Generation Challenge Programme

CULTIVATING PLANT DIVERSITY FOR THE RESOURCE-POOR

2007 Project mid-year and final reports: Competitive and commissioned projects

Generation Challenge Programme

2007
Project mid-year and final reports: Competitive and commissioned projects

Table of Contents

Publisher's Note

This publication has been published in black-and-white without appendices. For colour images or appendices, please contact the appropriate Principal Investigator or Adriana Santiago, GCP Project Officer, at a.santiago@cgiar.org.

COMPETITIVE GRANTS

Subprogramme 1: Genetic diversity of global genetic resources

Competitive Project #10: Exploring natural genetic variation: developing genomic resources and introgression lines for four AA genome rice relatives

Principal Investigators:
Mathias Lorieux, IRD/CIAT
Joe Tohme, CIAT
Collaborating Scientists:
Susan R. McCouch, Cornell University
Claudio Brondani, CNPAF-EMBRAPA
Baboucarr Manneh, WARDA
Marie Noelle Ndjiondjop, WARDA
César P. Martinez, CIAT
Miguel Diago Ramirez, Fedearroz

Mid-year report
As the last project update was sent on March 6, 2007, there are no significant changes to mention. However, two points may be presented:

1. Interspecific cross *O. sativa* (cv. Curinga) x *O. meridionalis* (acc. OR44)
A total of 80 BC1F1 lines were genotyped with a set of 123 SSRs and backcrossed to the recurrent parent Curinga during the second semester of 2006 to generate the BC2F1 population. Based on the genotypic data, a total of 43 BC1F1 lines were selected with the help of the CSSL Finder software. These 43 lines bear 59 chromosomal segments covering the 12 chromosomes. Thirteen Lines were selected to have two introgression fragments and one line was selected for three introgressions fragments. The criteria of selection for lines was based on the size of the target segments, approximately 12 Mb corresponding to a suite of 4 markers of the Universal Core Map. Moreover, the segments were chosen to overlap in at least one marker.

The genotyping of the BC2F1 population has started at Cornell University and the data should be ready by the end of June 2007. Field crosses for the BC3F1 population are being held at CIAT, so as soon as the molecular data is ready we can start soon the sowing of the corresponding seeds and molecular analysis for this generation.

Additionally, a group of SSRs chosen from Gramene database, are being optimised for the original cross to be screened on the BC1F1 population in order to saturate the genetic map obtained in the first year and to get accurate data ready to be published.

2. Interspecific cross *O. sativa* (cv. Curinga) x O. rufipogon (acc. IRGC105491)A set of 130 SSR markers were chosen from the Universal Core Map and used to genotype the BC_1F_1 population (80 lines). This genotyping work allowed the identification of the introgressed fragments from the donor parent in each line. Using the genotypic data and the programme CSSL Finder, 42 plants from the 80 BC_1F_1 plants were chosen to be backcrossed to the recurrent parent (Curinga). As a result, we obtained 42 BC_2F_1 families (representing in 504 plants). This 42 BC_1F_1 plants were selected to represent the entire donor genome of IRGC 105491 (*O. rufipogon*) in small overlapping chromosomal segments in the genetic background of the recurrent parent.

In order to track the plants with the desire introgressions, nineteen of these BC_2F_1 families have been genotyped with 61 SSR markers, corresponding to the introgressed fragments of interest from chromosome 1 to chromosome 5 (5 families for chromosome 1, 3 for chromosome 2, 4 for chromosome 3, 3 for chromosome 4 and 4 for chromosome 5). The same work will be done for the rest of the families and chromosomes during the months of May and June of 2007. At the same time and according to data obtained, plants from each family having the desirable introgressed segments will be selected to evaluate its genetic background in order to optimise the recovery of the desirable genetic background. All this according to later development of the Chromosome Segments Substitution Lines set.

Tangible outputs delivered

- We pursued the development of six populations of CSSLs (Chromosome Segment Substitution Lines) that bear introgressions from the AA-genome species *O. glaberrima, O. barthii, O. meridionalis, O. rufipogon* and *O. glumaepatula*. These populations will constitute a valuable tool for genetic analyses and will allow us to identify key genomic regions that are associated to agronomically important traits. The populations are now at the BC3 stage and they should be ready for use by middle 2009.
- We are almost finished with the development of a *Universal Core Genetic Map* for rice. This map has already been demonstrated as a very useful tool to help at designing introgression populations, particularly in the case of interspecific crosses. It is based on microsatellite markers that we discovered and choose with the help of several bioinformatic packages, including some that we develop at CIAT.
- A database, *Rice Universal Core Map*, was developed and is available upon request to the author (contact: m.lorieux@cgiar.org). This database aims at providing means to easily and quickly choose a series of genetic markers to be used to genotype a population derived from a specific cross.
- Five first-generation backcross segregating populations were genotyped and five interspecific genetic maps were developed from these data. These maps will be useful to assess the recombination rates for every wild species we use and will facilitate the localization of important genes or QTLs. All the map we generated were based on the Universal Core Genetic Map.
- In order to fully exploit the information given by the genetic mapping analyses carried out using crosses that involve the *O. glaberrima* species, we collaborated with the Arizona Genomics Institute to develop a *library of Bacterial Artificial Chromosomes* (BAC) for this species. The library is available to the international community of plant genomicists, and it will constitute the basis of *positional cloning* approaches to identify and characterize important genes for *O. glaberrima* (this work was also supported by USAID funds).
- We designed a computer program that helps geneticists at creating CSSL populations. The program is called *CSSL Finder* at is available for download as a freeware at http://mapdisto.free.fr/.
- A bioinformatic tool to facilitate the discovery of single-nucleotide polymorphisms (SNPs) was set up.
- The first SNP validation tests were successful and are encouraging. They let us hope that we will soon count with a tool for fast and reliable genotyping system that will make our marker-aided breeding programs more efficient.
- Seven students and four research assistants were trained.
- Four students do shuttle research between their respective centers and Cornell University.

- The international collaboration between several ARIs, CG centers and NARS was strengthened.
- Several publications are in preparation.

Deviations from the workplan
There is no major deviation from the work plan.

Data availability
Various data sets have been produced so far:
- Genotyping data (SSR markers) of 5 interspecific BC1F1 populations
- Five interspecific genetic maps aligned to the Nipponbare sequence
- Various sets of phenotypic data collected on two *O. sativa* x *O. glaberrima* ILs populations
- Fertility data collected on a BC1F1 *O. sativa* x *O. glaberrima* population,
- Interspecific SNP locations and primers
- SSR polymorphism for a collection of cultivated and wild rice accessions
- SSR genomic location of the Universal Core Genetic Map
- All of these data will be posted on the public GCP database after their publication, if applicable

Competitive Project #13: Development of informative DNA markers through association mapping in maize to improve drought tolerance in cereals

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborators:
Edward Buckler, (Co-PI), Cornell University
Alain Charcosset (Co-PI), INRA
James Gethi (Co-PI), KARI
Pichet Grudloyma (Co-PI)
Esther Khosa, (Co-PI), SIRDC
Tim Setter, (Co-PI), Cornell University
Li Wanchen, (Co-PI), Sichuan Agriculture University
José Crossa, (Collaborator), CIMMYT
Xu Yunbi, (Co-PI), CIMMYT
Magorokosho Cosmos, (Collaborator), CIMMYT
Jose Luis Araus, (Co-PI), CIMMYT

Mid-year report
The phenotypic data for this project has been collected, submitted, and tested for data quality for China, Kenya, Zimbabwe, and Thailand. CIMMYT data has been collected and is being tested now. This represents the first year (of two) that will be included in this study. Unfortunately, the data from Zimbabwe presented too many uncorrectable problems, so collaborators from Zimbabwe were asked to sign a good-faith contract to improve their technique, and specific, step by step instructions were sent to them again. Hybrids were created again from the chosen 350 inbreds and seeds were sent to all collaborators, along with the field planting plan, in April, for collection of the second year's phenotypic data.

Final analysis of the metabolites are underway now. Dried tissues were sent previously to Cornell, and the analyses have begun in the lab of Dr. Tim Setter. All we be completed by September 1, 2007.

Sequencing of drought candidate genes have been finished for this project; a total of 79 drought candidate genes were finished for this project whereby 1 – 3 amplicons per gene, and 2 – 3 SNPs per amplicon have been found and submitted to the Illumina company. These SNPs, and 530 other SNPs from genes of

interest to this project and to others, will be developed into a chip that will be used to screen the 350 inbred lines in this study.

Selected SNPs preparatory by Ed related with drought tolerance

	pathway	**SNP number**
1	**CarotenoidBiosynth**	**27**
2	**drought related**	**461**
3	Flowertime	140
4	Inflorescence	58
5	Kernel	53
6	Leaf Morphology	68
7	Starch	89
8	TocopherolBiosynth	95
total		991

1+2=127 amplicons represent 76 unigenes, Hope another 3-4 genes can be added at last.

Tangible outputs delivered
Phenotypic data for 350 hybrids from five field locations.
SNP chip containing 991 SNPs in genes of interest and 545 high information content SNPs created by the Illumina company and is publicly available.

Deviations from the workplan
Zimbabwe phenotypic data for 2006-2007 will not be used in the final analysis, but the other 4 field sites look good and should be sufficient for the analysis.

Data availability
None yet ready for public distribution.

Competitive Project #14: Characterisation of genetic diversity of maize populations: Documenting global maize migration from the centre of origin.

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborators:
S. Taba, Genetic Resources Programme, CIMMYT
Sarah Hearne, IITA
Alain Charcosset, INRA
Zachary Muthamia, National Genebank, KARI
S.H. Zhang, CAAS
B. M. Prasanna, ICAR
Sutrisno, Indonesian Department of Agriculture
Pichet Grudloyma, Nakhon Sawon Field Crops Research Center
Phan Xuan Hao, National Maize Research Insitute

Mid-year report

Since January of 2006, this project has made very good progress in three main areas: Capacity-building and reporting; genotyping or structural characterisation; and functional genotyping.

Capacity-building and reporting: A workshop was held in Delhi, India, April 1 – 4 for training of new National Research scientists in the bulked method of genetic characterisation. The workshop report is attached. In addition, a working meeting was held from April 5 – 6. Minutes from the meeting are also attached. This meeting was to report progress to date, finalise workplans for the rest of the project, and do a preliminary analysis of data gathered to date. In addition, a template for reporting of genotyping data, as well as phenotyping and passport data of the maize landraces, was developed and provided to all collaborators to make data integration easier.

Structural characterisation: China has delivered all their SSR data for all landraces in the study. In addition, China has characterszed an additional 82 landraces with the same markers and methodologies, which can be directly compared to the original landraces. CIMMYT, France, IITA and India have nearly finished the SSR characterisation for all landraces, but have some missing data to re-run or difficult data to re-analyse; all will be finished by May 31. Thailand extracted DNA, but is unable to do the genotyping, so will send primers, DNA, and budget to China. China and Indonesia will finish genotyping by July 31. Vietnam and Kenya had no genotyping responsibilities, and so are finished with their part of the structural characterisation work. Seeds from the Chinese landraces were sent to CIMMYT for storage in the CIMMYT genebank, but unfortunately, not all paperwork was sent on time and the entire shipment was sent back again. It will be sent again.

Functional characterisation: As per our previous agreement, work is being done on the functional characterisation of Pro vitamin A genes, not drought genes, as none have been delivered to date. In addition, rather than work on the current set of populations for the functional characterization work, the previous mini-composite set of maize inbred lines (chosen by the GCP) and some high carotene inbreds (chosen by Harvest Plus) are being used. Since January, DNA has been re-extracted from all lines in this study (since previous DNA quality was not sufficiently high for sequencing). Three genes have been chosen for characterisation, which together account for nearly 80% of the variation seen in pro-vitamin A content in a wide range of tropical and temperate maize. Primers for sequencing have been ordered for these three genes. All lines in the panel have been characterised with SSR markers to study population substructure. This project is going well, and all data collection will be done in only a few more months.

Tangible outputs delivered

- Training material for the bulked characterisation of maize populations is available from the GCP website at: http://www.generationcp.org/capcorner.php?da=0763932
- SSR data for 200 maize landraces and 20 teosinte populations have been obtained, and the data set is 80% completed. Passport data for all landraces have been collected, and some phenotypic data is available as well.
- Primers have been designed that amplify several different regions within three genes in the carotenoid pathway, and one of those genes has PCR primers associated with it that explain useful variation in total pro-vitamin A levels in maize inbreds. In addition, SSR marker data for 40 SSRs and 170 inbred lines have been generated.
- New software tool: FtoL, which converts the frequency data generated by the bulk analysis, to data from individuals, which is needed by several analyses programmes. Its manual is presented as well. Both can be found on the training website: http://www.generationcp.org/capcorner.php?da=0763932

Deviations from the workplan
Thailand was unable to do the genotyping work, so they will send their DNA, primers, and genotyping budget to China, who will complete the work.

Data availability
Training material for the bulked characterisation of maize populations is available from the GCP website at: http://www.generationcp.org/capcorner.php?da=0763932

Competitive Project # 17: Allele mining based on non-coding regulatory SNPs in barley germplasm

Principal Investigator:
Michael Baum, ICARDA
Collaborators:
W. Powell, NIAB
K. Stamati, NIAB
Salvatore Ceccarelli, ICARDA (Germplasm Programme)
Stefania Grando, ICARDA (Germplasm Programme)
Sripada M. Udupa, ICARDA (Germplasm Programme)
Wafaa Choumane, Tishreen University
P. Langridge, Australian Centre for Plant Functional Genomics Pty Ltd
Mark Tester**,** Australian Centre for Plant Functional Genomics Pty Ltd
J. K. Eglinton, School of Agriculture and Wine, University of Adelaide

Mid-year report

1. Sequencing of candidate genes and SNP primer development
In order to identify SNPs in abiotic stress related genes, we have sequenced more than 80 barley genes in 8 barley lines (Arta, H. spontaneum 41-1, Tadmor, Oregon Wolfe barley dom and rec, WI3408, Sloop, Alexis). The selection was based on data available at http://germinate.scri.ac.uk/barley_snpdb/dbStats contig.html. Genes include among others nitrate stress genes and genes from the sugar and photosynthesis metabolism. In addition, we selected genes that are presumably involved in stress tolerance. Fourteen new genes were selected because their expression was influenced by drought (Guo et al. 2007), see 2. In addition, some of them are being analysed in the GCP-ADOC project. These genes were sequenced across the eight parental lines (OWBd, OWBr, Alexis, Sloop, Tadmor, Arta, WI, *H. spontaneum*). The functionality of PCR primers and extension primers was tested by genotyping the SNPs across the eight parental lines.

