been able to decipher the diversity at population, species, and genus levels.

Novel genetic combinations with implications for abiotic stress were identified and isolated, using the widely distributed mangrove species, *Avicennia marina* (Forsk.) Vierh., and the wild rice *Porteresia coarctata* (Roxb.) Tateoka. Enriched gene libraries constructed from these two species are used to identify and isolate stress-tolerant genes. Many novel genetic combinations have been identified, sequenced, and characterized from these libraries. Efforts to identify unique genes in mangroves have also been undertaken, using large-scale genome sequencing and differential expression analysis. Some of the isolated genes from mangroves were characterized and analyzed for their expression levels under varying saline conditions.

Methodologies for constructing vectors for transformation and transformation systems have been established for tobacco, rice, black gram, and mustard. An integrated approach, from gene isolation to development of transgenics in locally adapted cultivars, and the integration of prebreeding with participatory breeding is the focus of ongoing work, which underlines the basic principle of ensuring diversity with efficiency. This will help strengthen the stability and sustainability of the fragile coastal ecosystem, as well as the productivity and profitability of coastal farming systems.

Developing and using molecular markers at the Plant Biotechnology Centre, IARI

T. Mohapatra

National Research Centre on Plant Biotechnology, IARI, New Delhi, India

Much of the following text has been published elsewhere, for example, at www.nrcpb.org/markers.html and in several refereed international journals.

At the National Research Centre on Plant Biotechnology, we have been using RFLP, STMS, RAPD, ISSR, and AFLP molecular markers for various purposes, for example:

1. *Genome mapping and gene tagging in mustard.* The mustard genome map was constructed and used to tag a gene, designated as *Ac2(t)*, that conferred resistance to the white rust pathogen, *Albugo candida* (Pers.) Kuntze. The same linkage map was also used to map, at marker intervals, common, linked, and independent loci that influence the levels of oleic, linoleic, linolenic, and erucic acids.
2. *Physical mapping in rice.* India, as a partner in the International Rice Genome Programme and with the collaboration of Clemson University Genomics Institute, USA, first constructed a BAC-based physical map using the mapped markers as probes and then sequenced about 57 cM of the long arm of chromosome 11 of rice.
3. *Marker-assisted selection in rice.* Bacterial blight (BB) caused by *Xanthomonas oryzae* pv. *oryzae* (Ishiyama) Swings et al. is a major disease of rice in tropical Asia. Because all Basmati rice varieties are highly susceptible and the disease is prevalent in the entire Basmati-growing region of India, BB is a severe constraint. In a collaborative study with the Division of Genetics, IARI, two BB resistance genes *Xa13* and *Xa21* present in IRBB55 were combined with the Basmati quality traits of Pusa Basmati-1 (PB-1). Molecular marker-assisted selection for the two resistance genes was carried out in BC₁F₁, BC₁F₂, and BC₁F₃ generations. The presence of the two genes was confirmed in all the desirable BC₁F₅ recombinants. Enhanced level of resistance in the recombinants was confirmed by disease reaction after inoculation. One promising line that outyielded the national check variety is now in the advanced stage of testing in the All India Co-ordinated Trials.
4. DNA fingerprinting for germplasm characterization and variety identification.
 - i. DNA fingerprinting technology was developed for rice, wheat, and *Brassica*. Markers such as RFLP, RAPD, STMS, ISSR, ASSR, and AFLP were standardized for variety identification in rice (including the Basmati varieties), wheat, and mustard. DNA fingerprints were created for 120 varieties, including the commercially important, traditional, Basmati rice cultivars, and 137 wheat and 42 *Brassica* cultivars, thus providing molecular descriptors for variety identification.
 - ii. Molecular marker analysis of species clones and commercial varieties in sugarcane. AFLP markers were employed to establish species relationships in commercial sugar-cane varieties, and to understand their genomic constitution, together with the contribution of the progenitor species *Saccharum officinarum* L. and *S. spontaneum* L. A large number of species-specific AFLP markers were identified, which could be used for identifying varieties and interspecific hybrids, in gene mapping, and in introgressing useful genes from different species. Cross transferability of maize microsatellite markers to sugar-cane species and varieties was demonstrated.
 - iii. Evaluation of genetic diversity and variety identification in jute. Jute (*Corchorus* spp.) is an important fibre crop that has ruled the packaging sector for over 150 years in India. To understand its pattern of diversity, 20 exotic germplasm lines and 20 commercial varieties of the two cultivated species (*C. olitorius* L. and *C. capsularis* L.), and two wild relatives (*C. aestuans* L. and *C. trilocularis* L.) were characterized, using STMS, ISSR, and RAPD markers. The level of polymorphism

was found to be significantly low within species. The commercial varieties, particularly those of *C. capsularis*, had an extremely narrow genetic base.

- iv. AFLP-based genetic diversity analysis and variety identification in ber (*Ziziphus* spp.). Ber is an important fruit crop that grows naturally under harsh environmental conditions. Genetic diversity among 41 accessions of cultivated *Z. mauritiana* (Lam.) and its wild relative *Z. nummularia* (Burm. f.) Wight & Arn. was determined, using AFLPs. The wild relative, which is being used as rootstock, showed greater diversity than the cultivated species, the commercial cultivars of which had a broad genetic base.
- v. Differentiation of sex in kiwi fruit, a dioecious fruit plant that has become commercially important in the low and mid-Himalayan regions of India. In collaboration with the University of Horticulture and Forestry, Solan, India, we used RAPD profiling to characterize male and female plants of this fruit crop. We identified diagnostic markers that could efficiently differentiate male from female genotypes and thus could be used as markers to identify the sex of planting materials.

Using AFLP markers and passport data when sampling a core collection and estimating genetic diversity

Kioumars Ghamkhar¹, and R. Snowball^{1,2}

¹CLIMA, University of Western Australia, Crawley, WA

²Department of Agriculture Western Australia, South Perth, WA

Clovers are widely used as annual pasture legumes. One that is currently receiving considerable attention for southern Australian conditions is the bladder clover, *Trifolium spumosum* L., a species from the Mediterranean Region. Although molecular techniques such as AFLPs, microsatellites, DNA sequencing, and isozymes have been used to assess the genetic diversity found in other *Trifolium* species or within the genus generally, there is no record of using these techniques to develop core collections within the genus or evaluate and characterize the genetic diversity available within this particular species.

Our study aimed to develop an effective core collection of *T. spumosum* that would act as a model for other pasture legumes. We first selected 317 accessions, with near-complete ecogeographical data, from the Australian ex situ collection of *T. spumosum*. Missing data, including details on ecosystems and patterns of infraspecific diversity were completed. The accessions were then grouped into five geographical regions, which thus provided a basis for determining how to maximize the sampling of genetic diversity. A preliminary core of 30% of the collection was selected.

To obtain primers, fluorescent AFLP (fAFLPs) were used. We screened four of the most diverse accessions to obtain primers that produced the highest number of bands and polymorphism to further screen the preliminary core collection. From this screening we expect to develop the final core collection, which would contain 30% of the preliminary core.

To date, the first ecogeographical subset of 100 accessions has been selected and mapped on Arcview and was found to effectively represent the original collection. A final core of 32 accessions will be selected, using ecogeographical data and AFLP markers with different fluorescent labelled EcoR-I primers to show different levels of diversity. The genetic profiles of the preliminary core will be scored and recorded in a database with ecogeographical data.

We expect our study to demonstrate that a combination of AFLP markers and ecogeographical data can be used to develop an effective core collection that maintains most of the genetic diversity. Such a core will allow breeders to more effectively select cohorts for field testing and enable gene bank managers to more efficiently conserve germplasm. It will also help identify deficiencies in genetic diversity, thus resulting in a more effectively targeted collection. The quality of ecogeographical data of collection sites will be critical to the success of this approach to developing a core collection. This model can also be used to develop core collections for other pasture legume species and even other crops with collections that are too large for efficient use.

Agriculture and Agri-Food Canada

Campbell G. Davidson

AAFC, Ottawa, Ontario, Canada

Canada is an important contributor in the preservation of genetic resources and the scientific application of genetics and plant breeding in a wide variety of crop species. The AAFC has expertise in the preservation of genetic resources, including gene bank management, crop variety development, seed storage, gene mapping and distribution. It fosters the scientific applications of genetics, including pest and disease management, molecular genetics, breeding, and improved germplasm. Development and marketing of nutritional and functional foods are expanding areas of activity. AAFC also has experience with intellectual property, patents, and technical and market assessments for commercialization. Its biotechnology expertise is often maximized by partnerships with universities, provinces, and the private sector.