Table 1: Fourteen new genes with SNPs in the eight analysed parents

Contig/Acc number	Gene name	SBE	OW Bd	OW Br	Alexis	Sloop	Tadmor	Arta	WI	Hsp
AJ300144	srg6 gene for stress responsive gene protein 6	sbe	T	T	T	C	T	T	T	C
		sbe1	T	T	T	A	T	T	T	A
		sbe2	A	A	A	G	A	A	A	G
Contig10029_at	putative heat shock protein	sbe	G	G	G	A	G	G	G	G
		sbe1	C	C	C	A	C	C	C	C
		sbe2	T	T		C	T	T	T	T
		sbe3	G	G	G	G	G	G	A	
Contig15719_at	putative protein kinase	sbe	G	C	G	G	C	C	C	C
		sbe1	C	C	A	A	C	C	C	C

Contig/Acc number	Gene name	SBE	OW Bd	OW Br	Alexis	Sloop	Tadmor	Arta	WI	Hsp
Contig2924_s_at	aldehyde dehydrogenase homolog Dha1	sbe	G	G	G	G	G	G	G	A
Contig7787_at	gibberellin-stimulated transcript	sbe	C	C	C	C	T	C	C	C
		sbe1	G	G	G	G	A	G	G	G
		sbe2	G	C	G	G	G	G	G	G
Contig3499_at	60 kDa jasmonate-induced protein	sbe	G	G	C	C	G	G	C	G
		sbe1	T	G	T	T	T	T	T	G
Contig8246_at	putative amylase	sbe	G	G	G	G	G	T	G	G
M55448	Rubisco activase	sbe	G	C	C	C	C	C	C	C
	Susy	sbe	G	G	G	A	G	G	G	G
		sbe1	T	T	T	T	T	T	C	T
		sbe2			T	C	C	C	T	?
AY150676	bZIP transcription factor ABI5		-	-	-	-	-	-	-	-
Contig11041_at	LEA		-	-	-	-	-	-	-	-
Contig8961_at	abscisic acid- and stress-induced protein - rice g		-	-	-	-	-	-	-	-
Contig4853_at	Delta 1-pyrroline-5-carboxylate synthetase		-	-	-	-	-	-	-	-
Contig6505_at	Low-temperature induced protein lt101.1 (Blt101) (Salt-stress induced hydrophobic peptide ESI3)		-	-	-	-	-	-	-	-

In addition a total of 48 genes containing several SNPs mapped in the barley population Steptoe x Morex were selected from those available at the "Barley SNP Database" (http://bioinf.scri.ac.uk/barley_snpdb/). After sequencing the region containing those SNPs in Arta and *H. spontaneum*, a total of 72 SNPs and 17 Indels were found in 23 different genes. Where possible, two SNPs per gene were selected (37 Snps) and primers designed for genotyping by SBE ("Single Base Extension") (Table 2). These SNPs are currently being used to test for allelic imbalance in the Arta × *H. spontaneum*.

Table 2: SNPs found in 23 genes by sequencing in Arta and *H. spontaneum* cultivars and selected for further studies (validation by genotyping using the SBE method in progress).

Contig name	Gene name	SBE	Arta	*H. sp*
ABC00089	tubulin alpha-3 chain	sbe	C	T
		sbe1	C	T
ABC00680	plasma membrane H^+-ATPase	sbe	C	G
ABC00970	proline-rich protein	sbe	A	C
		sbe1	A	T
ABC01243	unknown	sbe	A	G

Contig name	**Gene name**	**SBE**	**Arta**	***H. sp***
ABC01259	unknown protein	sbe	A	C
		sbe1	C	T
ABC01650	phosphoglucomutase	sbe	T	C
		sbe1	C	T
ABC01797	phenylalanine ammonia-lyase	sbe	G	A
		sbe1	G	T
ABC02116	peroxidase	sbe	C	A
		sbe1	A	C
ABC02258	histone H1-like protein	sbe	G	A
ABC02555	oryzain alpha chain precursor	sbe	C	T
		sbe1	G	C
ABC02614	unknown protein	sbe	A	G
		sbe1	T	G
ABC02895	putative 37kDa chloroplast inner envelope membrane polypeptide precursor	sbe	G	A
ABC03275	putative cellulose synthase catalytic subunit	sbe	A	G
		sbe1	G	A
ABC04240	putative late embryogenesis abundant protein	sbe	T	C
ABC04853	putative Delta 1-pyrroline-5-carboxylate synthetase (P5CS)	sbe	T	C
		sbe1	G	A
ABC04923	unknown protein	sbe	C	T
		sbe1	C	T
ABC04992	succinate-semialdehyde dehydrogenase	sbe	A	G
ABC06931	GDSL-motif lipase/hydrolase (proline-rich protein APG-like)	sbe	C	T
		sbe1	A	G
ABC07010	putative protein arginine N-methyltransferase	sbe	T	C
ABC08004	integral membrane protein-like	sbe	A	G
		sbe1	G	A
ABC08208	unknown	sbe	A	C
		sbe1	T	A
ABC14826	putative glucosyltransferase	sbe	A	G
ABC15559	unknown protein (similar to cytosolic factor)	sbe	C	T

2. Analysis of gene expression under drought using the Barley Affymetrix Array (funded under BMZ)

To investigate and identify the genes associated with drought tolerance can facilitate the understanding of the molecular mechanism of drought tolerance in barley and serve for biotechnology-assisted genetic improvement of barley (Guo et al. 2007a). The 22K Affymetrix GeneChip Barley 1 array was used to monitor transcriptional changes in leaves of reproductive stage under drought stress condition in three barley genotypes, Martin and HS41-1 (both of them are considered as drought tolerance genotypes), and Morocco 9-75, a drought-sensitive genotype. Comparison of gene expression pattern responding to drought stress among three genotypes, eighteen of genes were identified to be significantly regulated

across three genotypes by imposing water deficit stress. These genes with known annotations had ever been proposed to play roles in acclimation to diverse stresses. Since these genes are cooccured in three genotypes and can be considered as common genes responsive to drought stress in both tolerant and sensitive genotypes of barley, suggesting either that theses genes are not involved in drought tolerance, or that they may be necessary, but alone are not sufficient to confer drought tolerance in barley. Seventeen genes were highly induced only in two drought tolerant genotypes of Martin and HS41-1 under drought stress. Based on the putative functions reported in previous studies, these genes could be involved in the control of stomatal closure via carbon metabolism, in synthesis of osmoprotectant of glycine-betaine, and in generation of protectant in reactive oxygen species (ROS)-scavenging, as well as in products for membrane and protein stability, and suggesting these may be the reasons to result in drought tolerance of Martin and HS41-1. These results could provide new insights to elucidate the mechanism for drought tolerance in barley during the reproductive stage. The candidate genes identified in this study are being tested for allelic imbalance.

3. Testing of allelic imbalance under drought stress

A greenhouse experiment was conducted in an ICARDA temperature-controlled greenhouse with 16 hours daylight at 30°C and 8 hours dark period at 20°C. Eight crosses were subjected to the treatment: Tadmor/Sloop, Sloop/Tadmor; Tadmor/WI, WI/Tadmor; Sloop/WI, WI/Sloop and Arta/H. spontaneum, H. spontaneum/Arta. Drought stress was imposed in the vegetative stage before flowering and at the heading stage by reducing available water in the soil from 70% (non-stress condition) to 10% (severe drought stress). Total RNA was extracted from three biological replicates per F1 or RF1, and every replicate consisted of pooled leaves from three plants.

Twenty-five different genes were tested for allelic imbalance on the RNA extracted from the F1 and RF1 derived from Tadmor/Sloop and Sloop/Tadmor treated in at the generative stage (Table 2). Of the twenty-five tested genes thirteen showed significant differences in allele expression (Fig 2, Table 2). For example, the putative heat shock protein showed dominant expression of the Tadmor allele under control conditions (0.81, 0.92), but under drought the allelic expression of the Sloop allele increased and both alleles were expressed in approximately equal amounts (Table 3). For the Jasmonate induced protein the Tadmor allele was expressed less than the Sloop allele (F1:0.27, RF: 0.4) under control conditions. Under drought, the expression of the Tadmor allele was further significantly reduced to 0.09 and 0.18 in the F1 and the RF1, respectively. At the same time allelic expression was different between the F1 and RF1 and thus showed imprinting. In addition, eleven genes exhibited allelic imbalance with a dominant expression of the Tadmor allele for six genes and a dominant expression of the Sloop allele for another five genes. For these ten genes the differences in allelic expression were, however, observed under control conditions and drought treatment, and cis-regulation was therefore independent of the drought stress.

We are currently analysing the expression of the same genes in the Tadmor/Sloop cross subjected to the drought treatment in the vegetative stage. Out of five genes analysed so far, two showed changes in allelic expression from the vegetative to the generative stage. Gene ABC871, for example, exhibited a dominant expression of the Tadmor allele in the vegetative stage (0.8), while in the generative stage the expression of the Sloop allele was dominant (expression Tadmor allele: 0.3).

In addition, we are testing allelic expression in the crosses Tadmor/WI and Sloop/WI, Arta/H. spontaneum to investigate allelic expression in different genetic backgrounds and in a wide cross involving exotic germplasm.

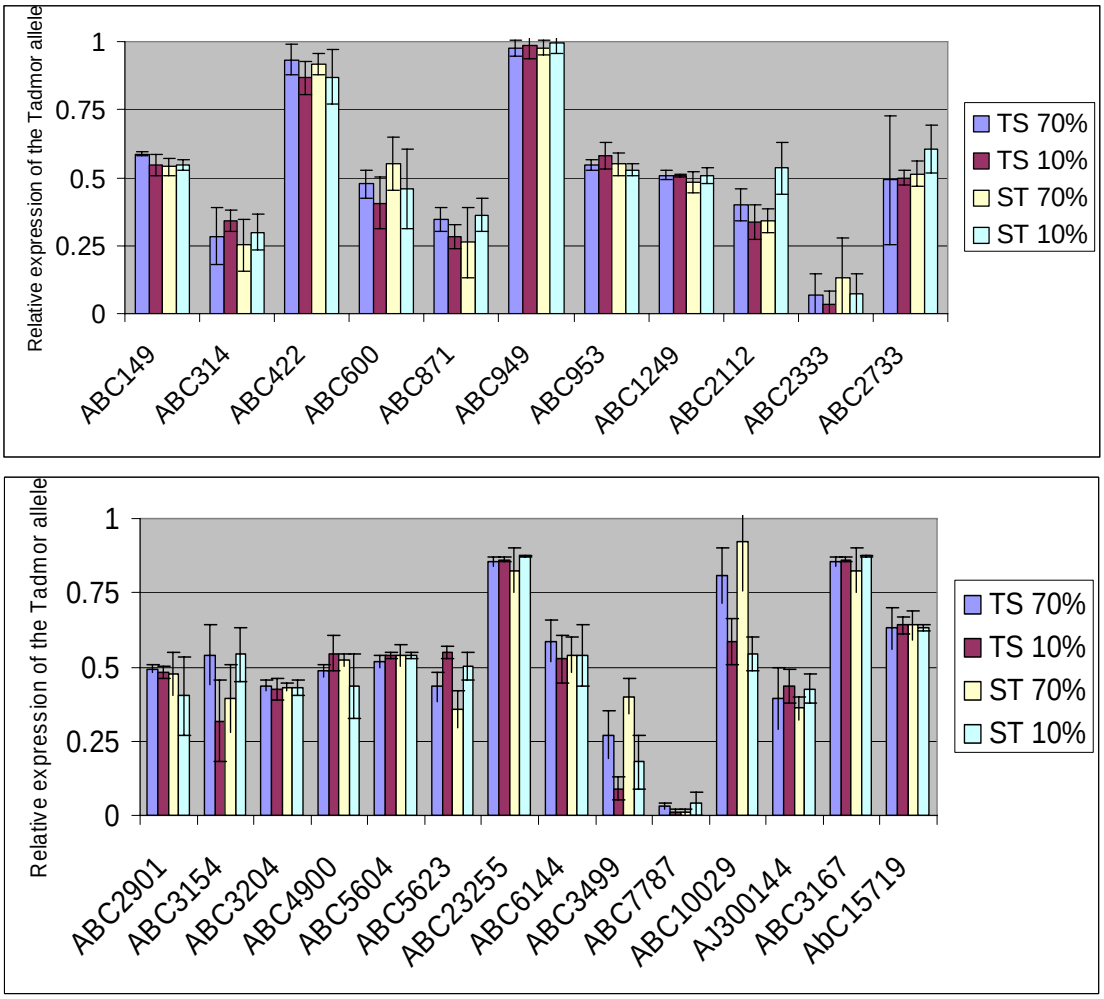


Figure 2: Relative expression of the Tadmor allele in the hybrids derived from Tadmor/Sloop (TS) and the reciprocal hybrids Sloop/Tadmor (ST) under control conditions (70%) and drought stress (10%)

The obtained data demonstrated that half (50%) of the barley genes are regulated in cis. Since the analysis focused on leaf tissue and one cross we expect that the number of genes regulated in cis will be even higher when testing further barley crosses and tissues. However, the analysed genes represent a selection of genes potentially involved in stress response. As these genes are possibly involved in adapting the plant to its environment, a preponderance of cis regulation can be expected. Using the same technique in maize, overall cis-regulation is estimated with 70%, when considering expression in different tissues and under different environmental conditions. However, imbalance of allelic expression in single maize tissues is around 40% which is comparable to our data. Similar results were obtained for humans where differential allelic expression ranges from 18 to 50% according to the tissue analysed.

Table 3. Examples of allelic expression in ten genes for the cross Tadmor/Sloop

		Tadmor/Sloop		**Sloop/Tadmor**	
Gene	Contig	Control (70%)	Drought stress (10%)	Control (70%)	Drought stress (10%)
Glucose 6 phospate isomerase	ABC06144	0.58 ± 0.07	0.53 ± 0.08	0.54 ± 0.06	0.54 ± 0.1
chlorophyll a/b-binding protein	**ABC00422**	**0.93 ± 0.06**	**0.86 ± 0.06**	**0.92 ± 0.04**	**0.86 ± 0.1**
Unknown protein	ABC00953	0.55 ± 0.02	0.58 ± 0.05	0.55 ± 0.04	0.53 ± 0.02
Peroxidase	**ABC02112**	**0.40 ± 0.06**	**0.34 ± 0.06**	**0.34 ± 0.05**	**0.54 ± 0.09**
Putative phytoene dehydrogenase precursor	ABC05604	0.52 ± 0.02	0.54 ± 0.01	0.54 ± 0.04	0.54 ± 0.01
Unknown protein	**ABC23255**	**0.85 ± 0.02**	**0.86 ± 0.01**	**0.83 ± 0.07**	**0.87 ± 0.01**
Jasmonate induced protein	**ABC3499**	**0.27± 0.08**	**0.09±0.04**	**0.4±0.06**	**0.18±0.09**
Gibberellin-stimulated protein	**ABC7787**	**0.03±0.01**	**0.01 ± 0.01**	**0.01±0.01**	**0.04±0.03**
Putative heat shock protein	**ABC10029**	**0.81± 0.09**	**0.59±0.07**	**0.92±0.16**	**0.54±0.06**
Putative protein kinase	**ABC15719**	**0.63± 0.07**	**0.64± 0.03**	**0.64± 0.05**	**0.63± 0.01**

4. Mapping of gene expression in the population Arta/H. spontaneum

In order to complement information on agronomic and physiological performance, as well as on allelic imbalance under drought with information on QTL for gene expression, the RIL population Arta/*H. spontaneum* was grown under two drought regimes. The population was subjected to drought stress at the generative stage by reducing water in the soil from 70% (non-stress condition) to 10% (severe drought stress). Expression differences between lines and between stress and control will be tested using RT-PCR. Based on results from the previous mapping efforts (Baum et al. 2003, Guo et al. 2007b) and from the allelic imbalance assays, expression analysis will focus on genes within QTL 'hotspots', such as on 2HL or on genes showing allelic imbalance. To complement this approach 37 SNPs are being mapped in this population using tetra-ARMS ("Amplification Refractory Mutation System) PCR (Ye et al., 2001). This is a convenient and economical procedure for genotyping SNPs which has been previously used in barley (Chiapparino et al., 2004).

Meiotic mapping of genes that putatively harbour cis-acting elements in the Arta/*H. spontaneum* population together with information on quantitative measures of gene expression under stress will give a valuable insight into the regulation and genetic control of drought tolerance.

Literature:

Baum M, Grando S, Backes G, Jahoor A, Sabbagh A, Ceccarelli S (2003) QTLs for agronomic traits in the Mediterranean environment identified in recombinant inbred lines of the cross 'Arta' x *H. spontaneum* 41-1. (Theor Appl Genet (2003) 107:1215–1225.

Chiapparino E., Lee D., Donini P (2004) Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS–PCR. Genome 47:414-420.

Guo P., M. Baum, S. Grando, R. K. Varshney, A. Graner, S. Ceccarelli, G. Bai, R. Li, J. Valkoun (2007) Expressional Analysis of Barley Genes Responsive to Water stress between Drought Tolerance and Sensitive Genotypes during Reproductive Stage. In prep.

Guo P, M. Baum, R.K. Varshney, A. Graner, S. Grando, S. Ceccarelli (2007b) QTLs for chlorophyll and chlorophyll fluorescence parameters in barley under post-flowering drought, in prep.

Ye S., Dhillon S., Ke X, Collins A.R., Day I.N.M. (2001) An efficient procedure for genotyping single nucleotide polymorphisms. Nucleics Acids Research 29 (17):e88.

Tangible outputs delivered

1. F1 and RF1 plants derived from the Oregon Wolfe barley parents were subjected to cold stress and RNA was extracted from leaves and roots of stressed and control plants. Of nine genes tested three showed allelic imbalance. Significant differences in allelic expression were independent of the treatment and tissue.

2. F1 and RF1 derived from twelve different crosses were subjected to drought stress at the developmental and generative stage, and RNA was extracted from leaves and roots. So far, four genes were tested for imbalance in the Tadmor/Sloop and Sloop/Tadmor hybrids. For one gene, dehydrin12, allelic imbalance was detected in stressed and control leaves at the generative stage.

3. In order to select the most promising candidate genes for allelic imbalance assays a consensus map containing QTL for agronomic and physiological performance under drought was compiled. This consensus map showed QTL clusters for performance under drought and genes mapped close to these QTL were prioritized for further analyses.

4. The RIL mapping population Arta/H. spontaneum was subjected to the same stress treatment as that used for the allelic imbalance assays, and a QTL analysis was conducted for parameters reflecting the photosynthetic status of the plants.

5. The Arta/H. spontaneum RIL population has been grown under control and drought conditions for the purpose of mapping expression under drought. This allows to test whether eQTL coincide with structural genes or rather point to an unlinked regulatory locus.

6. Parental lines (8) were sequenced for 70 genes and primers were designed to target the detected SNPs.

7. Training session on allelic imbalance took place. *(list the outputs that have been delivered so far)*

8. Additional genes were selected that are presumably involved in stress tolerance (fourteen new genes identified in a microarray study, genes that are tested in the GCP-ADOC project. These genes were sequenced across the eight parental lines and the functionality of PCR primers and extension primers was tested by genotyping the SNPs across the eight parental lines.

9. Another set of 48 genes from the "Barley SNP Database" were sequencing across Arta and *H. spontaneum*, and a total of 72 SNPs and 17 Indels were found in 23 different genes. SNPs were selected (37 SNPs) and primers designed for genotyping by SBE ("Single Base Extension"). These SNPs are currently being used to test for allelic imbalance in the Arta × *H. spontaneum.*

10. 37 SNPs are being mapped in this population using tetra-ARMS ("Amplification Refractory Mutation System) PCR.

Deviations from the workplan

Timelines and Milestones.

2005: **Develop genomic, expression imbalance assays**	**Activities carried out**	**Milestone completed**
1. Establish robust protocol to identify and quantify frequency of *cis*-acting regulatory elements.		
a. SNP discovery primer design and develop genomic, expression imbalance assays. (NIAB, University Udine)	Ongoing, to be continued in 2006	Continous
b. Co-ordination and curation of populations, hybrids and mRNA extraction protocols. (ICARDA)	Hybrids continuously established, Mapping population Arta/H.spontaneum 41-1 provided to NIAB	Yes
c. Hybridity of material confirmed via microsatellite analysis. (ICARDA)	..and verified by SSR analysis	Yes
d. Assembly of candidate genes for evaluation (All participants).	Ongoing effort	Yes
e. Organise Annual workshop to review progress and disseminate information and knowledge.	Meeting took place in San Diego, PAG, January 2006	Yes
2006: **Design and execute factorial experiments to test the influence of various stress treatments on expression of alleles in heterozygous condition.**	**Activities carried out**	**Milestone completed**
a. Parental and hybrid RNA extraction completed. (ICARDA)	Hybrids need continuously to be established and hybridity tested when leaf samples are collected	Yes
b. Frequency of cis acting regulatory elements quantified for 72 reciprocal hybrids (Udine and Adelaide) based on root tissue subjected to drought stress.	This seems to be unrealistic. Not all the SNP assays developed for one cross are applicable to other crosses. This was too ambitious from the beginning, but we aim to assay as many hybrid combinations as possible	No, Ongoing activity
c. Based on SNPs in LD with expression imbalance meiotically map polymorphim in D.H. population. (ICARDA).	Mapping population has been grown in 2006 at ICARDA and leaf samples have been collected for RNA extraction, needs to be completed in 2007	Ongoing
d. Organise Annual workshop.	ICARDA and NIAB scientists met at Udine to review protocols	Yes

2006: **Design and execute factorial experiments to test the influence of various stress treatments on expression of alleles in heterozygous condition.**	**Activities carried out**	**Milestone completed**
e. Ensure that at least 2 national programme scientists are enrolled in Masters course at University Adelaide.	Through transfer of activities from Adelaide to NIAB this is not applicable, however, one Moroccan PhD students has been enrolled in Adelaide, one Ethiopian student has completed his PhD in Australia.	Yes
f. Review opportunities for publication of findings	Postponed to 2007	Ongoing

2007 **Meiotically mapping *cis*-acting regulatory variation in doubled haploid population of barley.**	**Activities carried out**	**Milestone completed**
a. Extend quantification of imbalance assays to developing embryo (6-8 dpa), developing endosperm and leaves. (Udine and NIAB).	Ongoing. Imbalance is now checked and tested in different tissue(roots and leaves), stages (vegetative and generative).	Ongoing
b. Meiotically map polymorphism and correlate with segregating phenotypic traits (ICARDA).	Already started in 2006, However, How many expression differences genes can be mapped**?**	Ongoing
c. Validate the efficacy of regulatory haplotype blocks in different genetic backgrounds as surrogates for novel allele detection. (All).	Ongoing in 2007	Ongoing
d. Disseminate new knowledge on regulatory variation as a source of new polymorphisms for deployment in breeding.	Later in 2007	
e. Devise 'haplotype tag' based approaches to enable the deployment of sequence based diagnostics in breeding programmes	Later in 2007	

Data availability

1 (a). Sequencing data:

We have sequenced more than 100 barley genes in 8 barley lines (Arta, H. spontaneum 41-1, Tadmor, Oregon Wolfe barley dom and rec, WI3408, Sloop, Alexis). Primers have been designed and are currently being tested for the imbalance assays. Additionally, 37 genes have been sequenced in Arta/H.sponateneum.

1 (b). Where the data sets have been posted, if not where and when they will be posted (local/public GCP database):

Sequences and primer information have been registered in the GCP database. They will were posted May 2007.

2. Microaray data
Expression data were generated as part of a BMZ funded project. They will be made available once the data are published in refereed journals.

3. Data on allelic imbalance
Data will be posted on GCP database after publication.

Subprogramme 2: Comparitive Genomics for Gene Discovery

Competitive Project #1: Identifying genes responsible for failure of grain formation in rice and wheat under drought

Principal Investigator:
John Bennett, IRRI

MID-YEAR REPORT NOT SUBMITTED

Competitive Project #2: Revitalising marginal lands: discovery of genes for tolerance of saline and phosphorus deficient soils to enhance and sustain productivity

Principal Investigator:
Abdelbagi M. Ismail, IRRI
Collaborators:
David J. Mackill, IRRI
Michael Thomson, IRRI
Sigrid Heuer, IRRI
Xiaochun Lu, IRRI
Glenn Gregorio, IRRI
Rakesh Kumar Singh, IRRI
Matthias Wissuwa, JIRCAS
Eduardo Blumwald, University of California
Zeba I. Seraj, Dhaka University
Masdiar Bustamam, ICABGRRD,
Timothy J. Close, University of California
Ghasem H. Salekdeh, ABRII
Massahiro Yano, NIAS

Mid-year report

SALINITY (*Saltol*)

Objective 1: Further precision mapping of the *Saltol* locus

IRRI: *Fine mapping Saltol.* To further saturate the molecular map at the *Saltol* locus, 12 additional SSR markers from the International Rice Genome Sequencing Project (IRGSP) were tested and six of them were found to be polymorphic. These six markers span 11.0 to 11.5 million basepairs (Mb) on the short arm of chromosome 1 (RM10694, RM10696, RM10701, RM10711, RM10713, RM10740). In addition, the previously described marker AP3206 (from Niones, 2004) was found to be a good replacement for RM8094, which had been difficult to score. To identify additional recombinants across the *Saltol* region, advanced backcross populations of IR29/Pokkali, IR29/FL478, BR28/FL478 and IR64/FL478 were genotyped with 3 to 6 SSRs at *Saltol*. The marker data was used to select 169 individuals for further study, and these were genotyped with 20 SSRs across chromosome 1, and 16 additional markers across the rest of the genome. This data is currently being analysed to identify recombinant individuals that

contain different Pokkali introgressions across the *Saltol* region, and these will be phenotyped to identify tolerant plants to more precisely define the location of *Saltol*.

Objective 2: Development and validation of genetic markers for MAS

Dhaka Univ*: Identification of markers for introgression of Saltol into farmer-popular rice varieties*. Tolerant and sensitive extremes (seedling tolerance in hydroponics) of all available 10 families of BC_3F_3 near isogenic lines (NILs, between IR29 and Pokkali) provided by IRRI have been analysed at the *Saltol* loci using polymorphic markers between IR29 and Pokkali. Positions of linkage in the *Saltol* region of rice chromosome 1 have been re-confirmed as ~11.25-11.45, ~12.28-12.34, ~12.73-12.79 million base pairs (mbp) and an additional region of strong linkage at ~12.09 mbp. Respective markers at these positions are called CatCh-NC (11.25 mbp), Met Syn NC (12.34 mbp), AP3143-72760, (or Rice BAC 12.735, 12.79 mbp) and ABC transporter NC (12.09 mbp). Marker-trait co-segregation among the tolerant progeny varied from 52-86% among the different families and positions. Among the different families of NILs, the marker linkage for tolerant progeny at 11.25 mbp was 63-81%, low to 86% at 12.09 mbp (Figure 1), low to 55% at 12.34 mbp and 52-63% at 12.79 mbp. At all other positions linkage was less than 10%. Corresponding SSR markers close to these positions are still being optimised.

Dhaka Univ: *Progeny testing in breeding populations*. F5 and F4 progeny of two breeding populations, IR60494 × BR29 and IR52724 × BR36, respectively, were screened at the seedling stage in hydroponics. Both Pokkali and Nona Bokra are used as donors of salt tolerance of IR60494, whereas only Pokali was the donor for IR52724. Tolerant and sensitive extreme genotypes were analysed with 5-7 markers in the *Saltol* locus. Greater than 75% marker-trait linkage was found with markers from 12.2-12.34 mbp in the IR60494 ×BR29 population. Marker RM493 at position 12.2 showed 93% linkage (Figure 2). A larger population (F6) will be tested for confirmation. The marker Met Syn NC at position 12.34 mbp showed the highest (73%) marker-trait linkage for IR52724 × BR36 (Figure 3). Here also a larger population (F5) will be used for confirmation.