Some research initiatives carried out by the AAFC are:

- The AAFC's Cereal Research Centre (CRC) is providing new genetic tools, information, and genetic resources to chemists, breeders, and pathologists in Canadian cereal breeding programmes. For example, by using high-throughput genomics technology, the initiative identifies genes involved in disease resistance and seed development, and thus genes that would provide durable broad-spectrum disease resistance and quality traits tailored for particular end uses.
- The CRC is also collaborating with Canadian universities to develop genetic maps for wheat and barley. The maps are anchored with microsatellite markers and also include AFLP markers for higher marker density in specific regions. The primary traits being studied and mapped include (1) for wheat, genes for resistance to fusarium head blight and leaf rust, component traits of bread-making quality, yield, and kernel colour; (2) for barley, resistance to scald and stem rust, and malting quality; (3) for oat, resistance to crown rust; and (4) for peas, resistance to lodging.
- Identifying Canadian wheat varieties, using SNP chip technology. DNA microarrays with spotted oligonucleotides are developed to genotype Canadian cultivars. These oligonucleotides are SNP markers tested among cultivars by analyzing the 3'-end untranslated segments of more than 1000 wheat genes from 60 cultivars. We plan to explore diagnostic platforms that demonstrate high throughput, quantitatively detect contaminating seeds, and distinguish hundreds of cultivars in a single test.
- The AAFC's Eastern Cereal and Oilseed Research Centre (ECORC) is involved in several research activities, including identifying barley varieties, using RAPD, RFLP, and SSR markers, and developing identification keys; molecular systematics in barley species, including *Hordeum spontaneum* K. Koch, using RAPDs; and QTL analyses for traits in barley such as good adaptation, disease resistance, other agronomic traits, physiological traits, malting quality, and feed quality.
- Activities for identifying potato cultivars, particularly Canada's heritage potato varieties, some of which have long been suspected as being the same clone or genotype but with different names. Protocols for distinguishing genetically close varieties have been established, based on microsatellite DNA regions of the potato genome.

The AAFC is also involved in international inter-institutional activities that help link national programmes to global challenge efforts, exchange information and ideas, and develop collaborative activities of common concern. It also focuses on increasing the value of both national and international collections to the plant breeding community, collaborating on intellectual property rights issues, and participating in other international initiatives such

as the FAO Commission on Genetic Resources, Global Crop Diversity Trust as it relates to the development of crop and regional PGR strategies.

Molecular marker research on wheat germplasm at the Ch. Charan Singh University

H.S. Balyan and P.K. Gupta

Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut, India

For the past 10 years, we have been developing and using molecular markers for wheat genetics and breeding. This work included the development and use of random genomic SSRs, EST-SSRs, SAMPLs, AFLPs, and SNPs.

The random genomic SSRs were used for (1) preparing genetic and physical maps, (2) gene tagging/QTL analysis for several traits, and (3) diversity estimates. The marker–trait associations detected included those of *wmc41* on chromosome 2D with grain protein content (GPC), *wmc104* on chromosome 6B and *MST101* on chromosomes 7D with preharvest sprouting tolerance (PHST), and *wmc333* on chromosome 1A with grain weight (GW).

The QTL analyses conducted involved the use of single-locus analyses through simple regression analysis (SMA); simple interval and composite interval mapping (SIM and CIM, respectively); and two-locus analyses for detecting epistatic QTLs (E-QTLs). Many markers associated with QTLs were detected that may prove useful in marker-assisted selection (MAS). These studies also suggested that most of the genetic variation for GPC was due to E-QTLs and $Q \times E$ interactions, while that for PHST was controlled mainly by main-effect QTLs and E-QTLs. For PHST, a major new QTL (*QPhs.ccsu-3A.1*), independent of red-grain colour and explaining up to 78.03% of phenotypic variation, was also detected on 3AL.

For diversity estimates, a variety of molecular markers (SSR, EST-SSR, AFLP, and SAMPL) were used on a collection of >50 exotic bread-wheat genotypes. These studies suggested that the data from more than one type of random genomic markers and those from the expressed portion of the genome should be used for obtaining more reliable diversity estimates. SNP and haplotype analyses have also been conducted, using EST-contigs in bread wheat.

Recently, we also initiated programmes on (1) Eco-tilling for mining alleles for traits of adaptive value (including drought and heat tolerance) and domestication-related traits, (2) high resolution mapping of genomic regions containing important QTLs for GPC and PHST that we detected in bread wheat, and (3) introgression of important QTLs for GPC, PHST, and GW into elite Indian bread-wheat genotypes through marker-assisted backcrossing.

Marker-assisted selection for drought tolerance in rice

H.E. Shashidhar

University of Agricultural Sciences–Bangalore, India

My project is to develop drought-resistant rice (*Oryza sativa* L.) that can maintain yield levels while requiring less labour and less water. My strategy is to select for the deep-rooting habit among segregants from a hybridization programme, using a phenotyping methodology that simultaneously screens several hundreds of lines in replicated trials. Two PCR-based markers were developed and are now being used, together with 23 other markers chosen according to QTL mapping studies conducted in the University's laboratory and elsewhere.

With these two methodologies, I obtained stable F8 lines that perform very well when grown under aerobic conditions, that is, when directly sown into non-puddled soils. Because puddling and levelling become unnecessary, huge savings of water and labour are made, to as much as 50%—a significant amount, as about 2000 litres of water are needed to produce one kilogram of rice under puddling conditions. Labour and costs are further saved by sowing seeds directly into cropland, bypassing the need to establish seedling nurseries.

Crop establishment of the F8 segregating lines is excellent, with no symptoms of deficiencies or toxicities. Because planting densities are lower, crops are relatively healthier. Irrigation is by furrow, as with maize or sorghum, and can be limited to once every 5 or 6 days. Yield averages 4 to 5 tons per hectare. The lack of standing water eliminates the incidence of pests such as planthoppers (both brown and white). Some lines are attacked by blast and root grubs, which are both easily controlled.

The direct planting technology and the F8 aerobic rice lines are now being evaluated on farm with farmer participation. Field days are organized as each crop matures. Farmers provide feedback and select lines. These lines are then evaluated, together with those selected by breeders. At present, three lines are undergoing advanced trials in several sites across the State of Karnataka and the nation.

Current status and recent progress of conservation and use of PGR in the Republic of Korea

Ma kyung-ho

Korean National Institute of Agricultural Biotechnology, Suwon, Republic of Korea

The Korean Rural Development Administration (RDA) plays a key role in planning and supporting the national programme to conserve and use crop genetic resources. The RDA Gene Bank (RDAGB) was established in 1988 and organized as a division of genetic resources in the National Institute of Agricultural Biotechnology of the RDA. The Gene Bank undertakes its own research and collaborates with other institutions on germplasm collection, characterization, evaluation, conservation, documentation, and use of PGR, including for crop improvement for food and agriculture. It also aims to achieve the three objectives of the Convention on Biological Diversity (CBD), pursuing the proposed activities for implementing the Global Plan of Action adopted by the International Technical Conference on Plant Genetic Resources in Leipzig, 1996.

Efforts are being made to develop additional databases for germplasm, including an image database, DNA profiles, and molecular characterization. The RDAGB has begun conserving mutant and genetic stocks for genomic and functional research. To upgrade characterization and evaluation PGR data, genetic and biochemical tools, including gene expression displays such as DNA microarrays, are being used.

Among its activities, the RDAGB surveys the *in situ* distribution of wild relatives of crop species and assesses the diversity of weedy types found in farm fields, for example:

- Wild adzuki bean (*Vigna angularis* var. *nipponesis* (Ohwi) Ohwi & H. Ohasi), which is a progenitor of the cultivated type (*V. angularis* var. *angularis* (Willd.) Ohwi & H. Ohasi) and *V. nakashimae* (Ohwi) Ohwi & H. Ohasi, another wild relative. Seed samples of 165 accessions were characterized in 2002.
- Wild soybean. About 1000 accessions, collected during 1991–1995, are preserved, and another survey implemented to manage the wild species *in situ*.
- Weedy rice. A survey across five different river regions was conducted in 2002, resulting in 3599 accessions being sampled from 9 provinces. This collection complements the weedy collection held by the Yeungnam University. These accessions will be re-evaluated by comparative analysis based on phenotype and DNA analysis during 2004–2005. Incidence of their duplication and on-farm conservation status will also be studied.

Crop landraces are disappearing very fast from Korean farms. Compared with 1985, only 26% of landraces could be found in 1993, with losses ranging from almost 84% to 97%, depending on the region. Major reasons for such losses are their rapid replacement by a much narrower spectrum of improved, modern cultivars, simplified cropping systems, migration of young people to the city, and mechanization of the agricultural system.

Ex situ collections of wild relatives of crops include wild soybean (950 accessions), medicinal plants (714 accessions), *Allium* species (1466 accessions), wild *Raphanus* (1304 accessions), wild *Vigna* (76 accessions of wild adzuki bean, 88 *V. nakashimae*, and 1 weedy type of *V. angularis*), and weedy rice from farm fields (3599 accessions).

The RDAGB also has a seed storage facility that conserves 113,702 accessions of cereal crops, 18,273 of industrial and medicinal plants, 13,820 of vegetables and fruit trees, and 3947 of forage crops and other germplasm, collected within the country and abroad. It also holds, for vegetatively propagated germplasm, 45 field gene banks, maintaining 21,170 accessions of fruit trees, flowers, vegetables, medicinal plants, tuber crops, shrubs, grasses, and other germplasm. *In vitro* conservation for recalcitrant seeds and vegetatively propagated PGRs is being developed and adopted at the RDAGB, together with cryopreservation (using liquid nitrogen) for potato, tea, and garlic. Lastly, the RDAGB provides long-term conservation of

mutants (about 10,000 accessions), while the active collections are operated by each university or institute that had developed mutant populations.

So scientists can characterize, evaluate, and use germplasm, the RDA had developed descriptors by partially modifying the IBPGR/IPGRI descriptions. So far, about 72% of the stored germplasm has been for various traits including physiological responses, disease and pest susceptibility, yield productivity, and other agronomic features.

In 1991, the RDAGB was designated as the centre for world sesame germplasm collection by FAO and IPGRI and, since then, 1600 accessions have been characterized and evaluated.

The RDAGB is developing an integrated information system on the conserved germplasm. It is also active in capacity building, and recognizes its "international obligations in sharing benefits arising from the use of introduced PGRs" and the policies established by the CBD and the International Treaty on Plant Genetic Resources for Food and Agriculture. The RDAGB is currently working to become a member of the Treaty.

Genetic improvement of rice in Tamil Nadu, India*Sabariappan Robin, S. Rajeswari, K. Mohana Sundaram, and T.S. Raveendran**Centre for Plant Breeding & Genetics, TNAU, Coimbatore, India*

In post-Independence India, rice production steadily increased, keeping up with population growth, thanks to the proactive dominant roles played by agricultural scientists in synergy with policy makers. However, the growth rate of rice production, which had reached more than 4% in the 1980s, dropped to less than 2% in the 1990s and was feared to further decline to the point of importing rice to feed the Indian population.

In contrast, in Tamil Nadu, a strong crop improvement programme, supported by advances in crop management and protection practices, led to a quantum jump in rice production such that self-sufficiency was achieved. Rice has been bred in Tamil Nadu for more than 100 years. Crop breeding first exploited the natural genetic variations through systematically planned collections, evaluation, and selection exercises. Rediscovery of Mendel's theories led to the adoption of pure-line selection (PLS) based on Johansson and mass selection and gave rise to a series of rice varieties being released.

The first official rice variety, GEB 24, was a pure-line selection from a local variety, Konamani (GEB 24 was later found to be a deviant of Konamani, possibly a spontaneous mutant). Other successful rice varieties developed through PLS included CO 13, CO 18, CO 19, ADT 3, ADT 10, ADT 23, ASD 1, ASD 5, and TKM 1, which were popular for their high yields. PLS has helped purify and conserve several traditional rice varieties.

The most effective breeding method for rice was recombination breeding, involving the selection of parents, hybridization, and selection of superior segregants from F₂ and subsequent generations. The first rice variety developed through this method was CO 14, popularly known as 'Perunthandu', and released in 1940. Other varieties, including ADT 20 (hybrid 'Kuruvai') were released by many regional centres affiliated with TNAU. The search for favourable parents led to attempts in hybridizing with wild rice, including a cross between *O. perennis* Moench and GEB 24, which gave rise to variety CO 31 ('Ottu Manavari'), released in 1963 for the upland ecosystem. TNAU's Aduthurai Station then released, in 1965, a much acclaimed indica × japonica cross derivative 'ADT 27' that formed the basis for the 'Green Revolution' in Tamil Nadu. Simultaneously, several superior rice varieties were also released through introduction from IRRI (Philippines), of which IR20, IR36, and IR50 are still preferred by farmers for their excellent yield potential.

Fixed genotypes (i.e. homozygous lines) were found to plateau in yield. Cross-pollination in rice was achieved through the use of male-sterile lines received from IRRI, and locally adapted hybrids were synthesized, with three hybrids—CORH 1 (1994), CORH 2 (1998), and ADTRH 1 (1999)—being so far released.

To date, TNAU has released 145 rice varieties or hybrids, which contributed to the state's tremendous increase in rice production, from 1,529,000 tonnes in the 1920s to 7,366,000 tonnes in 2000/01. The rice crop improved genetically in terms of its yield potential, fertilizer responsiveness, pest resistance, biotic stress tolerance, and also got upscaled in marketability, and, hence, profitability.

Using molecular marker technology for mapping quantitative trait loci, functional genomics, and genetic diversity analysis

M. Maheswaran

Department of Plant Molecular Biology and Biotechnology, Centre for Plant Molecular Biology, TNAU, Coimbatore, India

Our group works with rice to understand the phenotypic and genetic structures behind three important agronomic traits of this crop: resistance to planthoppers, nitrogen uptake and nitrogen-use efficiency (NUE), and grain quality parameters. We also aim to assess the genetic diversity of the rice germplasm available to us and of important medicinal plants, using various types of molecular markers. Results of our research include:

- Many major genes are now available to manage the brown planthopper (BPH) *Nilaparvata lugens* Stål, a major pest of rice, using host-plant resistance. We used a double-haploid (DH) population derived from the cross IR64 × Azucena to map the quantitative trait loci (QTLs) associated with resistance to BPH, seedling resistance, antibiosis, and tolerance. We then evaluated a set of 94 DH lines, using a series of phenotypic tests that examined seedling resistance and resistance mechanisms such as antixenosis, antibiosis, and tolerance. Four QTLs in chromosome 1 showed association with tolerance. Two QTLs in chromosome 7 showed association with seedling resistance and antibiosis. The phenotypic contribution of the QTLs ranged from 10.4 to 17.6%.
- Mapping genomic regions associated with nitrogen uptake in rice cultivars with improved NUE even under suboptimal levels of applied nitrogen. We used 190 F₂ individuals of the cross Basmati370 × ASD 16. A parental survey with 264 SSR primer pairs and 9 ISSR primers showed 52% polymorphism between the parents. The segregation patterns of 60 SSR markers were scored on F₂ population and a framework map constructed. The F₃ families derived from the 190 F₂ lines were evaluated in the field without added nitrogen. Nitrogen content in grain and straw, grain yield, and dry matter production were estimated and the nitrogen uptake and translocation efficiency were derived for parents and segregating progenies.

ASD 16 was more efficient in N uptake and use than was Basmati370. Wider variation with transgressive segregation for different traits was observed among the F_{2,3} families. By adopting single-marker analysis, SSR markers linked with each phenotypic trait were identified, indicating their potential usefulness in working with the complex network of genes associated with NUE in rice.

- To map QTLs associated with grain quality parameters in rice, we adopted a strategy of combining DNA pooling from selective segregants and genotyping. The number of individuals forming the bulk influenced the identification of putative marker(s) for each trait. The association of these putative markers was established by single-marker analysis (SMA), which revealed that SSR markers RM225 on chromosome 6 and RM247 on chromosome 12 showed significant association with grain breadth (GB) and cooked grain breadth (CGB), respectively.
- To assess the genetic purity of seed lots of four rice varieties from different State Seed Farms of Tamil Nadu, we used a set of randomly amplified polymorphic DNA (RAPD) and simple sequence repeat (SSR) markers, generated by using five random primers and four SSR primer pairs. To detect heterogeneity, we tried out a simplified procedure involving two steps: bulked seed sample analysis (BSSA) and individual seed sample analysis (ISSA). Marker analysis, using five random primers on 18 bulked seed samples of the four rice varieties, showed heterogeneity on comparing the marker profiles with those of DNA extracted from breeder seeds of the test varieties. The ISSA also confirmed heterogeneity within the seed lots. In contrast to the two-step DNA testing method, the

four SSR primer pairs did not show heterogeneity in the 18 bulks of four varieties analyzed.

- To determine genetic diversity in *Phyllanthus*, we used 16 RAPD and eight ISSR primers to survey a collection of 54 genotypes of *P. amarus* Sch. et Thonn. and three genotypes each of *P. debilis* Klein ex Willd. and *P. virgatus* Forst. f. from different districts of Tamil Nadu. The average polymorphism percentage of the RAPD and ISSR markers was, respectively, 68.18 and 69.67 across the 54 genotypes of *P. amarus* for the 16 and eight primers used. When the *P. debilis* and *P. virgatus* genotypes were included, the level of polymorphism percentage increased to 96.46 for the RAPD markers and 97.51 for the ISSR markers.

Cluster analysis of the 54 genotypes of *P. amarus* based on RAPD and ISSR markers did not establish a clear-cut pattern according to place of collection. However, the cluster analysis of the 60 genotypes of all three species revealed a clear-cut grouping of genotypes at the species level. Both RAPD and ISSR markers separated *P. virgatus* from the two other species.

Session VI: The Long-Generation Crops Project: population structure, phenotypic information, and association studies

Population structure, phenotypic information, and association studies in long-generation crops: A project overview

M. Carmen de Vicente

IPGRI–Americas, c/o CIAT, Cali, Colombia

This Project attempts to integrate association analyses (in a broad sense, the comparison between molecular and phenotypic diversity) into the course of ongoing characterization and breeding activities in those cases where accurate phenotypic data are available for materials amenable to linkage disequilibrium mapping.

This alternative approach to conventional linkage mapping could prove especially useful for those crops where refined materials for genetic analysis such as inbred lines, NILs, or sets of substitution lines cannot be produced and for those crops where generation time is long. The crops involved in the Project respond to these characteristics.

For each crop the expected outcomes are:

1. Undertake a description of the phenotypic data available, and filling gaps as necessary to prepare germplasm sets for association studies;
2. Gather a description of the germplasm evaluation (phenotyping) system that is existing or accessible;
3. Obtain an estimation of the structure and degree of LD in the materials whose phenotypic description will be used; and
4. Document the feasibility and exploitation plans in subsequent breeding activities.

These elements, after their assembly and analysis, will serve to diagnose the prospects of association studies using current populations in these crops. In the most favorable cases, whole-genome scanning could later be undertaken, for detecting local haplotypes related to superior traits, thus providing positional information and allele comparisons in the search for genes of interest.