IRRI: *Analysis of allelic variation at Saltol markers across multiple Pokkali accessions.* Since there are many different accessions having the same name "Pokkali," we undertook a study to examine the allelic variation across different Pokkali accessions compared with our RILs and NILs. We conducted DNA fingerprinting analysis of 8 individual plants from seven Pokkali accessions stored in the T.T. Chang Genetic Resources Centre (IRGC Accession numbers 8948, 15388, 15602, 15661, 26896, 28609, and 108921) and an analysis of 46 chromosome 1 SSRs and 32 background SSRs was used to define the relationships between different Pokkali accessions compared with our control lines IR29, FL478, NIL-17 and NIL-30. A tight cluster consisted of four accessions was identified based on the SSR data: an IRRI Pokkali used to make crosses in the breeding programme, the Pokkali accession used as a control at UC Davis, and the two accessions IRGC26896 and IRGC108921 (Figure 4). All four of these accessions are highly tolerant to salinity, in contrast to the intermediate tolerance of accessions IRGC28609 and IRGC8948 and the susceptible IRGC15602 (data not shown). This data is also being used to define the origin of the Pokkali introgressions found in our RILs and NILs across the *Saltol* region. For example, an unusual Pokkali introgression from 11.2 to 11.4 Mb that was identical to the allele size of IRGC28609 but none of the other Pokkali accessions was found in the tolerant line FL478 (Figure 5).

Objective 3: Identification and validation of candidate genes

UCD: To understand the contribution of the specific introgressed locus of Pokkali at the *Saltol* region/ QTL in salinity tolerance at seedling stage, NILs of IR29/ Pokkali at BC_3F_5 were characterised physiologically. Also, expression analysis of the candidate genes selected at the *Saltol* region of chromosome I was performed on selected NILs. Nine NILs were selected based on the following criteria: 1) presence and absence of Pokkali introgression at the *Saltol* region from 11.6 -12.7 Mb on chromosome 1; 2) Na/K content ratio of shoots; and 3) scoring of seedling salt tolerance. Among the nine NILs, lines

17, 23 and 24 were selected for having fragments of the *Saltol* loci, and lines 30 and 35 were selected for lacking the *Saltol* loci.

The eight genotypes (5 NILs and 3 control lines) were grown hydroponically in Yoshida solution (Gregorio, 1997) and after seven and 14 days in 120 mM NaCl nutrient solution, leaves and root tissues were harvested. Plant tissues harvested from seven days under saline conditions were assessed for ion content, metabolite profile and the relative abundant transcripts of selected genes from the *Saltol* region.

Physiological Characterisation of the NILs and Parental lines. Based on the biomass production and damage scores, progenies FL478 (RIL), NIL17 and NIL30 (NILs) are able to adapt in saline nutrient solution of 120 mM NaCl relative to Pokkali (Figure 6). The NILs (NIL 17 and NIL 30) performed equally in the presence of salt at the seedling stage. Less damage and moderate growth were observed in the two NILs. After 14 days in salinity, Pokkali showed vigorous growth as compared to IR 29. Growth slowed down in the sensitive genotypes during salt stress.

Ion content in the shoot and root tissues after 7 days in 120 mM NaCl. The two NILs (NIL 17 and NIL 30) displayed different sodium uptake from root to shoot (Figure 7). The NILs 17 displayed low shoot/ root Na^+ ratios, indicating an ability to exclude Na^+ ions from the shoot, similar to Pokkali, a salt-tolerant variety known to be sodium excluder. On the other hand, NIL 30 displayed high shoot/ root Na^+ ratios, similar to IR 29, a very salt-sensitive indica cultivar. The NIL 30 was able to withstand concentrations of up to 120 mM NaCl. The salt-tolerance of NIL 30 suggested its ability to compartmentalize sodium into its older leaves. The ion data suggests two possible mechanisms of salt tolerance in Pokkali.

Secondary metabolite production in the root and shoot during salt stress. The photoassimilates produced under salt stress are utilised to support processes such as growth, maintenance and osmotic adjustment. The availability of carbohydrate is not limiting for growth under saline conditions, although the regulation of carbon allocation and partitioning may have an important influence in the maintenance of growth rate and yield (Munns and Termaat, 1986). Both IR29 (sensitive) and Pokkali (tolerant) displayed high levels of sucrose in the shoot (Figure 8). But biomass production was negatively correlated with the abundance of photoassimilates in IR29. These results suggest that sensitive genotypes used its photoassimilates for survival not for growth. Moreover, high accumulation of trehalose in Pokkali roots under salt stress and high levels in the shoot in unstressed condition suggests a high capacity for the translocation of the trehalose produced in the leaves to the roots in order to protect (osmoprotectant) the tissue that was exposed to salt stress (Figure 8). The accumulation of inositol in the leaves in all genotypes suggested that photoassimilates can be mobilized from the leaves to the roots, thus linking photosynthesis to root metabolism (Nelson et al., 1998). For instance, the abundance of lactic acid in the root in all genotypes is basically due to anaerobic growing condition.

UCD: Table 1 is the list of genes that were gathered from the Exon array data provided by our Korean collaborator. Candidate genes were selected based on expression profile and their putative function, possibly associated with salt tolerance. The candidate genes used in our RT-PCR analysis are putatively involved in transporting ion (SKC/HKT, Cation chloride co-transporter) amino acid (SEC A and Sec 23/24) and sugars (sugar transporter/ Multifacilatory transporter), S-Adenosyl-methionine synthase which could be associated with polyamines biosynthesis or, it could function as a good donor of methyl group for production of osmolytes. Moreover, an example of serine threonine kinase gene and WD 40 were also used because their involvement in signal transduction pathways might be associated with the response to salt stress. However, the expression data are not yet conclusive (Figures 9 and 10). Evaluation of the expression of other genes through Real-time PCR is on-going for other set of genes like transcription factors (transcriptional regulation) and plant formin (involved in Cell mobility and integrity).

Objective 4: Functional confirmation of *Saltol* and impact assessment

Dhaka Univ: *Breeding for introgression of Saltol*. Breeding has been initiated from July 2006 at BRRI to introgress the salinity tolerance traits of Pokkali and its derived tolerant RILs FL378 and FL478 into T. Aman varieties, BR11, BR23 and BR41 and Boro varieties, BR28 and BR29. MAS will be applied to two backcross populations, one for *Saltol* introgression into BR11 and one for introgression into BR28, popular varieties of the monsoon and the boro season, respectively. Markers (being identified above) will be applied at BC1 to allow identification of individual progeny containing *Saltol* with the maximum genome of BR11 or BR28. F_1 progeny of the monsoon crop have already been planted and F_1 crossing for the boro crop completed. F_1 seeds have already set for the latter. Background and foreground marker information has been provided by IRRI. DNA isolation and testing of BC_1 will commence in August 2007.

IRRI: *Graphical genotypes of contrasting NILs analysed with SSR markers.* Two NILs were selected by Dr. Blumwald's lab at UC Davis for functional characterization, and DNA samples of these NILs were sent to IRRI for SSR genotyping to confirm the Pokkali introgressions in these lines. DNA samples from two bulk extractions for each line were analysed with 78 SSR markers (46 from chromosome 1 and 32 from the rest of the genome) along with the multiple Pokkali accessions described above. In contrast to previous SSR data of these lines, several fragments of Pokkali introgressions were detected in both NIL-17 and NIL-30, though with different patterns (Figure 5). The SSR data from these lines at *Saltol* and at background loci will be further analysed and compared with the phenotype and gene expression data to better characterise the effects of the Pokkali introgressions.

Transfer of Saltol into popular varieties using MAB: Two MAB populations using FL478 as the salt tolerant donor and BR28 and IR64 as the recurrent parent were genotyped with *Saltol* SSR markers and selected individuals were backcrossed. BC_2F_2 seeds are being phenotyped to confirm the presence of the *Saltol* introgression, and BC_3F_1 seeds for the BR28 population and BC_2F_1 seeds for IR64 are available for the next round of SSR genotyping for foreground and background selection.

PHOSPHORUS DEFICIENCY (*Pup1*)

Objective 2: Development and validation of genetic markers for MAS

IRRI: *Molecular marker development for Pup1*. The testing of Simple Sequence Repeats (SSR) markers for *Pup1* was continued and so far, 223 markers polymorphic between Kasalath and Nipponbare were identified (169 SSR and 54 STS markers). The physical map of the *Pup1* sister NILs 14-4 (+Pup1) and 14-6 (-Pup1) is illustrated in Figure 11. Based on this data, it is now possible to further purify the *Pup1* NIL 14-4 from additional Kasalath background introgression. This is especially important for a Kasalath introgression on Chr.8 which is associated with root growth (Wissuwa *et al*., 2002) and that might mask the effect of *Pup1*. F1 plants derived from a NIL14-4 x Nipponbare backcross are currently growing for seed increase and F2 plants will be genotyped using selected markers to identify plants with fewer additional introgressions. Preliminary data showed that the RM511 marker, indicative of a QTL associated with yield under drought (Bernier et al., 2007; see below), is not polymorphic between Nipponbare and Kasalath. This is currently being analysed in more detail.

The SSR3 marker located in the *Pup1* region is currently being tested in a wide range of germplasm for polymorphism and usefulness as marker for the introgression of *Pup1* into breeding lines. The development of allele specific markers for *Pup1* candidate genes is still ongoing.

Phenotyping system for Pup1 and introgression into recipient parents. The polymorphic SSR markers developed for Kasalath and Nipponbare (see above) were successfully used to confirm presence of *Pup1*

in NIL14-4 and it's absence in NIL14-6 in the seeds used at IRRI. All plants used in experiments are now routinely being genotyped. In addition, in the drought experiment (see below) clear differences between the *Pup1* NILs were observed for the first time. Based on this progress it seems now feasible to identify suitable recipient parents and to develop a high throughput phenotyping system for *Pup1*. Putative recipient parents that do not possess the *Pup1* locus were already identified earlier using *Pup1* dominant gene specific markers (see last report), however, the P- efficiency of these varieties is currently unclear. We will now screen these accessions under P-deficiency in combination with drought to determine which varieties might benefit from *Pup1* introgression.

1. **ICABIOGRAD:** *Identification of suitable recipient cultivars based on preferred local varieties of selected NARES***.** To confirm last year's results, we repeated the evaluation of upland varieties under P deficiency at two screening sites (Sitiung-West Sumatra and Kentrong-West Java). Two greenhouse experiments using soil sample from both field experiments were set to compliment the field experiment. These experiments are on going. Although nutrient solution was not recommended for phenotypic evaluation under P-deficiency during the last planning meeting held at IRRI, we still used this test to evaluate the responses of test varieties to to both Al toxicity and P- deficiency. Molecular analysis using more molecular markers as suggested by Dr. Sigrid Heuer at IRRI meeting for detecting polymorphism between recipient and donor plant was completed (Figure 12). These molecular data was supported by performance of these material under greenhouse conditions using soils from sites identified as suitable based on field experiments. Therefore we had come to conclusion that those three Indonesian upland varieties can be use as recipients of *Pup-1* (Table 2). Moreover, to develop NILs between *Pup-1* donor and Indonesian upland varieties, 9 backcross combinations were made and are being advanced for further analysis and selections (Figure 13). One problem encountered in the field this year is that Kasalath and all lines carrying *Pup-1* allele are affected by leaf blast and Neck blast in Sitiung, West Sumatra.

Objective 3: Identification and validation of candidate genes

JIRCAS: *Quantification of gene expression in candidate genes*. RNA from roots and shoots of soil grown Nipponbare and NIL-*Pup1* plants were used in an attempt to identify candidate genes based on the gene regulation pattern. One gene was strongly induced under P deficiency in NIL-*Pup1* but not in Nipponbare. This gene is similar to a Nucleoside-diphosphate-sugar epimerase and may be involved in root cell wall metabolism, particularly in growing root tissue close to the root tip. This gene is now one of the top candidates and will be used to transform Nipponbare once the full-length cDNA sequence has been determined.

IRRI: *Pup1 candidate genes*. The assessment and short listing of the *Pup1* candidate genes is still ongoing and a summary of the candidate genes expression using semi-quantitative RT-PCR is shown in Figure 14. Gene specific primers were designed for quantitative RT-PCR analyses and sent to M. Wissuwa at JIRCAS.

In order to clarify whether gene #5 is part of gene #4 or, as predicted, located within the gene #4 intron in reverse direction; RT-PCR analyses were performed using primers targeting specific exons of genes #4 and #5 (Figure 15). This analysis so far confirmed (i) presence of a predicted 5'UTR of gene #4 in a distant exon thereby creating the intron where gene #5 is located, (ii) confirmed a distinct gene #4 EST found in the database with an exon (exon II) that is not predicted for gene #4, (iii) this exon II is not present in gene #5 in reverse direction (iv) confirmed the expression of gene #5 by targeting the first predicted exon (exon I). The same exon though smaller was predicted before based on an old gene annotation (old gene #28) that provided first evidence for the existence of gene #5. A small, second exon predicted for gene #28 was not confirmed with the primers used for this analysis. SMART PCR will now be performed to isolate the full length cDNA of gene #5.

Five genes (#4, #5, #18, #26, #43) have been selected for Kasalath transformation and the construction of the vectors is currently ongoing (Figure 16). This approach aims at the down regulation of the genes in Kasalath using the RNAi technique and it is expected to see a loss of function phenotype in Kasalath

transgenic plants. For gene #5 we will specifically target the exon I whereas for gene #4 exon II and an additional downstream exon is targeted that is presumingly not present in the EST transcript. This will ensure specificity of gene silencing.

In parallel, preparations are ongoing to isolate the full length cDNAs of the candidate genes from Kasalath for overexpression in Nipponbare.For this, primers were ordered that amplify the entire coding region based on the predicted gene structure. In parallel, SMART PCR will be performed to ensure that cDNAs are full length and to confirm the predicted gene models.