Population structure, phenotypic information, and association studies for coconut in Vanuatu*Luc Baudouin and Patricia Lebrun**CIRAD, Montpellier, France*

Vanuatu is an archipelago of scattered islands in the South Pacific. Coconut is at the centre of the economic and cultural life of its remote islands, often offering the only monetary resource. The local coconut population, known as Vanuatu Tall or VTT, has a high degree of diversity and, because of a viral disease (coconut foliar decay) that destroys virtually all exotic varieties, forms the basis for domestic coconut breeding. The cost of seed production and transportation forms a major obstacle to the diffusion of genetic progress, which could be lessened by identifying QTL markers that could be selected for in populations growing in remote islands and demonstration plots.

The Vanuatu coconut population also seems to have resistance to the Ghanaian form of lethal yellowing. We collected 22 samples for DNA analysis from each of the first three generations of the VTT breeding population at the Vanuatu Agricultural Research and Technical Center (VARTC; Luganville), namely unselected plantation materials, first and second generations of mass selection. We selected 152 trees from the advanced material resulting from progeny testing. Individual observations will be performed on fruit and bunch return, fruit components, and stem and leaf measurements. Ease of removing albumen from the shell will also be tested. Finally, a training session at CIRAD is being programmed for a breeder from Vanuatu.

Yam in Vanuatu

Jean-Louis Noyer¹ and Vincent Lebot² (presented by Jean-Christophe Glaszmann³)

¹CIRAD–AMIS, Montpellier, France

²VARTC, Espiritu Santo Island, Vanuatu

³Leader of SP1, GCP; CIRAD, Montpellier, France

Phenotypic data were scored for 331 *Dioscorea alata* L. accessions from the Vanuatu collection since 2000 (SPYN project 1999–2003). From these 331 accessions, 117 were characterized through an AFLP analysis, from which 124 loci were retained from patterns of 5 primer pairs. Phenotypic and molecular data analyses were achieved at the beginning of 2005 and the collection re-organised essentially by the suppression of 110 duplicates. Two hundred and twenty-one remaining accessions were planted at the end of July 2005 to continue the phenotypic evaluation, together with new accessions from ongoing collections in the Vanuatu Islands.

Phenotypic characterization is carried out as follows: 26 descriptors are used to describe the Vanuatu collection, 9 for tubers, 14 for aerial traits, and 3 for resistance to pest and diseases. The available results confirm high phenotypic variability within the species *D. alata*. Correlations between these characters were calculated from the data obtained during 2000 to 2003 and will be reconsidered at the end of 2005 when additional data (2004 and 2005) will be included.

The molecular data obtained so far include 51 SSRs, which were screened in a small interspecific sample, with 16 being retained for their capacity to exhibit useful patterns for all cultivated species belonging to the *Dioscorea* genus. DNA sequences were available in the EMBL Gene Bank since July 2005. The 35 remaining SSRs are currently being screened for their ability to reveal useful patterns within *D. alata*.

Association analysis in the course of varietal improvement

*M. Bonierbale, Marc Ghislain, E. Mihovilovich, Jorge Nuñez, J. Landeo, and M. Gastelo
CIP, Lima, Peru*

At CIP, activities with potato include exploring association mapping as an efficient means of relating variation at marker loci with agronomic phenotypes in the course of breeding potato for resistance, quality, and nutritional traits. Tetraploid potato cultivars and breeding clones are not inbred, but originate from crossing heterozygous parents. From the resulting F1 progenies, superior recombinant genotypes are selected and fixed by vegetative propagation. Although potato is not managed as a 'long-cycle' crop, genetic features resulting from population admixture, vegetative propagation, limited meioses, and selection history may influence the success of association mapping in ways similar to those of perennial species. The objectives and main activities for 2005 to mid-2006 are:

- To assemble passport/pedigree and trait data for 2 populations (one a 2× germplasm collection, and one an advanced 4× breeders' population) selected for this purpose.
 - To complement SSR data on the potato composite genotype set (CGS) by genotyping additional individuals of the selected populations with 50 SSRs from throughout the genome, and sets of closely linked markers in two or more genomic regions.
 - To estimate genetic structure and the degree of linkage disequilibrium in at least the 2× population.
 - To assemble candidate genes/primers hypothesized to underlie variation for selected resistance, quality, and nutritional traits.
 - To conduct gap-filling phenotyping for trait assessment in the 2× and 4× populations.
- The plant materials to be used are:

1. For the 4× population, genotypes will be selected from Population B1. This population had been developed through a breeding scheme that upgraded levels of resistance to late blight found in native tetraploid *Andigena* germplasm through five cycles of recurrent selection, while seeking to maintain post harvest quality and improve earliness, yield, and tuber aspects. Being based on 57 modern pedigree-wise unrelated genotypes and having undergone simultaneous selection for several independent characters, Population B1 probably provides and maintains significant allele and haplotype diversity, compared with the more-bottlenecked collections of North American or European varieties. Population B1 is an excellent source of new varieties, and may yield heterotic combinations when crossed with *Tuberosum*.

Considering the lack of admixture or introgression from distinct gene pools into this population, structure is expected to be relatively homogeneous (non-stratified). This should minimize the identification of spurious associations between the markers and traits it varies for. Increased recombination through five sexual cycles may further lead to increased resolution on marker–trait associations, and thus opportunities for finer scale mapping with respect to primitive or even pedigree-bred populations of this vegetatively propagated crop. A history of selection for late blight resistance will make this population ideal for positional candidate gene association analysis if sufficient variability remains, as indicated by parameters calculated by classical genetic means. However, a lack of algorithms for assessing LD in tetraploids and added complications in haplotyping heterozygous individuals may complicate the approach. One hundred clones of Population B1 are being assembled for study.

2. The second population selected for this study comprises about 120 clonal accessions of group of Phureja germplasm (2×), with moderate variability for disease resistance, and outstanding quality and nutritional traits. Preliminary analysis of the CGS suggests some internal structure of this group and, although its origin is undocumented, it has probably

undergone fewer meioses than Population B1, being a collection of traditional landrace genotypes. However, it has not been subject to selection in a modern breeding programme and therefore may contain greater allele and haplotype diversity than the B1 population.

For phenotypic data, significant proportions of both populations have been evaluated for resistance to late blight in replicated trials over two or more seasons. Data on resistance to PVY and bacterial wilt are partially available and could be completed if early studies show promise for the approach. Morphological descriptors for tuber shape, eye depth, and pigmentation; tuber yield, dry matter, and chipping quality data and concentrations of reducing sugars, vitamin C, iron, zinc, and other micronutrients are partially available and can be completed in the study. Both populations also vary for carotenoid content, but additional resources would be required for precise determinations.

Population structure, phenotypic information, and association studies in long-generation crops: cassava update

Paula Hurtado, Janneth Patricia Gutiérrez, Isabel Moreno, Edgar Barrera, and Martin Fregene

CIAT, Cali, Colombia

Cassava is a tropical staple crop that grows well under marginal conditions, where poverty is severe, making it an important source of food security in developing countries. It is an outbreeding crop, vegetatively propagated, and affected by diseases that lower the yield of preferred varieties. Cassava improvement is field-based and can be delayed by problems related to the crop's heterozygosity and reproductive cycle. To ensure cost-effective breeding, we are using readily available molecular marker tools to determine traits of interest, using marker-assisted selection (MAS).

In the same way, the need for genetic markers can be accelerated, using approaches that do not demand the mapping of populations, for example, association mapping. The CIAT cassava team has compiled phenotypic data on 800 lines generated by the Centre's breeding programme. These materials have been evaluated over several years at different sites and in replicated trials for traits of agronomic importance such as fresh and dry root yield, root dry matter content, starch content, number of roots, and resistance to pests and diseases.

Of the 800 lines, 117 varieties (75 advanced breeding lines and 42 landraces/elite) were selected as a sample for the preliminary determination of cassava linkage disequilibrium (LD). The accessions included in the reference sample were evaluated with 36 SSR markers and DArTs to estimate genetic diversity (this activity was part of the first year activities in SP1 Cluster 2). To have a general idea of LD in cassava, LD was calculated, using the information generated by the 36 SSR markers (212 alleles).

Initial results revealed LD in alleles of SSR loci on different linkage groups, suggesting unrelated germplasm in the sample of 117 varieties and that a new subsampling must be done, based on genotypic and phenotypic factors, to compose a new sample. To better approach the structure and degree of LD, new SSR markers (every 100–500 kb) are being developed on a genome-wide basis, using BAC contigs currently being built across the cassava genome. The new markers will be used to do a genome-wide evaluation that will allow us to calculate LD and use the genotypic data generated, as well as the information compiled on phenotypic factors in the association study.

Assessing the genetic diversity of *Musa* in Congo Basin

FC Baurens¹, K Tomekpe², JL Noyer¹

¹*CIRAD-AMIS, Montpellier, France*

²*CARBAP, Douala, Cameroon*

Plantain (*Musa* subgroup AAB), a particular type of cooking banana, presents high phenotypic diversity in the Congo Basin, a region that is considered to be a secondary diversification zone. Using nine microsatellite loci and AFLP markers, we assessed the genetic diversity of a group of 30 plantain landraces that constitute a representative sample of phenotypic diversity. The results confirmed a very narrow genetic base for this cultivar group.

We then conducted methylation-sensitive amplified polymorphism (MSAP) analysis to survey the cytosine methylation status at CCGG sites. A higher degree of polymorphism was revealed, allowing classification of the samples into three clusters. Although MSAP seems a useful molecular tool for highlighting differences within the plantain subgroup, we observed no correlation between the phenotypic classification (based on the phenotype of the inflorescence) and methylation diversity.

The possible transmission of methylation in progenies of banana involving plantain parents is being studied. Agronomic and biochemical characterization is being conducted for diploid populations derived from crosses between plantain parents and the highly polymorphic diploid banana. Potentially polymorphic and well documented populations are being analyzed by molecular markers to examine association mapping in light of the latest compiled genetic map.

These research activities are carried out in partnership with CIRAD

Session VII: Perspectives

The need for allele mining: perspectives of the System-wide Genetic Resources Programme (SGRP)

Ruaraidh Sackville Hamilton¹, Moussa Sie², Jan Valkoun³, Hari Upadhyaya⁴, and Visvanathan Mahalakshmi⁵

¹IRRI, Manila, Philippines

²WARDA, Cotonou, Benin

³ICARDA, Aleppo, Syria

⁴ICRISAT, Patancheru, Andhra Pradesh, India

⁵IITA, Ibadan, Nigeria

The SGRP is a consortium of all CGIAR gene banks. These banks aim to (1) conserve diversity efficiently, including and especially rare alleles and genotypes; (2) ensure efficient use of collection by delivering appropriately selected subsets of germplasm to users, maximizing the chance of giving users the set of alleles or genotypes that they need; and (3) ensure that the entire collection remains available for use. The SGRP exists to ensure a consistent, integrated, system-wide approach to the efficient achievement of these goals.

At this meeting, the SGRP, represented by five CGIAR centres —ICARDA, ICRISAT, IITA, IRRI, and WARDA—considered the following main topics:

1. Data, especially issues on publicizing standardized data through SINGER.
2. Policy on conservation and use of germplasm in terms of, for example:
 - i. Germplasm exchange and the intellectual property aspects of genetic resources,
 - ii. The Convention on Biological Diversity,
 - iii. The International Treaty on Plant Genetic Resources for Food and Agriculture, and
 - iv. Protection against the unintentional presence of transgenes in gene banks.
3. Generic methodologies for improved efficiency of conservation and use of genetic resources, for example, GIS technologies and molecular technologies for characterizing genotypes.

In particular, molecular tools are seen as vital for improving the efficiency of conservation and use of genetic resources. Molecular characterization represents the biggest scientific advance in PGR concepts since the 'Green Revolution', and it is an opportunity to revolutionize the way we operate. For this reason, the SGRP strongly supports the GCP and, since the inception of the idea, was involved in planning the GCP (then GRCP or Genetic Resources Challenge Programme). We expect the GCP's results to enable a major improvement in the way CGIAR gene banks operate.

An issue of particular interest to CGIAR gene banks is, what about the rest of the CGIAR collections? SP1 has defined, and is now characterizing, diversity in representative composite collections of the major crops, combining germplasm from as many gene banks as possible. These composite collections contain a large percentage of crop gene pools, and the information will be used to understand the ecogeographic distribution of diversity in each crop. Refining the composite collections to smaller reference samples will further facilitate the discovery of new functional genes and, once the functional genes are identified, we can discover the range of alleles included in the composite collections. The value of this information is exceptionally great.

Yet, as for rice, only about 2% of the germplasm held in the gene bank at IRRI is included in the composite collection; and only 0.2% is in the reference collection. The collection at IRRI contains many accessions that are of interest to our users but are not included in the composite collection. We must not forget that breeding is a 'numbers game', and finding the

rare valuable alleles or genotypes can involve screening many more accessions than are included in the composite collections. We need to find a way forward from genotyping the composite collections to true 'allele mining'—efficiently screening other accessions outside the composite collections to discover new valuable alleles.

Observers' viewpoint

Campbell G. Davidson², François Balfourier¹, Ajay Parida³, Kioumars Ghamkar⁴, Andreas Graner⁵, and Gael Pressoir⁶

¹INRA, Clermont-Ferrand, France

²AAFC Ottawa, Ontario, Canada

³MSSRF, Chennai, India

⁴CLIMA, University of Western Australia, Crawley, WA

⁵IPK, Gatersleben, Germany

⁶IGD, Cornell University, Ithaca, NY, USA

Overview:

Subprogramme 1 is charged with the exploration of the genetic diversity of global germplasm collections of the 22 mandate crops of the CGIAR. This exploration helps identify novel and diverse variants of genes involved in complex stress tolerance traits, thus providing the base for the research activities of the other 4 subprograms

From the observers' viewpoint, the benefits obtained were many and varied and included an increased familiarity with the goals for tier 1 crops.

The workshop provided an excellent overview of the GCP, updates on individual GCP activities, and the coordination of activities between the GCP and national projects. Additionally there were many suggestions for further development of national research and development activities, along with identifying potential future partners.

Some of the more specific issues commented on by the Observers included:

- Examining the potential to link with existing national programs and perhaps gaining access to matching funds. These linkages could expand the impact and potentially help to solve some of the critical challenges the GCP is facing. Early development of technology transfer will be essential to maintain the GCP program momentum.
- The GCP is developing valuable protocols to ensure access to and benefit sharing for the products that are derived from germplasm accessions. Gaining a solid understanding of the role of intellectual property rights (IPRs) and how these may affect various programs (e.g. freedom to operate or research exemptions). IPRs have the potential to assist and hinder the attainment of the goals and steps should be undertaken to further clarify important steps required.
- Identifying a clear route for the next steps and challenges, focusing on those that are "Mission Critical" in creating value for the plant breeding community. The linkage back to the overall GCP objectives and seeing how the various parts and components link together is important as well as keeping connected to the progress made in the other projects.

The observers also express their appreciation of the participants' excellent project presentations. They also noted and welcomed the participants' vision and direction, their spirit of collaboration, and the clear identification of future benefits, both personal and professional. The presentations and research undertaken demonstrated a consistent system-wide approach, where linkages and cooperation are being nurtured, results shared, materials made available, and forward-looking strategies developed.

NARS' perspectives

Emma K. Sales¹, Lalith Perera², Liwayway M. Engle³, T. Mohapatra⁴, H.S. Balyan⁵, H.E. Sashidhar⁶, M. Maheswaran⁷, Robin Sabariappan⁷, Kyung-Ho Ma⁸, and Kioumars Ghamkhar⁹

¹*University of Southern Mindanao, Kabacan, Cotabato, Philippines*

²*Genetics and Plant Breeding Division, CRISL, Lunuwila, Sri Lanka*

³*Genetic Resources and Seed Unit, AVRDC, Shanhua, Tainan, Taiwan*

⁴*National Research Centre on Plant Biotechnology, IARI, New Delhi, India*

⁵*Dep't of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut, India*

⁶*University of Agricultural Sciences–Bangalore, India*

⁷*TNAU, Coimbatore, India*

⁸*National Institute of Agricultural Biotechnology, Suwon, Korea*

⁹*CLIMA, University of Western Australia, Crawley, WA*

After discussion, the NARS representatives established the following viewpoints. Many NARS now maintain and conserve their own germplasm collections. Most, if not all, have acquired the capacity to use molecular tools to support plant breeding programmes. Diversity studies on different crops are conducted in each country. Where germplasm is endemic to given geographical areas, the GCP should establish linkages with institutions in these countries, ensuring that such linkages are not confined to a particular institution per country. Ideally, to complete the picture of global genetic diversity, the GCP should tap many different institutions in each country, making the effort to include as partners all institutions and individuals doing significant work on crops of priority to the GCP.

Apparently, within the GCP, data are not uniformly acquired across laboratories. Experience with AMBIONET suggests that establishing a format would be best, as data sets across laboratories can then be compared and combined. Phenotype and genotype data could be integrated 'on the go' and thus be more informative and alleles used as they are being discovered. For more meaningful data, the following is also suggested:

- Integration, visualization, and access within the GCP, through development and establishment of allele kits.
- Use of germplasm identifiers (GIDs) in all GCP work, as now is the best time to link the availability of materials and data with the role and potential of SP4 to implement the system. If a germplasm material represents a new genotype, it can be assigned a (-) GID and the curator of the database advised.
- The GCP should apply a 'Google' type search tool to make all databases accessible by keywords. Thus, a query would go across molecular, phenotypic, geographic, and topographic data, and visual tools become readily available to interested parties.
- Tier 2 crops should include fuel-yielding crops, considering the rate at which global fuel prices are increasing.
- Use heterologous markers to determine and/or establish considerable homology between crops (e.g. barley and rice, potato and tomato).
- Have a 'test user', meaning that one particular laboratory or group would initially validate and assess the usefulness of given techniques and the data generated by these techniques.
- Ideally, the numbers of markers should be based on genome size, or at least be proportional to it.
- The GCP should develop phenotypic platforms, that is, uniform phenotypic data that are specific, whether by crop, trait, or situation.

- Caution is needed in determining the constitution of composite and core collections or subsets, that is, too small is dangerous because of the possibility of losing useful and unique alleles. The ploidy level, agronomic uselessness, diverse phenology, etc., should also be determined. Hence, ploidy level based on cytological and flow cytometry techniques should be done first, that is, before or while embarking on serious molecular marker work (e.g. cassava and banana)
- And the most important—linkage and funding opportunities!
The NARS' wish list is as follows:
 1. For NARS to propose project proposals in line with the GCP's objectives. If this is not possible in the current phase, then perhaps this could be made so in the next phase?
 2. Training in procedures used by the GCP so to meet standards set by the GCP and ensure data are compatible with the GCP's data (including software use).
 3. Pre-doctoral fellowships for students.
 4. Help in acquiring equipment, supplies (e.g. markers), and software.