Objective 4: Functional confirmation of *Pup1* and impact assessment

JIRCAS: *Studies on root growth/ physiology/ morphology as related to Pup1*. A first pilot study to evaluate the effect of the *Pup1* gene under combined stresses of P deficiency and water deficit indicated that the advantage of *Pup1* is more pronounced in drying soil and that this advantage was due to a root-growth effect. More detailed studies on the interaction of P deficiency, water deficit and root growth are planned for 2007.

IRRI: *Association of Pup1 and drought tolerance*
Recently, converging evidence suggest that *Pup1* might be associated with drought tolerance:
(i) *Pup1* overlaps with a major QTL (RM511) for yield under drought. Fine mapping of this QTL is currently ongoing at IRRI (Bernier et al., unpublished).
(ii) High representation of the *Pup1* locus in IRRI accessions scored as drought tolerant
(iii) *Pup1* phenotype does not show under irrigated field conditions and in liquid, hydroponics culture solution (Wissuwa and Ae, 2001; this study).
(iv) The *Pup1* phenotype did not show in an experiments conducted in Japan when soil was well watered (M. Wissuwa, personal communication, also see above)
An experiment is currently ongoing to further address this association using the *Pup1* NILs 14-4 and 14-6, as well as contrasting NILs for the drought QTL RM511 (data not shown). In order to assess interrelation between P-availability and drought, plants were grown at 100%, 85% and 60% field capacity in a P-deficient Philippine soil (from Kapatalan, Luzon) without and with addition of P-fertilizer (equivalent to 60 kg ha^{-1} P). Preliminary analyses of already available data indicate that

(i) The soil is generally suitable for P-deficiency screenings since differences in tiller number under +P and –P conditions are evident. However, under well watered conditions no differences between the contrasting *Pup1* NILs were observed, i.e. both reduced tiller number equally (+P: 24 tillers, -P: 18 tillers).
(ii) NIL14-6 (-*Pup1*) is delayed in tiller development under +P and well watered conditions, but eventually reaches the same tiller number. This delay might indicate that presence of Pup1 is beneficial even under +P conditions, or that the amount of P-fertilizer supplied was too low, and plants still experienced P-stress in the +P treatment. An experiment with different P-fertilizer applications (60-200 kg ha-1) is currently being set up to address this question. It is also possible that the additional introgressions present in NIL14-6 account for the observed difference (Figure 17).
(iii) Differences between the *Pup1* NILs were observed under water stress conditions. At 85% and 60% field capacity and +P conditions, NIL14-4 maintains a higher tiller number (22 tillers both treatments) compared to 20 tillers and 18 tillers, respectively, in NIL14-6. Under –P conditions, no significant differences between the NILs are observed (Figure 17).
This might indicate that the *Pup1* QTL confers drought tolerance independent of P supply. Similar results were obtained by M. Wissuwa.
(iv) No differences in plant height are currently obvious between the *Pup1* NILs in any treatment. Since the plants were only recently shifted from 16 h light to flower inducing (natural) short day conditions, differences in plant height between the NILs might show later.
(v) The RM511 drought QTL NILs (V3: + drought QTL, V4: – drought QTL) that were included in this experiment so far did not show the expected behavior since V4 outperforms V3 under stress

conditions (data not shown). The plants are currently being genotyped to validate the presence of the RM511 QTL.

Germplasm survey for phosphorus efficiency tolerance under lowland field conditions: A total of 75 rice accessions are currently being screened for performance under phosphorus deficiency. These lines include 11 P-efficient breeding lines identified before at IRRI (currently being genotyped for *Pup1*), 22 genotypes selected before for their drought tolerance at IRRI (genotyped for *Pup1* locus); 21 accessions consisting of pairs of IRRI lines that differ in specific *Pup1* alleles but presumably have similar genetic background (assumption based on accession designation); and 21 checks that had been phenotyped and genotyped for *Pup1*. Trials were set up in a P-deficient lowland field in Pangil (Luzon, Philippines) and a subset of the accessions was planted in an irrigated (demo) plot at IRRI. The demo plot contains Pangil soil that was transferred to IRRI some years ago. Experiments were set up in randomized complete block design with four replications and two P-fertilizer treatments (0 kg and 60 kg ha^{-1} P). Plots were otherwise fertilized with N, K and Zn. Tiller number, plant height and growth stages were recorded regularly for five plants in each replication. After harvest, dry weight, grain weight and spikelet fertility will be determined and samples will be analysed at least for total NPK content. The trial will be repeated during wet season 2007. The Pup1 NILs 14-4, 14-6 and 6-4 were not included in this screening because all accessions are photosensitive and do not perform under short day conditions. More details on this experiment will be given in the next report when the first data set is available.

Soil samples of both fields (Pangil and demoplot) will be taken and analysed for macro-elements (NPK) and zinc content at IRRI. Soil samples will also be taken from an area adjacent to the lowland field in Pangil in order to determine if this area would be suitable for phosphorus deficiency screenings under upland conditions. This experiment will help identify accessions that are potential varieties/donors with high P-uptake efficiency under low land conditions.

Objective 6: Build the capacity of participating NARES partners in MAB and other needed molecular techniques.

Three students are currently pursuing their PhD programmes under this project. One at Dhaka University, Bangladesh; one at Bogor University, Indonesia; and one at UC Davis. The fourth student at IRRI/UPLP dropped from the programme for personal reasons. In an affiliated GCP programme, M. Phil. student (Suhaila Rahman) is receiving training at UC Davis, California, USA for 9 months. One Ph.D. student received 7-day training at IARI, New Delhi, India. Same student will be spending 6 months at IRRI, training on MAS on an affiliated GCP programme (Habibul Bari Shajib); One more Ph.D. student will be receiving 2 week training in breeding methods at IRRI in May, 2007 (Rokeya Begum).

Tangible outputs delivered

ICABIOGRAD

1. Selected three Indonesian rice varieties as recipients of Pup-1 based on genotypic and phenotypic evaluation, and generated the first backcrosses for further advancement and analysis using MAS.
2. 10 breeding line from linked GCP-SP2 #42, "*Targeted discovery of superior disease QTL alleles in the maize and rice genomes"* reveal good information on rice blast resistance as well as against P- deficiency.

Dhaka University

1.Three varieties bred for the salinity affected areas in Satkhira for Boro (dry winter season) are being tested again in the saline zone in Satkhira for the 3rd consecutive year.
2. Three markers identified for introgression of *Saltol* into elite varieties by Student from BRRI (Sazzzadur Rahman) will be used in MAS programme starting August, 2007. Background and foreground marker information were provided by IRRI.

3. Identified popular varieties as recipients of *Saltol* , crosses made with *Saltol* donors and are being advanced for further analysis using MAS.

IRRI and UCD:

1. Good progress made in identifying tightly linked as well as flanking markers at the two loci, background markers were also identified and MAB system initiated to introgress the two QTLs into popular varieties
2. Twenty-two genes identified as potential candidates in *Pup1* region, 4 genes were short-listed for further analysis using complementation and RNAi

Deviations from the workplan

ICABIOGRAD: Because breeding material from IRRI to be tested in Indonesia was not yet available, the activity on "Impact assessment of *Pup1* in multiple-stress environments (saline/drought/P-deficient)" is yet to be conducted. This activity can be done in the future in the greenhouse. Field test can only be conducted in another wet season (with the assumption that the seed become available as soon as possible so that we have enough time for the seed increase).

Dhaka University: 1. Additional time would be required (no-cost extension) to complete MAB programme

Data availability

Most of the project data is still being compiled and undergoing extensive analysis, and this data will subsequently be made available.

Competitive Project #8: Targeted discovery of superior disease QTL alleles in maize and rice genomes

Principal Investigator:
Rebecca Nelson, Cornell University
Principal Collaborators:
Peter Balint-Kurti, NCSU
Darshan Brar, IRRI
Hei Leung, IRRI
Casiana Vera Cruz, IRRI
Masdiar Bustamam, ICABGRRD
James Gethi, KARI
Jedidah Danson, KARI
Jane Ininda, KARI
Jan Leach, CSU
Margaret Smith, Cornell University
Utut Suharsono , IPB

Mid-year report

Evidence for multiple disease resistance (MDR) at the gene level (Balint-Kurti and Wisser, NCSU; Kolkman and Nelson, CU)

- CU, NCSU: In the 2006 annual report for Project 8, we reported a significant correlation among levels of resistance to southern (Maydis) leaf blight (SLB) and gray leaf spot (GLS) in the Buckler association mapping population (n=308 lines, Flint-Garcia et al., 2005). We now add that our dataset for northern (turcicum) leaf blight (NLB) also correlated significantly with the SLB and GLS data for 250 lines from the same panel (Slide 2). Even accounting for population structure and kinship, the genetic correlations found imply that some genes contribute to resistance to multiple diseases.

- NCSU, CU: In an effort to identify genes conditioning quantitative resistance to single and multiple diseases, we will score the 5,000-line nested association mapping panel, in collaboration with the groups of E. Buckler and J. Holland, for SLB and NLB this field season. This will be our second year of data on this population.
- NCSU, CU: From a comparison of the datasets for SLB, GLS and NLB, several genotypes with MDR at the genotypes level were identified.

Genetic analysis of MDR (Gethi, KARI)

- KARI: Towards the development of heterogeneous inbred families and introgression lines, BC_3F_1 (87.5% of recurrent parent) lines were produced, and now enter the inbreeding phase through single-seed descent. By the end of this year, the material will have been inbred for at least two cycles. We can then start identifying lines possessing different introgressions from the donor genome.
- KARI: Towards dQTL mapping in materials of central interest to the KARI programme, the Gethi group advanced the F_1s obtained from the diallel study through selfing three plants per plot and obtained 42 different F_2 populations. After the field evaluation of the 42 single-cross hybrids under artificial and natural infestations at Kiboko, Kabete and Kakamega, six single-cross hybrids were selected based on their performance in three areas. These F_2 populations have been planted at Kakamega for characterisation of GLS in the F_2 populations. The plants' reactions to GLS will be evaluated and tissues will be harvested for genotyping with 30 SSR markers that are linked to known GLS QTL, and for which we have obtained marker data on the parents. In addition to this population, we will generate $F_{2:3}$ populations to allow for replication of the phenotyping activities at Kakamega.

Genetics of resistance to brown spot of rice (Banu, Brar, Leung, Vera Cruz, IRRI)

- IRRI: The traditional tropical japonica cultivar Dinorado (IRTP12568), which is resistant to brown spot, was crossed with the susceptible IR36. Seedlings of the F_1, F_2, F_3, BC_1F_1 and BC_2F_1 populations were screened with an aggressive *Bipolaris oryzae* isolate. Phenotypic segregation of 200 F_3 progenies showed that resistance to brown spot of rice is governed by two recessive genes. F_2 progenies were used as mapping populations using SSR markers. A parental survey using 160 SSR markers showed 92% polymorphism. Bulked segregant analysis was used to analyse 186 F_2 lines with 4 SSR markers from chromosome 12. Genes for resistance to brown spot are located between 8.7 and 18.2 MB in chromosome 12. The resistance was governed by two recessive genes tentatively designated as *bs1* and *bs2*. This is the first report on molecular mapping of genes for brown spot resistance located on chromosome 12 of rice.

Selection mapping of quantitative disease resistance in maize (Poland, Kolkman and Nelson, Cornell; Wisser and Balint-Kurti, NCSU [with Seth Murray, Cornell University])

- CU: Using selection mapping, we identified a locus under putative selection in the dQTL hotspot region at bin 2.06. When we tested for association of the selected locus with resistance to northern (turcicum) leaf blight (NLB) in greenhouse tests of an F_2 population derived from the recurrently selected population, the selected allele was found to be significantly associated with susceptibility rather than resistance to NLB. We hypothesised that susceptibility to NLB might be "trumped" by resistance to common rust, since that same region had been associated with rust QTL in previous studies. Phenotypic assays of the same F_2 population using the rust pathogen confirmed that the selected allele is indeed associated with resistance to rust in this population (Slide 4). In dicotyledonous systems, salicylic acid- and jasmonic acid-controlled disease resistance pathways have been shown, in some cases, to be antagonistic. Our previous rice dQTL synthesis provided evidence for antagonism at the resolution of chromosomal segments; co-localising dQTL for different diseases were identified where alleles from the same

parent provided a positive effect on disease development for one disease but a negative effect for another. Our recent results suggest that QTL in maize may also have opposite effects on different types of pathogens. This is relevant to our underlying hypothesis that dQTL are diverse in their mechanisms of action and their merits in crop protection. It reinforces our argument that to understand quantitative disease resistance, it is important to look at several pathogens with different pathogenesis strategies.

- CU: The effect of the dQTL allele at bin 2.06 was found to depend on the presence of another selected allele at bin 1.07. The epistatic interaction is illustrated in Slide 5.
- NCSU and CU: Simulation was used to provide insights into the process of selection. Results indicated that, when a locus is under selection, the best allele will not necessarily exhibit significant increases in frequency over the remaining alleles. Simulation analysis (using an additive effect model) showed that this is dependent on the relative allele effects, or the allele effect distribution. It was also found that the distance from the marker locus used to detect a gene under selection may be less important than the extent of LD in the base population used for subsequent selection. Our current interpretation for the application of selection mapping is that it is ideal to find multiple linked marker loci exhibiting deviations from genetic drift – that would represent strong evidence that one is close to the actual gene(s) under selection.

<u>Characterisation of near-isogenic lines (NILs) of maize differing in quantitative disease resistance (Zwonitzer and Balint-Kurti, NCSU: Chung and Nelson, CU)</u>

- NCSU: We now have a set of n=50 NILs in the B73 genetic background. Each line is >95% B73 plus one of six QTL derived from the elite resistance source NC250. These are now in field trials at NCSU.
- NCSU, CU: We have created several F_2 families derived from crosses between NILs shown to differ for disease resistance. These F_2 families are now being trialed to assess whether the presence/absence of resistance QTL can be reliably scored on a single plant basis and to confirm that the presumptive QTL really confers disease resistance.
- CU, NCSU: A set of chromosomal segment substitution lines (CSSLs) from Tx303 x B73 has been used for analyzing and dissecting quantitative resistance for NLB. Previous disease evaluation (Aurora, 2006) suggested that Tx303 alleles in bins 1.03-1.04, 1.07, 5.07-5.09, 6.05, 7.04, and 10.06-10.07 contribute resistance, and B73 alleles in bins 1.02, 1.04, 2.07-2.09, 3.01, 3.05, 4.01, 4.07, 5.04, 7.01, 7.05, 8.03, 8.08, 9.05, and 10.03 confer resistance to NLB. Linkage analysis in five F_2 populations (derived from the cross of selected CSSLs and B73) validated a weak QTL effect at umc1568 (141.8 cM) in bin 1.02 (greenhouse, 2007). The B73 allele at this locus was associated with delaying incubation period by 0.7 days ($P = 0.003$), as well as lowering lesion expansion by 0.47 mm per day ($P = 0.012$).
- CU: Slides 5, 6 and 7 report the results of histopathological analysis of CML52 (source of multiple-disease resistance used for extraction of several near-isogenic line pairs, as previously reported), Tx303 (donor for CSSL set), B73 (recurrent parent for CSSLs and other NILs), and TBBC3_42, a line near-isogenic to B73 carrying an introgression from Tx303 that increases disease susceptibility. Through microscopic investigation of initial plant-pathogen interactions on a selected CSSL (TBBC3_42) and B73, we determined this QTL has no effect on decreasing infection efficiency of *E. turcicum*, but can induce a larger necrotic area surrounding the infection site, and restrict the spread of hyphae growing into the xylem vessels. In the future, these differential phenotypes during pathogenesis will be examined on other selected CSSLs for QTL characterization.
- NCSU: In two cases, in bins 1.09/10 and 3.04, we have identified families with closely overlapping introgressed regions, which together cover the entire disease QTL (dQTL) region (e.g. Slide 8). These are currently being assessed in order to fine map the QTL.
- NCSU: Based on our maize dQTL synthesis a region in bin 2.06 was found to be of particular interest for MDR (and for reasons noted above as well). This region was identified as the most significant

dQTL hotspot with dQTL reported for eight different diseases. In a population derived from an NC250 x B73 cross, we have shown that the largest effect dQTL for both SLB and GLS resistance co-localized in bin 2.06 with resistance being derived from NC250 in each case. We have now identified several independently-derived ~97% B73 NILs with NC250-derived DNA introgressed in the bin 2.06 region. These NILs will be examined at field sites this summer for their resistance to NLB, SLB, and GLS.

Genomic analyses of rice oxalate oxidases (Carrillo, Leung, Vera Cruz, IRRI; Leach, CSU)

- IRRI: BLAST similarity search for the cupin protein, oxalate oxidase Y14203 from barley identified several orthologs from the rice genome. A PERL script designed to identify proteins sharing the same motifs as OsOxo identified 70 sequences that can be divided into two groups depending on the copy number of the conserved motifs of the cupin domain: monocupins and bicupins. There are 45 monocupins in the rice genome. These include oxalate oxidase and oxalate oxidase-like protein; the other 25 proteins contain two copies of the cupin domain and are called bicupins (Slide 9). The monocupins can be divided into six subfamilies – rhicadhesin-like proteins, oxalate oxidases, nectarin 1 precursor proteins, auxin-binding proteins, germin-like proteins and oxalate oxidase-like proteins
- IRRI: Of the six subfamilies, oxalate oxidases and oxalate oxidase-like proteins have been mapped to the rice genome and are associated with quantitative resistance to rice blast. The nectarin 1 precursor proteins are similar to the nectarin 1 proteins in tobacco, the majory nectary protein with a possible antimicrobial role in the nectar of tobacco flowers because of its superoxide dismutase activity. Oxalate oxidase and oxalate oxidase-like proteins were genetically mapped to dQTLs in chromosomes 3 and 8, respectively, and physical mapping on the rice genome also confirm the presence of these genes in the same region.
- IRRI: Oxalate oxidases form a unique subclass composed of four members - LOC_Os03g48750 (OsOxo1), LOC_Os03g48760 (OsOxo2), LOC_Os03g48770 (OsOxo3), and LOC_Os03g48780 (OsOxo4). These genes exhibit > 90% similarity at the nucleodtide and amino acid levels. All four genes are intronless, consistent with reports that stress responsive genes usually do not have introns. EST clones were identified only for OsOxo1, OsOxo3 and OsOxo4 while full-length cDNA clones are available only for OsOxo1 and OsOxo4. Expression analysis using resistant and susceptible Vandana x Moroberekan advance backcross lines reveal that only OsOxo4 is expressed during rice-*Magnaporthe grisea* interaction. This is consistent with EST data available and gene structure analysis.

Sequence variation in the oxalate oxidase locus (Carrillo, Reveche, Leung, Vera Cruz, IRRI; Leach, CSU)

- IRRI: We analysed the sequence variation of OsOxo1, OsOxo2, OsOxo3 and OsOxo4 from 62 rice cultivars belonging to six distinct rice subgroups - *indica*, *aus*, *aromatic*, *japonica*, *deepwater subtype 3,*and *deepwater subtype 4.* Analysis of the amino acid substitution rates for each of the OsOxo gene showed that they are undergoing purifying selection in contrast to resistance genes which are undergoing rapid evolution. While in the previous analysis that the four OsOxo genes are > 90% similar to each other, phylogeny of the four genes germplasm show that each OsOxo gene cluster with the same OsOxo gene from other germplasm sources. Within a cluster, genes from the same subtypes tended to cluster together, though this was not the rule as there were groupings of genes that from different subtypes (Slide 10).
- IRRI: We examined the sequence variation and promoter elements 1kb upstream of the genes. For OsOxo2, we observed a 450 bp Pong subclass transposon insertion for all members of the Aus subtype (Slide 11).
- IRRI: The OsOxo4 1kb upstream region also showed a 23 bp deletion in Vandana and analysis of this missing region revealed *cis*-elements related to root nodulation in bacteria-infected cells (Slide 12).

Identification of deleted regions in rice mutants exhibiting altered interactions with pathogens (Leach, Diaz, Bruce et al., CSU; Wisser, NCSU)

- CSU: Deletions were detected using commercially available Affymetrix arrays (details in next bullet). The analysis developed involves flagging of probe-sets on arrays that vary from wildtype with varying stringency. The flagged probesets are scrutinized for the number and location of probes within the probeset that are presumed deleted. Putative deletions could be rapidly confirmed using PCR. Because the Affymetrix probe-sets have been mapped to the rice genome, the hybridization results allow rapid localisation of deletions. Deleted regions (corresponding to groups of probe-sets) on a chromosome are revealed, allowing inferences to made about the relative size of deletions (Slide 13). The Leach lab optimised hybridisation and data analysis such that it is only necessary to use one array hybridisation per line, making the work more cost-effective. The availability of an allelic series of a mutation allows for the identification of a few genes as candidates for the phenotype (Slide 14)
- CSU: Eleven rice lines that harbor chemical- or physical-induced mutations were analysed, leading to the discovery of mutated regions in each. The best detection was found for those mutations induced by fast neutrons and gamma rays, because the deletions induced by these agents are relatively large. DEB-derived mutants can also be detected, but this requires closer scrutiny and depends on probe-set coverage per gene, the location of mutations within a gene, etc. The number of mutations predicted per line was not related to the type of mutagen. Mutations were detected on all chromosomes, as expected, but with unequal distribution across the chromosomes (unexpected). We would like to determine whether this is because of the annotation state used for the development of the Affymetrix array, or because mutants were pre-selected for disease resistance/susceptible phenotype. Some regions were deleted in several mutants (Slide 15). It is unclear whether this is because these regions are hotspots for deletions, whether selection favored mutations in specific genes, or whether this is attributable to probeset design in those regions.
- CSU: To discover mutations associated with dQTL, seven loss-of-resistance and one gain-of-resistance mutant were hybridised to the Affymetrix arrays. All of these exhibited deletions in regions on multiple chromosomes (Slide 16). Validation of randomly selected deleted probesets by PCR verified deletions in some cases, but was equivocal in others (Slide 17; note less intense bands). This may be because the lines are heterozygous for the mutation.
- CSU and IRRI: Next steps include validation of randomly flagged mutations in various lines (how reliable can we predict deletions at various stringencies?); validation of gene discovery (*spl1*); confirmation of the association of a deletion with a dQTL region; backcrossing lines to develop isogenic sets with deletions in particular regions (already started at IRRI); developing primers to track deletions in progeny; testing publicly-available (NSF) oligo spotted array for feasibility.
- IRRI: We will also apply the sliding window analysis of aggregated expression patterns (developed under the SP4 data analysis project) as a means to statistically identify contiguous regions with reduced hybridization signals (i.e. spanned by deletions). By applying several analytical "filters", we hope to reduce the number of false positives and increase the confidence of declaring deleted regions based on hybridization data.
- All: With the accumulation of results from QTL analysis (in maize and rice), deletion mapping, and NIL mapping, we can combine the datasets to develop a comparative map with common or distinct regions (within species and between species) that have significant effects on disease resistance in rice and maize.

Towards analysis of resistance to *Aspergillus flavus*

- CU: Because aflatoxicosis is an important problem for maize-dependent populations subject to drought conditions, we are seeking genes that contribute to per-se resistance to infection, disease development and aflatoxin accumulation. Towards this end, we are developing a quantitative PCR assay for *A. flavus*. We have achieved specific amplification of *A. flavus* and are fine-tuning the quantitative performance of the assay. We are developing crosses for genetic analysis of resistance to

A. flavus.

Development of pathogen collections for analysis of multiple-disease resistance (Hsieh and Nelson, CU)

- CU: Gray leaf spot (GLS) of maize is caused by *Cercospora zeae-maydis* and *C. zeina*, of putative North American and African origin respectively. To obtain representative strains of both species from NY for use in our experiments, we systematically collected 500 isolates from around an area where the two species had been detected in a sample of n=2. To date we have screened 77 monoconidial isolates using species-specific histone primer pairs and *Cercospora*-specific mating-type primer sets. We found both species in southern NY, often both at the same locality (about 60% were *C. zeae-maydis*). Both MAT1-1-1 and MAT1-2 alleles were identified for each species. We will conduct AFLP analysis and greenhouse experiments to further assess the genetic variation and the differences in pathogenicity among the New York isolates of the two species.
- IRRI: Towards understanding the population structure of *Bipolaris oryzae* causing brown spot of rice and selecting tester strains useful for screening resistance of varieties and breeding lines, we have systematically collected *Bipolaris oryzae* in a single field and in upland and rainfed areas in the Philippines. The genetic diversity of a collection of *Bipolaris oryzae* isolates from the Philippines was estimated with a Variable Number of Tandem Repeat (VNTR) marker. Several different lesion types were observed in the field, but isolates from each lesion type could produce a range of different lesion types when inoculated onto rice cultivar IR72, which is susceptible to *B. oryzae*. All the isolates originating from different lesion types also belonged to the same VNTR haplotype, supporting the view that the different lesion types were not a result of genetic differences in the fungus. Intensive sampling from a single field showed that the *B. oryzae* isolates had a continuous range of virulence to IR72. All those isolates had the same VNTR haplotype, except for two, each of which belonged in a separate haplotype. A collection of 325 isolates obtained from multiple locations on the islands of Luzon and Mindanao and from the Visayas islands could be divided into 50 VNTR haplotypes giving a somewhat high genetic diversity (H_T) value of 0.89. However, there were 3 or fewer isolates in 39 of the haplotypes, and 80% of the isolates belonged to only 8 haplotypes that had 10 to 71 isolates each, indicating a prevalence of clonality. No major relationship between haplotype and geographical location or host variety was observed. These results indicate that rice fields may contain *B. oryzae* populations with considerable genetic diversity, but the majority of infections arise from a much smaller subset of these that form clonal populations. It appears that most lesions arise from secondary inoculum that is produced asexually. The source of genetic variation resulting in the large number of haplotypes is not known as the sexual stage of *B. oryzae*, *Cochliobolus miyabeanus*, has not yet been reported in the Philippines.

Breeding for disease resistance (Gethi, KARI; Bustamam, Trijatmiko et al., Indonesia)

- KARI: We continue to further test the 30 hybrids obtained from the MDR panel tested in the past season (Slide 18). Some of these hybrids have been crossed to a third parent to generate three way crosses (TWCs). These have now been planted at seven locations (three at the coastal belt, one in western Kenya, three in the drylands of Katumani). Hybrids developed initially from the MDR panel with GLS resistance were evaluated for drought and there are indications to suggest this material may have some level of tolerance to water stress.
- KARI: From the diallel study conducted by the student from the University of Nairobi, the best sources of GLS resistance was identified as CML 373 and TZMi 711. Initially CML 373 was not in our MDR panel. In the initial screening, CML 373 and TZMi 711 had the highest level of resistance in the three regions where it was tested. The general and specific combining ability of the various parents with TZMi 711 and CML 373 being the best combiners. As expected, reciprocal crosses were not significant for grain yield.
- KARI: The best specific combining ability for GLS was found in the TZMi 711/TZMi 712, though these did not perform well in terms of grain yield.

- Indonesia: A total of 11 QTLs for rice blast resistance located on chromosomes 3, 4, 7, 8, 9, 11 were detected in a BC_2F_2 population derived from Way Rarem x Oryzica Llanos-5. The range of R^2 value of these QTLs was 10.7 – 45.1 %. The Oryzica Llanos-5 allele was favorable for 36% of the QTLs.
- Indonesia: Ten $BC_2F_{2:5}$ families of Way Rarem x Oryzica Llanos-5 were evaluated in a blast hotspot in Sitiung, Sumatera, Indonesia, along with genetic materials for phosphorus tolerance (under another GCP project). They performed well under a severe neck blast infection in comparison to the susceptible check. Surprisingly, Oryzica Llanos-5 was also infected (Slide 19).

Tangible outputs delivered
Publications:

Jines M, **Balint-Kurti P,** Robertson-Hoyt L, Molnar T, Holland J, Goodman M (2006) Mapping Resistance to Southern Rust In a Tropical By Temperate Maize Recombinant Inbred Topcross Population. Theoretical and Applied Genetics 114:659-667.

Balint-Kurti P.J, Zwonitzer J.C, **Wisser R.J,** Carson M.L, Oropeza-Rosas M, Holland J.B, X Szalma S.J. (2007) Precise mapping of quantitative trait loci for resistance to southern leaf blight, caused by *Cochliobolus heterostrophus* race O, and flowering time using advanced intercross maize lines. Genetics (In Press).