SP5's activities*M. Carmen de Vicente**Subprogramme 5 Leader*

IPGRI–Americas, c/o CIAT, Cali, Colombia

A deep technological divide exists between developed and developing countries in contemporary, cutting-edge knowledge on the effective use of genetic resources. To overcome this distressing reality, the GCP's Subprogramme 5 (SP5) focuses on capacity building in the belief that education and knowledge form the basis for development. Moreover, the GCP's ability to create impact in farmers' fields is directly associated with the NARS' capacity to use the GCP's technical outputs in breeding programmes that address farmers and consumers' needs. As such, linking with NARS is essential for developing an effective delivery plan for GCP products. From now on, the GCP will require each of its projects to develop a delivery plan that details which partners, particularly the NARS partners, they will work with and what their capacity building needs are to be able to apply the outputs of the GCP research. The idea is to encourage scientists to clearly explain how their results and products will help users and generate future products for farmers.

Appendix I. Acronyms and abbreviations used in the text

Acronyms of organizations

AAFC	Agriculture and Agri-Food Canada
ABI	Applied Biosystems, California, USA
Agropolis	Pôle international de recherche et d'enseignement supérieur agronomiques, France
AMBIONET	Asian Maize Biotechnology Network
AMIS	Amélioration des méthodes pour l'innovation scientifique (<i>of CIRAD</i>)
ARS	Agricultural Research Service (<i>of USDA</i>)
AVRDC	Asian Vegetable Research and Development Center, Taiwan (<i>also 'The World Vegetable Center'</i>)
BecA	Biosciences eastern and central Africa, Kenya
BIOTROP	Programme biotechnologies et ressources génétiques végétales (<i>of AMIS</i>)
CAAS	Chinese Academy of Agricultural Sciences
CARBAP	Centre Africain de Recherche sur Bananiers et Plantains, Cameroun
CBD	United Nations Convention on Biological Diversity
CGIAR	Consultative Group on International Agricultural Research, Washington, DC, USA
CIAT	Centro Internacional de Agricultura Tropical, Colombia
CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo, Mexico
CIP	Centro Internacional de la Papa, Peru
CLIMA	Centre for Legumes in Mediterranean Agriculture (<i>of the University of Western Australia</i>)
CNPMPF	Centro Nacional de Pesquisa de Mandioca e Fruticultura Tropical (<i>of EMBRAPA</i>)
CRC	Cereal Research Centre (<i>of the AAFC</i>)
CRISL	Coconut Research Institute of Sri Lanka
CSIRO	Commonwealth Scientific and Industrial Research Organisation, Australia
ECORC	Eastern Cereal and Oilseed Research Centre (<i>of the AAFC</i>)
EMBL	European Molecular Biology Laboratory
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária, Brazil
FAO	Food and Agriculture Organization of the United Nations, Italy
FCRI	Field Crops Research Institute, Thailand
IARI	Indian Agricultural Research Institute
IBPGR	International Board for Plant Genetic Resources (<i>now IPGRI</i>)
ICARDA	International Center for Agricultural Research in the Dry Areas, Syria
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics, India
IFS	International Foundation for Science, Sweden
IGD	Institute for Genomic Diversity, New York, USA (<i>of Cornell University</i>)
IITA	International Institute of Tropical Agriculture, Nigeria
ILRI	International Livestock Research Institute, Kenya
INIBAP	International Network for the Improvement of Banana & Plantain, France
INRA	Institut National de la Recherche Agronomique, France
IPK	Institut für Pflanzengenetik und Kulturpflanzenforschung, Germany
IRRI	International Rice Research Institute, Philippines
ITC	INIBAP Transit Centre (<i>of INIBAP</i>)
JIRCAS	Japan International Research Center for Agricultural Sciences
KARI	Kenya Agricultural Research Institute
LI-COR	LI-COR Biosciences, Nebraska, USA

NCT	National Corn Testing, Philippines (<i>of the USM Agricultural Research Center</i>)
NPGRC	National Plant Genetic Resources Center (<i>of the Taiwan Agricultural Research Institute</i>)
RDA	Rural Development Administration, Republic of Korea
RDAGB	RDA Gene Bank
SCRI	Scottish Crop Research Institute, Scotland, UK
SGRP	The CGIAR System-wide Genetic Resources Programme
SINGER	The CGIAR System-wide Information Network for Genetic Resources
SP	Subprogramme (<i>of the GCP</i>)
1	Genetic Diversity of Global Genetic Resources
2	Comparative Genomics for Gene Discovery
3	Trait Capture for Crop Improvement
4	Genetic Resources, Genomic, and Crop Information Systems
5	Capacity Building and Enabling Delivery
TIGR	The Institute for Genomic Research, Maryland, USA
TNAU	Tamil Nadu Agricultural University, India
USM	University of Southern Mindanao, Philippines
USDA	United States Department of Agriculture
VARTC	Vanuatu Agricultural Research and Technical Center
WARDA	West Africa Rice Development Association, Benin

Other acronyms and abbreviations

AFLP	Amplified fragment length polymorphism
AMOVA	Analysis of molecular variance (<i>statistics</i>)
APSIM	Agricultural production systems simulator (<i>a farming systems model</i>)
ARI	Advanced research institution
ASSR	Anchored simple sequence repeats (<i>also SSR-anchored PCR</i>)
BAC	Bacterial artificial chromosome
BB	Bacterial blight (<i>of rice</i>)
BBMV	Banana bract mosaic virus
BBTV	Banana bunchy top virus
BC	Backcross
BLAT	BLAST-like alignment tool
BMV	Banana mosaic virus (<i>also 'cucumber mosaic virus'</i>)
bp	Base pair
BPH	Brown planthopper
BSSA	Bulked seed sample analysis
C1	Cluster 1 activities (<i>of the GCP</i>)
CD	Compact disk
CGS	Composite genotype set (<i>of the GCP</i>)
CIM	Composite interval mapping (<i>of QTLs</i>)
cM	Centimorgan (<i>genome mapping unit</i>)
convar.	convariety
CTAB	Cetyltrimethylammonium bromide (<i>buffer used in DNA extraction</i>)
DArT	Diversity array technology (<i>used in genotyping</i>)
DARwin	Dissimilarity analysis and representation for windows (<i>software</i>)
DFA	Discriminant function analysis (<i>statistics</i>)
DH	Double haploid (<i>e.g. population</i>)
DNA	Deoxyribonucleic acid
cDNA	Complementary DNA

EST	Expressed sequence tag
F ₂ (3, 4, etc.)	Filial generation (second, third, fourth, etc.)
FAFLP	Fluorescent AFLP
GID	Germplasm identifier
GIS	Geographic information system
GPC	Grain protein content
Gy	Gray (<i>unit of absorbed dose</i>)
GW	Grain weight
IARC	International agricultural research centre
ISSA	Individual seed sample analysis
ISSR	Inter-simple sequence repeat
LD	Linkage disequilibrium
kb	Kilobase
MAS	Marker-assisted selection
MCA	Multiple correspondence analysis (<i>statistics</i>)
MSA	Multiple sequence alignment (<i>genome mapping</i>)
MSAP	Methylation-sensitive amplified polymorphism
NARES	National agricultural research and extension systems
NARS	National agricultural research systems
NILs	Near-isogenic lines
NJ tree	Neighbour-joining tree (<i>statistics</i>)
NUE	Nitrogen-use efficiency (<i>in plants</i>)
PAGE	Polyacrylamide gel electrophoresis
PCoA	Principal coordinates analysis (<i>statistics</i>)
PCR	Polymerase chain reaction
e-PCR	Electronic PCR
PGR	Plant genetic resources
PHST	Preharvest sprouting tolerance
PI	Principal investigator
PIC	Polymorphism information content
PLS	Pure-line selection
QTL	Quantitative trait loci
E-QTL	Epistatic QTL
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
SAMPL	Selectively amplified microsatellite polymorphic loci
SD	Structural disequilibrium
SIM	Simple interval mapping (<i>of QTLs</i>)
SMA	Single-marker analysis
SMA	Simple regression analysis (<i>statistics</i>)
SNP	Single-nucleotide polymorphism
ssp.	Subspecies (<i>often "subsp."</i>)
SSR	Simple sequence repeat
STMS	Sequence-tagged microsatellite site
TILLING	Targeting induced local lesions in genomes
UPGMA	Unweighted pair-group method with arithmetic mean (<i>statistics</i>)
UTR	Untranslated region (<i>on a chromosome</i>)

Appendix II: Programme

International workshop on molecular markers for allele mining Chennai, India, 22–26 August 2005