Gao X, ShimW-B, Göbel C, Kunze S, Feussner I, Meeley R, **Balint-Kurti P**, Michael Kolomiets (2007) Disruption of a maize 9-lipoxygenase results in increased resistance to fungal pathogens and reduced levels of contamination with mycotoxin fumonisin. Molecular Plant-Microbe Interactions (In Press)

Wisser R. J, Murray S. C, **Kolkman J. M**, Ceballos H., **Nelson R. J,** Selection mapping of loci underlying the response to artificial selection for quantitative disease resistance in a diverse maize population. In preparation for Genetics.

Advanced rice lines for field evaluation

Six Vandana-like lines with added blast resistance from Moroberekan ready for testing in upland sites in India. Vandana is a drought tolerant upland variety but deficient in disease resistance in the upland environment.
IR78221 19-6-56 (Chitinase, Oxalate Oxidase, Peroxidase, Oxalate oxidase-like protein)
IR78221 19-6-99 (Oxalate oxidase)
IR78221 19-6-7 (Oxalate oxidase, HSP90, Peroxidase)
IR78221 19-6-33 (Oxalate oxidase, Peroxidase, Oxalate oxidase-like proteins)
IR78221 19-6-90 (Oxalate oxidase, Peroxidase, Oxalate oxidase-like proteins)
IR78222-20-7-148 (Chitinase, Oxalate oxidase, HSP90, Aldose reductase, Peroxidase, Oxalate oxidase-like protein)

Deviations from the workplan
- Our interest in resistance to *A. flavus* is relatively new and was not explicity mentioned in the proposal.
- We would like to officially request a one-year, no-cost extension for 2008.

Data availability
- P. Balint-Kurti submitted a dataset on SLB QTL mapping in the IBM advanced intercross population to the GCP. Other related datasets will follow.
- On behalf of the project team, the project leader acknowledges our obligation to provide GCP with datasets that will enable others to utilise the valuable genetic resources that have been produced

through the project, as well as other datasets of potential utility to others. We intend to provide such datasets by the end of the project term (at the end of the no-cost extension period, if necessary).

Competitive Project #11: Functional genomics of cross-species resistance to fungal diseases in rice and wheat (Cereal Immunity)

Principal Investigator:
Jean-Benoit Morel, Agropolis-INRA
Collaborators:
D Tharreau, Agropolis-CIRAD
JL Nottéghem, Agropolis-AgroM
E Guiderdoni, Agropolis-CIRAD
M Ferreira, EMBRAPA Genetic Resources and Biotechnology
G de Capdeville, EMBRAPA Genetic Resources and Biotechnology
S Scagliusi, Postdoc - EMBRAPA Wheat
A Bonato, EMBRAPA Wheat
Y Mehta, IAPAR – a state research institute collaborating with EMBRAPA
J Maciel, EMBRAPA Wheat
P Scheeren, EMBRAPA Wheat
A Mehta, EMBRAPA Genetic Resources and Biotechnology
MS Chaves, EMBRAPA Wheat
S Brammer, EMBRAPA Wheat
P. Ronald, UC Davis
KH Jung, UC Davis
R Sing, CIMMYT
M William, CIMMYT
S Kikuchi, NIAS
K Satoh, NIAS
L Boyd, JIC
H Tufan, JIC
C Feuillet, INRA (Clermont)
P Sourdille, INRA (Clermont)

Mid-year report
Activity 1 Phenotypic and cytological characterisation of non-host interactions in rice and wheat with *M. grisea* and *P. triticina.*
1A Non-host cytology in wheat(JIC)
Past achievements: cytology data suggest that cell death is not massively involved in wheat non-host resistance to *M. grisea* at 24 hpi (JIC). Early time point examination suggests that even earlier events may be critical (EMBRAPA).
The cytological analysis of *Magnaporthe grisea* isolates was further investigated (24 hpi previously analysed); the wheat cultivar Renan was used and the 32 hours after inoculation (hai) is now complete (Figure 1).

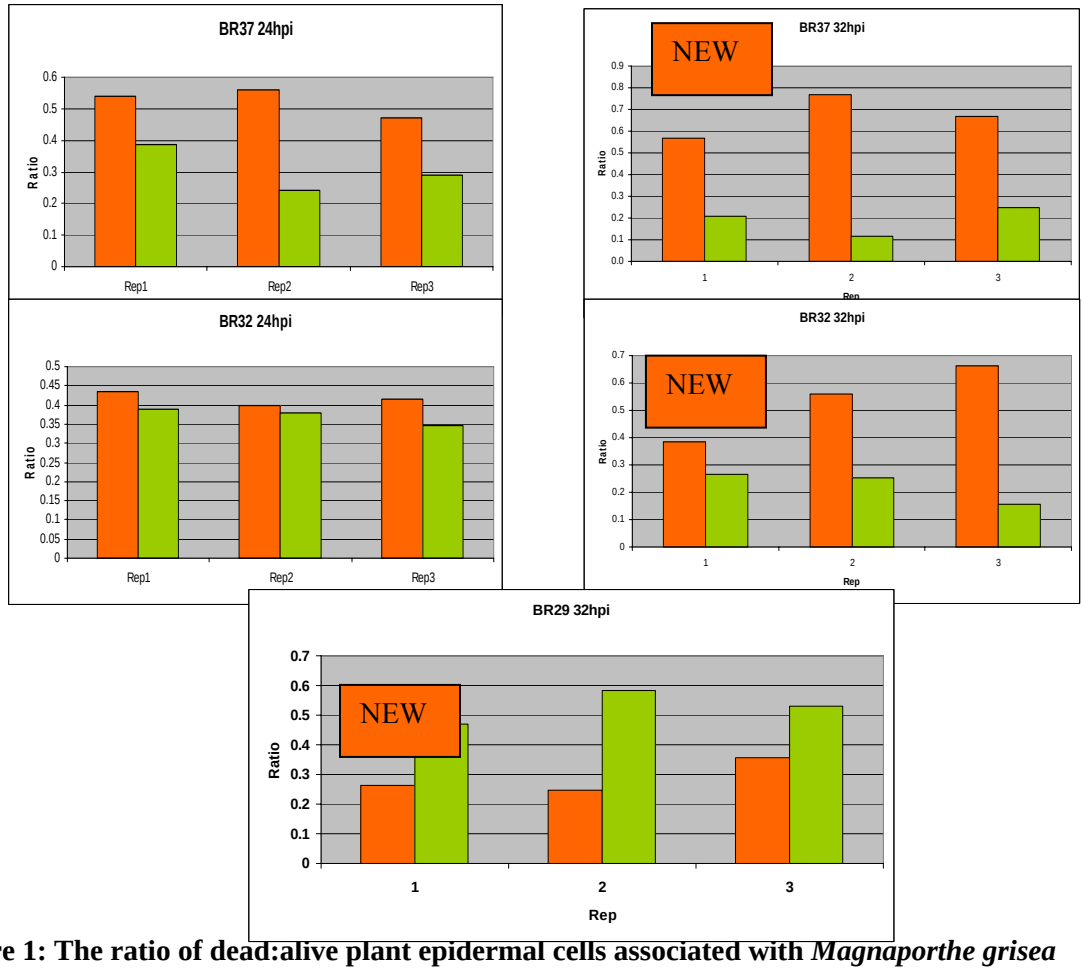


Figure 1: The ratio of dead:alive plant epidermal cells associated with *Magnaporthe grisea* appressoria.
Appressoria associated with dead epidermal cells.
Appresoria associated with live epidermal cells.
BR37 is a wheat isolate virulent on cv. Renan, BR32 a wheat isolate avirulent on Renan and BR29 is a non-host isolate from *Digitaria*.

As infection hyphae were seen to form in epidermal cells in all host (avirulent and virulent) and non-host (Digitaria and rice isolates) interactions, fungal appressorial development was measured in relation to plant cell death. The proportion of appressoria associated with plant cell death was calculated as the number of appressoria divided by the number of dead (or alive) epidermal cells associated with those appressoria (Figure 1).

With the host-avirulent isolate (BR32) at 24 hai there were more live epidermal cells associated with appressoria compared with BR37 (host-virulent isolate). With both host isolates the proportion of appressoria associated with dead cells increased between 24 and 32 hai. In the non-host interactions fewer appressoria were seen associated with epidermal cell death than in the host interactions, further suggesting previous data (24 hpi) that non-host resistance does not primarily involve a plant cell-death response.

Because of the poor visual resolution obtained using light microscopy and chlorazol black staining it was decided to transform the four *M. grisea* isolates with a GFP marker gene. This has now been done for all

four isolates (BR32, BR37, BR29 and FR13), and transformants are being selected that show good GFP expression using confocal microscopy, while maintaining the same macroscopic disease phenotypes on cv Renan.

The interaction between Renan and the host and non-host isolates will now be examined using confocal microscopy. This will enable the expression of non-host resistance genes (identified from the micro array and quantitative RT-PCR analyses) to be directly related to the pathogen's development and the plant's resistance phenotypes (including plant cell death).

Additional data are being produced by EMBRAPA.

Activity 2: Transcriptome analysis of non-host interaction in rice and wheat
Activity 2A) Microarray hybridsations in japonica rice (UCD, NIAS, Agropolis)
Past achievements: 488 and 492 potential non-host genes identified in IR64 indica (NIAS) cultivar and Nipponbare japonica (UCD) cultivar respectively.
EMBRAPA is organizing microarray experiments with UCD to get further insights in early time points.
Activity 2B) Microarray hybridisations in wheat (JIC, Agropolis)
Past achievements: Wheat non-host Affymetrix hybridization completed. First statistical analysis identified a set of 33 potential wheat non-host genes (JIC). Cross species analysis identified 5 genes potentially common to rice and wheat non-host resistance.
The 24 hour time samples from the following four treatments and all three biological replicates were sent to The National Arabidopsis Stock Centre (NASC) to be hybridised against the Affimetrix wheat array:

(1) Mock
(2) Wheat inoculated with host virulent isolate, BR37
(3) Wheat inoculated with host avirulent isolate, BR32
(4) Wheat inoculated with non-host *Digitaria* isolate, BR29.

These data sets were previously analysed using the dChip and TMEV (www.tm4.org/mev.html) and Significance Analysis of Microarrays (SAM) analysis, performed in TMEV. Subsequently Dr Andreas Magusin (Micro Array Support Service at JIC) has used R-stat software to repeat the analysis. From these second set of analyses, a consensus set of 38 wheat candidate non-host genes has been selected (Table 2), 19 of which were not previously found by SAM.
Conversely, from the 33 previously found by SAM, 12 were not found in this second set of analyses (Table 1).

Table 1. Gene that were only found with SAM analysis

Description	Accession	in second analysis
Weakly similar to putative plasma membrane integral protein	CA722332	no
cold acclimation protein WCOR615	U73217.1	no
Weakly similar to putative glucan endo-1,3-beta-D-glucosidase	CA648355	no
Affy probe, sequence not in Unigene	AJ609766	no
Moderately similar to expressed protein [Arabidopsis thaliana]	CA613215	no
Affy probe, sequence not in Unigene	CA682916	no
Transcribed locus	BJ274310	no
Strongly similar to protein kinase family protein [Arabidopsis thaliana]	CA659767	no
Affy probe, sequence not in Unigene	CA679566	no
Affy probe, sequence not in Unigene	CA743463	no
Transcribed locus	CA632620	no
Affy probe, sequence not in Unigene	CA636241	no

Some of these genes will still be investigated and two of them (CA682916 and CA743463) are already being mapped in wheat (see Activity 4B) since putative rice orthologues (Os08g42000, Os06g06420 and Os09g38910 respectively) were also found induced in non-host resistance in rice (see Activity 3A; Table 3).

Table 2. Upregulated genes identified with the R-stat software.