Monday, 22 August 2005		
	Introduction	
09:30–09:50	Welcome address and presentation of the MSSRF	M Velayutham
09:50–10:00	Objectives and structure of the workshop	Carmen de Vicente
10:00–10:20	The workshop within SP1 perspectives (results of Tier 1 crops + ongoing SP1 projects)	Jean-Christophe Glaszmann
10:20–10:25	Logistics	Carlos Tovar / Monique Costes
10:25–11:00	Introductions	
11:00–11:30	<i>Coffee break</i>	
	Review of results on Tier 1 crops	
	<i>Moderator:</i>	François Balfourier
11:30–12:00	Wheat	Susanne Dreisigacker
12:00–12:30	Cowpea	Visvanathan Mahalakshmi
12:30–13:00	Cassava	Paula Hurtado
13:00–14:00	<i>Lunch</i>	
	<i>Moderator:</i>	Lalith Perera
14:00–14:30	Maize	Susanne Dreisigacker
14:30–15:00	Rice	Ken McNally
15:00–15:30	Sorghum	Claire Billot
15:30–16:00	<i>Coffee break</i>	
	<i>Moderator:</i>	Merideth Bonierbale
16:00–16:30	Common bean	Mathew Blair
16:30–17:00	Barley	Zhang Jing
17:00–17:30	Potato	Marc Ghislain
19:00–21:00	<i>Welcome cocktail at hotel Residency Towers</i>	
Tuesday, 23 August 2005		
	Review of results on Tier 1 crops (continued)	
	<i>Moderator:</i>	David Hoisington
09:00–09:30	Chickpea	Hari D. Upadhyaya
09:30–10:00	<i>Musa</i>	Claire Billot
10:00–11:00	Tier 1 recapitulation and discussion	David Hoisington /

		J.-C. Glaszmann
11:00–11:30	<i>Coffee break</i>	
	Selected experiences	
	<i>Moderator:</i>	Moussa Sie
11:30–12:00	Wheat at INRA	François Balfourier
12:00–12:30	Barley at IPK	Andreas Graner
12:30–13:00	Maize at Cornell University	Gael Pressoir
13:00–14:00	<i>Lunch</i>	
	Review of other ongoing SP1 projects	
	<i>Moderator:</i>	Ajay Parida
14:00–14:20	Phenotyping; developing a platform; support from modelling	J.-C. Glaszmann
14:20–14:40	Ecotilling	Ken McNally
14:40–15:00	DArTs	J.-C. Glaszmann
15:00–15:30	Assessment of LD in sorghum	Claire Billot
15:30–16:00	<i>Coffee break</i>	
	<i>Moderator:</i>	Visvanathan Mahalakshmi
16:00–16:20	Assessment of LD in coconut	Luc Baudouin
16:20–16:40	Association studies in maize	Gael Pressoir
16:40–17:00	Development of new materials in rice and cassava	J.-C. Glaszmann
17:00–17:20	Maize out of America	Susanne Dreisigacker
Wednesday, 24 August 2005		
09:00–09:45	Discussion of Tuesday's presentations	
	Use of molecular data for subsampling	
	<i>Moderator:</i>	Andreas Graner
09:45–10:00	Introduction	J.-C. Glaszmann
10:00–10:30	SP4 project: orientations and progress	Marco Bink
10:30–11:00	Methodologies for controlling population structure in association mapping (LD mapping)	Gael Pressoir
11:00–11:30	<i>Coffee break</i>	
11:30–13:00	Working session on subsampling	
13:00–14:00	<i>Lunch</i>	
	Research activities of the guests from NARS	
	<i>Moderator:</i>	Claire Billot
14:00–14:10	Coconut Research Institute	Lalith Perera
14:10–14:20	University of Southern Mindanao	Emma Sales
14:20–14:30	AVRDC: The World Vegetable Center	Liwayway M. Engle

14:30–14:40	MS Swaminathan Research Foundation	Ajay Parida
14:40–14:50	NRC in Plant Biotechnology, IARI	T. Mohapatra
14:50–15:00	Centre for Legumes in Mediterranean Agriculture (CLIMA)	Kioumars Ghamkhar
15:00–15:10	Agriculture and Agri-Food Canada	Campbell Davidson
15:10–15:20	Choudhury Charan Singh University	H.S. Balyan
15:20–15:30	University of Agricultural Sciences (UAS)	H. E. Sashidhar
15:30–16:00	<i>Coffee break</i>	
	<i>Moderator:</i>	Jan Valkoun
16:00–16:10	National Institute of Agricultural Biotechnology of South Korea	Kyung-Ho Ma
16:10–16:20	Tamil Nadu Agricultural University (TNAU)	M. Maheswaran / Sabariappan Robin
16:20–16:30	IPR perspectives in PGR management	Malathi
16:30–17:00	General discussion	Laxmikumaran
Thursday, 25 August 2005		
	Review of the 'population structure, phenotypic information and association studies in long-generation crops' project	
	<i>Moderator:</i>	Ken McNally
09:00–09:10	Introduction	Carmen de Vicente
09:10–09:30	Coconut in Vanuatu	Luc Baudouin
09:30–09:50	Yam in Oceania	Carmen de Vicente / J.-C. Glaszmann
09:50–10:10	Potato	Merideth Bonierbale
10:10–10:30	Cassava	Paula Hurtado
10:30–10:50	<i>Musa</i> in CARBAP	Kodjo Tomekpe
10:50–11:30	<i>Coffee break</i>	
11:30–13:00	General discussion/The way forward	
13:00–14:00	<i>Lunch</i>	
14:00–14:45	Visit to MSSRF facilities and laboratories	
14:45	<i>Excursion to Mahabalipuram, a beach resort, for a special dinner</i>	
Friday, 26 August 2005		
	SP1 perspectives	
09:00–10:00	The viewpoint of germplasm curators	Ruaraidh Sackville Hamilton, Moussa Sie, Jan Valkoun, Hari D. Upadhyaya, Visvanathan Mahalakshmi

10:00–11:00	The viewpoint of observers	Ajay Parida, Andreas Graner, François Balfourier, Campbell G. Davidson
11:00–11:30	<i>Coffee break</i>	
11:30–12:30	The viewpoint of NARS representatives	Lalith Perera, Emma K. Sales, Liwayway M. Engle, T. Mohapatra, H.S. Balyan, H. E. Sashidhar, M. Maheswaran, Sabariappan Robin, Kyung-Ho Ma, Kioumars Ghamkhar, Malathi Laxmikumaran
12:30–13:30	General discussion on SP1 and SP5 perspectives	
13:30–13:45	Conclusion	Carmen de Vicente / J.-C. Glaszmann
13:45–14:30	<i>Lunch</i>	

Appendix III. Participants

Australia

Kioumars Ghamkhar
 Research Associate
 Centre for Legumes in Mediterranean
 Agriculture (CLIMA)
 University of Western Australia
 35 Stirling Highway
 Crawley WA 6009,
 Tel: (61 8) 6488 7120
 Fax: (61 8) 6488 1140
 Email: kioumars@cyllene.uwa.edu.au

Benin

Moussa Sie
 Lowland breeder
 Acting Head Genetic Resources Unit
 The Africa Rice Center (WARDAADRAO)
 WARDAADRAO-Le Centre du Riz pour
 l'Afrique
 S/c IITA Bénin - 01 BP 2031, Cotonou
 Tel: (229) 35 01 88 ext 329
 Cell: (229) 45 33 12
 Fax: (229) 35 05 56
 Email: m.sie@cgiar.org

Cameroon

Kodjo Tomekpe
 Director
 African Research Centre on Banana and
 Plantain (CARBAP)
 B.P. 832 Douala
 Tel: (237) 342 6052 / 3427129 / 7708117
 Fax: (237) 3425786
 Email: tomekpe@carbap-africa.org

Canada

Campbell G. Davidson
 International Science - Program Director –
 Genetic Resources
 Agriculture and Agri-Food Canada
 Room 762 Sir John Carling Building
 930 Carling Avenue, Ottawa, Ontario
 K1A 0C5
 Tel: (1 613) 694 2521,
 Cell: (1 613) 371 2994
 Fax: (1 613) 759-7771
 Email: CDavidson@agr.gc.ca

China

Zhang Jing
 Profesor
 Institute of Crop Science, Chinese
 Academy of Agricultural Sciences (CAAS)
 12#, South Avenue of Zhongguancun
 Beijing 100081
 Tel: (8610) 6218 6653
 Fax: (8610) 6218 9650
 Email: zhangjing@caas.net.cn

Colombia

Carlos Andres Tovar
 Administrative Assistant
 International Plant Genetic Resources
 Institute (IPGRI)
 Generation Challenge Programme
 Apartado Aereo 6713 Cali
 Tel: (57 2) 4450 048/9
 Fax: (57 2) 4450 096
 Email: c.a.tovar@cgiar.org

Carmen de Vicente
 Molecular Geneticist and Senior Scientist
 from IPGRI - GCP Subprogramme 5
 Leader
 International Plant Genetic Resources
 Institute (IPGRI) - Generation Challenge
 Programme
 Apartado Aereo 6713 Cali
 Tel: (57 2) 4450 048/9
 Fax: (57 2) 4450 096
 Email: c.devicente@cgiar.org

Mathew Blair
 Germplasm Specialist / Bean Breeder -
 Biotechnology Unit
 International Center for Tropical
 Agriculture (CIAT)
 Apartado Aereo 6713 Cali
 Tel: (57 2) 4450000
 Fax: (57 2) 4450073
 Email: m.blair@cgiar.org

Paula Hurtado
Research Assistant
International Center for Tropical
Agriculture (CIAT)
Apartado Aereo 6713 Cali
Tel: (57 2) 4450000
Fax: (57 2) 4450073
Email: p.x.hurtado@cgiar.org

France

Claire Billot
Research position
French Agricultural Research Centre for
International Development (CIRAD)
UMR PIA, CIRAD-Biotrop TA 40/03
Avenue d'Agropolis F-34398
Montpellier Cedex 5
Tel: (33 4) 6761 5858
Fax: (33 4) 6761 5605
Email: claire.billot@cirad.fr

François Balfourier
Head of "Management and Exploitation of
Wheat Biodiversity" Unit
National Institute for Agricultural
Research (INRA)
234 av. du Brézet, 63100 Clermont-Ferrand
Tel: (33 4) 7362 4346
Fax: (33 4) 7362 4453
Email: balfour@clermont.inra.fr

Jean Christophe Glaszmann
Director of PIA mixed research unit
(Cirad, Inra, AgroM) and GCP
Subprogramme 1 leader
French Agricultural Research Centre for
International Development (CIRAD)
Generation Challenge Programme
Avenue Agropolis, TA 40/03, 34398
Montpellier Cedex 5
Tel: (33 4) 6761 5922
Fax: (33 4) 6761 5605
Email: glaszmann@cirad.fr

Luc Baudouin Researcher
French Agricultural Research Centre for
International Development (CIRAD)
TA 80/03 Avenue Agropolis 34398
Montpellier Cedex
Tel: (33 4) 6761 7116
Fax: (33 4) 6761 7183
Email: luc.baudouin@cirad.fr

Monique Costes
Assistant
French Agricultural Research Centre for
International Development (CIRAD)
Avenue Agropolis, TA 40/03, 34398
Montpellier Cedex 5
Tel: (33 4) 67615925
Fax: (33 4) 67615605
Email: monique.costes@cirad.fr

Germany

Andreas Graner
Head Dept. of Genebank
Institute of Plant Genetics and Crop Plant
Research (IPK)
Corrensstr. 3, 06466 Gatersleben
Tel: (49 0) 39482 5521
Fax: (49 0) 39482 5155
Email: graner@ipk-gatersleben.de

India

Ajay Parida
Program Director of Biotechnology
MS Swaminathan Research Foundation
3rd Cross Street, Institutional Area,
Taramani, Chennai - 600113,
Tel: (91 44) 5529 9022 (Direct)/ 22541319
Res: (91 44) 5516 15614
Cell: (94 44) 024019
Fax: (91 44) 2254 1319
Email: ajay@mssrf.res.in

C. Tom Hash
Principal Scientist (Breeding)
International Crops Research Institute for
the Semi-Arid Tropics (ICRISAT)
ICRISAT Patancheru PO, Andhra Pradesh
502 324, India
Tel: (91 40) 30713071 ext. 2322, 3071, 3322
Fax: (91 40) 3071 3075
Email: c.t.hash@cgiar.org

David Hoisington
 Global Theme Leader – Biotechnology
 International Crops Research Institute for
 the Semi-Arid Tropics (ICRISAT)
 ICRISAT Patancheru PO, Andhra Pradesh
 502 324, India
 Tel: (91 40) 30713071 ext. 2322, 3071, 3322
 Fax: (91 40) 3071 3075
 Email: d.hoisington@cgiar.org

H. E. Sashidhar
 Associate Professor
 Marker-Assisted Selection Laboratory,
 Dept. Genetics & Plant Breeding, College
 of Agriculture, University of Agricultural
 Sciences (UAS)
 University of Agricultural Sciences,
 Bangalore 560 065
 Tel: (91 80) 2363 6201
 Fax: (91 80) 2333 0277
 Email: heshashidhar@rediffmail.com
 kodjo.tomekpe@tiscali.fr

H. S. Balyan
 Profesor
 Department of Genetics and Plant
 Breeding, Choudhury Charan Singh
 University
 Dr. H. S. Balyan
 Department of Genetics & Plant Breeding
 Ch. Charan Singh University
 Meerut (U. P.) 200004, India
 Tel: (91 121) 2763 564 (Res)
 (91 121) 2760 551, 2760 554 (Univ.)
 Fax: (91 121) 2760 577
 Email: hsbalyan@rediffmail.com

Hari D. Upadhyaya
 Principal Scientist
 International Crops Research Institute for
 the Semi-Arid Tropics (ICRISAT)
 Genetic Resources, ICRISAT,
 Patancheru 502 324, Andhra Pradesh
 Tel: (91 40) 30713333
 Fax: (91 40) 30713074/30713075
 Email: h.upadhyaya@cgiar.org

M. Maheswaran
 Associate Professor (Biotechnology)
 Department of Biotechnology, Centre for
 plant Molecular Biology, Tamil Nadu
 Agricultural University (TNAU)
 Tamil Nadu Agricultural University
 Coimbatore 641 003
 Tel: (91 422) 2431 222 ext. 353
 Fax: (91 422) 2433890
 Email: mahes@tnau.ac.in;
 mahesmalar@hotmail.com;
 suhirtha@eth.net

Malathi Lakshmikumaran
 Scientist
 Tata Energy Research Institute (TERI)
 Darbari Seth Block
 India Habitat Place Lodhi Road
 New Delhi 110003, India
 Tel: 468 2100
 Fax: 468 2144

Robin Sabariappan
 Associate Professor
 Centre for Plant Breeding & Genetics,
 Tamil Nadu Agricultural University
 Tamil Nadu Agricultural University
 Coimbatore TN 641 003, India
 Tel: (91 422) 247 4967
 Fax: (91 422) 551 1415
 Email: robin.tnau@gmail.com

T. Mohapatra
 Senior Scientist
 NRC in Plant Biotechnology, IARI
 Pusa, New Delhi 110 012
 Tel: (011) 25788783, 25841787, 25843554
 Fax: (011) 25843984
 Email: tm@nrpcb.org

Mexico
 Susanne Dreisigacker
 Postdoctoral researcher
 International Maize and Wheat
 Improvement Center (CIMMYT)
 Apdo. Postal 6-641, 06600, Mexico D.F,
 México
 Tel: (52 555) 804 2004
 Fax: (52 555) 804 7567 / 7558
 Email: s.dreisigacker@cgiar.org

Netherlands

Marco Bink
Senior scientist quantitative genetics -
Biometris Wageningen University &
Research Centre
PO Box 100, 6700 AC, Wageningen
Tel: (31 317) 477306
Fax: (31 317) 483554
Email: marco.bink@wur.nl

Nigeria

Visvanathan Mahalakshmi
Crop Physiologist
International Institute of Tropical
Agriculture (IITA)
PMB 5320, Ibadan
Tel: (234 2) 2412626
Fax: (234-2) 2412221
Email: v.mahalakshmi@cgiar.org

Peru

Marc Ghislain
Biotechnology Advisor Molecular
Biologist. Head Molecular Biology
Laboratory International Potato Center
(CIP)
Apartado 1558, Lima 12
Tel: (51 1) 3496017 ext. 3064
Fax: (51 1) 3175326
Email: m.ghislain@cgiar.org

Merideth Bonierbale
Leader, Germplasm Enhancement and
Crop Improvement División
International Potato Center (CIP)
Apartado 1558, Lima 12
Tel: (51 1) 3496017 ext. 3050
Fax: (51 1) 3175326
Email: m.bonierbale@cgiar.org

Philippines

Emma K. Sales
Associate Professor V,
Head Biotchnology Unit
University of Southern Mindanao,
Kabacan, Department of Plant Breeding
and Genetics, College of agriculture,
University of Southern Mindanao
Kabacan, Cotabato
Tel: (63 064) 248 2688
Fax: (63 064) 2482426
Email: ekalaw@yahoo.com

Ken McNally
Senior Scientist, Molecular Genetics -
Genetic Resources Center
International Rice Research Institute (IRRI)
DAPO Box 7777, Metro Manila
Tel: (63 2) 580 5600 ext 2766
Fax: (63 2) 580 5699
Email: k.mcnally@cgiar.org

Ruaraidh Sackville Hamilton
Head, Genetic Resources Centre
International Rice Research Institute (IRRI)
DAPO Box 7777, Metro Manila
Tel: (63 2) 5805 600 ext 2809
Fax: (63 2) 580 5699
Email: r.hamilton@cgiar.org

Republic of Korea

Kyung-Ho Ma
Junior Scientist
447-707 National Institute of Agricultural
Biotechnology
Rural Development Administration
Suwon Gyednggi
Tel: (82 31) 2991814
Fax: (82 31) 2946029
Email: khma@rda.go.kr

Sri Lanka

Lalith Perera
Head/Genetics and Plant breeding
División
Coconut Research Institute
Genetics and Plant Breeding Division,
Coconut Research Institute,
Lunuwila
Tel: (94 31) 2255300
Fax: (94 31) 2257391
Email: gpb@cri.lk

Syrian Arab Republic

Jan Valkoun
Head of Genetic Resources Unit
International Center for Agricultural
Research in the Dry Areas (ICARDA)
P.O. Box 5466, Aleppo
Tel: (963 21) 2213433 / 2225012 / 2225112
Fax: (963 21) 2213490
Email: j.valkoun@cgiar.org

Taiwán

Liwayway M. Engle
Geneticist and Head of Genetic Resources
and Seed Unit
AVRDC The World Vegetable Center
60 Yi-Ming Liao Shanhua, Tainan 74151
Tel: (886 6) 5837 801 ext. 530
Fax: (886 6) 5830 009
Email: lmengle@avrdc.org

USA

Gael Pressoir
Post-Doc associate
Cornell University, Institute for Genomic
Diversity Laboratories
Biotechnology Building, Room 175
Ithaca, NY 14853-2703
Tel: (1 607) 255-1809
Fax: (1 607) 255-6249
Email: ghp5@cornell.edu