Probe	Description	Accession	in 1st analysis
Ta.207.1.S1_at	S276 protein	AF031194.1	yes
Ta.7814.1.S1_at	moderately similar to NP_909926.1 putative nodule-specific protein [Oryza sativa (japonica cultivar-group)]	BJ250219	no
Ta.1989.1.S1_a_at	moderately similar to XP_462851.1 B1146F03.16 [Oryza sativa(japonica cultivar-group)	CA674427	no
TaAffx.8993.1.S1_at	weakly similar to NP_195056.2 disease resistance protein (CC-NBS-LRR class), putative [Arabidopsis thalian]	CA616697	no
Ta.6629.1.S1_s_at	moderately similar to XP_479099.1 acidic nuclear phosphoprotein-like protein [Oryza sativa (japonica cultivar-grou	CK162414	no
Ta.16790.1.A1_at	weakly similar to NP_468092.1 unknown protein [Oryza sativa (japonica cultivar-group)]	CA612656	no
Ta.8468.1.S1_x_at	Transcribed locus	CK216337	yes
Ta.28395.1.A1_at	moderately similar to NP_850169.1 expressed protein [Arabidopsis thaliana]	CA658354	no
Ta.7918.1.S1_at	moderately similar to XP_478523.1 unknown protein [Oryza sativa (japonica cultivar-group)]	BQ150910	yes
Ta.18525.1.S1_at	weakly similar to NP_921257.1 hypothetical protein [Oryza sativa (japonica cultivar-group)]	CA638213	no
Ta.23109.1.S1_at	Transcribed locus	CA732234	yes
Ta.9102.2.S1_x_at	weakly similar to NP_566528.1 expressed protein [Arabidopsis thaliana]	CA600902	yes
TaAffx.139741.1.S1_at	gb: CA626528 /DB_XREF=gi: 25204824 /DB_XREF=wl1n.pk0143.e4/CLONE=wl1n.pk0143.e4 /TID=TaAffx.1	CA626528	no
Ta.144.1.S1_at	wpk4 protein kinase	AB011670.1	yes
Ta.28553.14.A1_x_at	Ubiquitin-like mRNA, partial sequence	BJ273013	yes
Ta.7979.1.A1_at	moderately similar to NP_916206.1 OsPK7 [Oryza sativa (japonica cultivar-group)]	CK157049	yes
TaAffx.29251.1.S1_at	weakly similar to XP_479470.1 hypothetical protein [Oryza sativa (japonica cultivar-group)]	CA647097	yes
Ta.9102.1.S1_at	weakly similar to NP_566528.1 expressed protein [Arabidopsis thaliana]	CA738151	yes
Ta.25272.1.S1_at	weakly similar to NP_176059.1 GDSL-motif lipase/hydrolase family protein [Arabidopsis thaliana]	CA680512	no
Ta.18914.1.A1_a_at	moderately similar NP_568855.1 F-box family protein / LOV keich protein 1 (LKP1) [Arabidopsis thaliana]	CD875211	no
Ta.12137.1.A1_at	moderately similar to XP_483259.1 Cyt-P450 monooxygenase [Oryza sativa (japonica cultivar-group)]	BQ171664	yes
TaAffx.111818.1.S1_at	gb: CA632910 /DB_XREF=gi: 25211206 /DB_XREF=wle1n.pk0061.h7/CLONE=wle1n.pk0061.h7 /TID=TaAffx	CA632910	no
Ta.18914.2.S1_x_at	moderately similar NP_568855.1 F-box family protein / LOV keich protein 1 (LKP1) [Arabidopsis thaliana]	CA644075	no
TaAffx.79046.2.S1_at	gb: CA723390 /DB_XREF=gi: 25445183 /DB_XREF=wdr.1f.pk002.l 11/CLONE=wdr.1f.pk002.l 11 /TID=TaAffx	CA723390	yes
TaAffx.112912.1.S1_at	Transcribed locus	CA619860	no
TaAffx.25555.1.S1_at	weakly similar to NP_912423.1 Putative NAM (no apical maeristem) protein [Oryza sativa(japonica cultivar-group)	AJ614172	yes
Ta.11076.1.A1_x_at	weakly similar to XP_469705.1 putative inmediate-early salicylate-induced glucosyltransferase [Oryza sativa	BQ578840	no
TaAffx.25555.3.S1_at	gb: CA634931 /DB_XREF=gi: 25213227 /DB_XREF=wle1n.pk0088.d7/CLONE=wle1n.pk0088.d7 /TID=TaAffx	CA634931	yes
Ta.485.1.A1_at	strongly similar to XP_464447.1 putative Lipoxygenase2.3, chloroplast percursor [Oryza sativa (japonica cultivar-	CD454354	yes
TaAffx.38635.1.S1_at	Transcribed locus	BJ228820	yes
Ta.27725.1.S1_at	plasma membrane protein	AB030210.1	yes
Ta.1057.1.A1_x_at	weakly similar to NP_195738.1 proton-dependent oligopeptide transport (POT) family protein [Arabidopsis thaliana	BQ171746	yes
Ta.13336.1.S1_at	apetala2 protein	CK212277	yes
Ta.2492.1.A1_at	weakly similar to NP_567080.1 expressed protein [Arabidopsis thaliana]	BQ902022	no
TaAffx.23898.1.S1_at	sequence(s) not in UniGene	CA728707	no
Ta.1940.1.S1_x_at	moderately similar to NP_181018.1 homeobox-leucine zipper transcription factor (HB-14) [Arabidopsis thaliana]	BG604505	no
TaAffx.23898.1.S1_at	weakly similar to NP_922867.1 unknown protein [Oryza sativa (japonica cultivar-group)]	CA697658	no
Ta.23424.3.S1_at	moderately similar to XP_462851.1 B1146F03.16 [Oryza sativa(japonica cultivar-group)	CA632371	no

From the 5 genes common to rice and wheat, 3 were shown to be non-host genes in rice (Os08g42000, Os06g42000 and Os09g38910) (Activity 3A)
This activity is now completed.

Activity 3: validation of non-host genes in rice and wheat
Activity 3A) Validation of microarray data using Quantitative RT-PCR
Rice genes (Agropolis)
Past achievements: The specific expression of 10 rice non-host genes was confirmed by QRT-PCR (Agropolis).
The initial list of 10 genes was increased to 24 genes (Table 3). Some genes could still be induced in host resistance ("NHG?" genes). Five genes seem to be specific to non-host resistance to Puccinia ("Pu"), 5 to non-host resistance to *Magnaporthe* ("Mg") and 14 are common to both ("Mg & Pu").
A subset of 17 of these genes has been selected for further expression analysis for Activity 5 (see below).

Table 3. The rice genes that are up regulated in non-host and host resistance

TIGR	annot	microarray source	insertion line	insertion CIRAD	insertion UCD a	prog RNAi	diagnostic	NH type	origin
Os12g04740	expressed protein	rice	ASGG02 in intron?exon sp	yes	no		NHG?	Pu	IR64
Os01g66980	expressed protein/chaperone?	rice	AT in gene; other Ats in int	yes	yes	yes	NHG?	Pu	IR64
Os04g15690	DSBA-like thioredoxin domain containing protein, expr	rice	T-DNA RMD in intron	na	no	no	NHG	Pu	NB
Os05g46530	pectinesterase inhibitor domain containing protein, exp	rice	AENH06 in promoter	yes	na		NHG	Pu	IR64
Os07g30760	UDP-glucoronosyl and UDP-glucosyl transferase famil	rice	none	na	na	yes	NHG	Pu	IR64
Os04g10160	Cytochrome P450 CYP99A1, putative, expressed	rice	none	na	na	yes	NHG?	Mg & Pu	IR64
Os06g36210	Amino acid carrier, putative, expressed	rice	AT in promoter; other in int	na	no	no	NHG?	Mg & Pu	IR64
Os02g51930	Cytokinin-O-glucosyltransferase 2, putative, expressed	rice	AT all over	na	yes		NHG?	Mg & Pu	IR64
Os06g44010	WRKY2 protein, putative, expressed	rice	AJBA03 in 3'; AT at start	yes	yes	yes	NHG?	Mg & Pu	IR64
Os09g38910	wall-associated kinase-like 1	wheat	AQOC02	yes	yes	no	NHG?	Mg & Pu	IR64
Os07g47990	Peroxidase 2 precursor, putative, expressed	rice	none	na	na	no	NHG	Mg & Pu	NB
Os03g64260	AP2 domain containing protein, expressed	rice	AT in gene	na	yes		NHG	Mg & Pu	IR64
Os11g02370	nonspecific lipid-transfer protein 2 precursor (ltp 2)	rice	AT in prom; T-DNA RMD	na	yes	yes	NHG	Mg & Pu	IR64
Os01g60020	No apical meristem protein, expressed	rice	3 Ats in gene	na	yes	yes	NHG	Mg & Pu	IR64
Os08g34330	Strictosidine synthase family protein, expressed	rice	AT in promoter	na	yes		NHG	Mg & Pu	IR64
Os10g36160	nonspecific lipid-transfer protein precursor, putative, ex	rice	none	na	na	yes	NHG	Mg & Pu	IR64
Os01g68750	Adaptin N terminal region family protein, expressed	rice	AT in promoter	na	yes		NHG	Mg & Pu	IR64
Os06g37150	L-ascorbate oxidase, putative, expressed	rice	ATs in gene	na	yes	yes	NHG	Mg & Pu	IR64
Os08g42000	Nuclear transport factor 2	wheat	AWFB03	not ordere	no	no	NHG	Mg & Pu	NB
Os09g25070	WRKY DNA binding domain containing protein, expres	rice	none	na	na	yes	NHG?	Mg	IR64
Os10g39150	Thylakoid membrane phosphoprotein 14 kDa	rice	AT and Ds in intron	na	no	yes	NHG?	Mg	IR64
Os10g28120	Chitinase 1 precursor, putative, expressed	rice	Ds in promoter	na	no	no	NHG	Mg	NB
Os02g54590	serine threonine kinase 1, putative, expressed	rice	ALDG02; Ats and Ds	yes	yes		NHG	Mg	IR64
Os06g06420	lactoylglutathione lyase family protein	wheat	ATVC01, ACSG05	yes	no	no	NHG	Mg	NB
Os10g35810	uncharacterized low-complexity proteins, putative, exp	wheat		no	not ord	no	defense		NB
Os03g52390	type II proteinase inhibitor family protein, expressed	rice		no	no		defense		IR64
Os05g19910	Transferase family protein, expressed	rice		no	yes		defense		IR64
Os01g68650	plant-specific domain TIGR01615 family protein, expre	rice		no	no		defense		IR64
Os07g48020	Peroxidase 2 precursor, putative, expressed	rice	yes but…	no	yes		defense		NB
Os05g46020	OsWRKY7 - Superfamily of rice TFs having WRKY an	rice		no	yes		defense		NB
Os12g36880	major pollen allergen Bet v 1-D/H, putative, expressed	rice		yes	yes		defense		IR64
Os02g10120	lipoxygenase 2.3, chloroplast precursor, putative, expr	wheat	AIGG04	yes	no	no	defense		IR64
Os02g09510	limonoid UDP-glucosyltransferase, putative, expressed	rice		no	no		defense		IR64
Os06g08060	leucoanthocyanidin dioxygenase, putative, expressed	rice		no	yes		defense		IR64
Os04g43650	L-allo-threonine aldolase, putative, expressed	rice		no	yes		defense		IR64
Os01g51060	IAA-amino acid hydrolase ILR1-like 4 precursor, putati	rice		yes	yes		defense		IR64
Os06g04020	histone H1, putative, expressed	rice		no	no		defense		NB
Os01g71340	glucan endo-1,3-beta-glucosidase, acidic isoform prec	rice		yes	yes		defense		NB
Os10g40934	flavonol synthase/flavanone 3-hydroxylase, putative, ex	rice		no	yes		defense		NB
Os05g44060	expressed protein	rice		no	yes		defense		IR64
Os12g38230	expressed protein	rice		no	no		defense		IR64
Os01g72080	EF hand family protein, expressed	rice	yes but…	no	yes		defense		NB
Os02g36840	cytokinin-O-glucosyltransferase 2, putative, expressed	rice		no	yes		defense		IR64
Os11g05380	cytochrome P450 94A2, putative, expressed	rice		yes	no		defense		IR64
Os03g27080	calmodulin-binding transcription activator 2, putative, e	rice		no	yes		defense		NB
Os09g29710	blight-associated protein p12 precursor, putative, expr	rice		no	no		defense		IR64
Os12g26290	alpha-DOX2, putative, expressed	rice		no	yes		defense		NB

Microarray source indicates whether the genes were initially identified in rice or wheat. Diagnostic indicates whether the corresponding gene was found to be non-host specific ("NHG" and "NHG?") or more generally induced during infection ("defense").

This activity is now completed.

Wheat genes (JIC)

The expression levels of selected candidate non-host resistance genes from Table 2 are being measured in Renan leaf tissue using a quantitative RT-PCR approach. Leaf tissue was collected from the original *M. grisea* inoculation expt. used for the micro array samples. Time course samples were taken at 4, 7, 9, 24, 32, 48 and 72 hpi. RNA has been extracted from all wheat-*M. grisea* interactions and time points. The level of expression of target non-host genes in these samples is being measured.

Activity 3B) In vivo validation of non-host genes using rice mutants (Agropolis, UCD)

Past achievements: Insertion lines corresponding to rice and wheat non-host genes were identified in Agropolis and UCD mutant collections (see Table 3).

Seven insertion lines corresponding to 6 selected genes from Table 3 were sown (Table 4). DNAs were extracted and the genotype at the insertion locus was determined using PCR primers flanking the

insertion. The gene corresponding to the AIGG04 line (Os02g10120) was lately found involved in host resistance and may not be further studied.

Table 4. Rice insertion lines genotyped

Line name	Gene	Homozygous	Wt	Heterozygous	?	Total
AENH06	Os05g46530	0	7	5	7	19
ASGG02	Os12g04740	3	1	7	0	11
ALDG02	Os02g54590	1	6	13	0	20
ASQG05	Os01g66980	1	3	2	2	8
ACSG05	Os06g06420	no data	no data	no data	no data	20
ATVC01		1 ?	4	0	0	5
AIGG04	Os02g10120	1	1	8	1	11

Seeds have been amplified for all lines and phenotyping will begin soon (Agropolis).
Other insertion lines (UCD) and RNAi lines (Agropolis) for other genes targeted are underway.

Activity 4: non-host gene mapping in rice and wheat

Activity 4A) Rice non-host genes mapping in rice (Agropolis)

Past achievements: rice non-host genes from activity 2A were mapped to the rice genome and locally anchored to Agropolis private Genome Browser.
Files will be transferred to GCP database as soon as the transfer template will be made available.

Activity 4B) Generation and validation of wheat and rice NHG/HG functional markers (EMBRAPA, Clermont, CYMMIT, JIC)

Past achievements: 5 genes common to rice and wheat non-host resistance were first selected for mapping in wheat.
Homologues of five candidate non-host genes identified in the wheat micro array analysis were identified as candidate non-host genes in the rice micro array analysis (Table 5). These genes are now the principal targets for mapping in wheat. INRA (Clermont) is designing polymorphic markers for these candidate non-host genes.

The procedure is divided in three steps:

1- Design consensus sequence primer pairs (amplification of the 3 homoeologous copies) to amplify intron as a source of SNP polymorphisms.
2- Clone and sequence amplicons in the 12 considered population parents.
3- Design specific primer to perform AS-PCR (allele specific PCR) as a chip and efficient SNP detection method especially for haplotyping (large number of accessions, limited to 48 lines) and more precisely genetic mapping.

Table 5: ID of the five genes selected for genetic mapping

Original wheat sequence	Best blast rice sequence	sequence used for primer design (best blast mutual hit)
BJ273013	Os01g22490 [nt]	Os05g06770 [nt]
CA743463	Os09g38910*	Os09g3800 [nt]
CA600902	Os03g38950 [nt]	
CD454354	Os02g10120**	
CA682916	Os10g35810**	
	Os08g42000*	
	Os06g06420*	

*: validated as non-host specific in rice; **: non-host and host induced in rice; nt: not tested.

The *in silico* analysis of the wheat orthologous sequence for all the 5 rice gene accessions has been performed. On the basis of the 5 wheat/rice sequence alignments, consensus primers have been selected for intron amplifications. Step one is considered to be finished. Step 2 is under progress to select homoeologous-specific sequences. Genotyping will be performed on 12 wheat varieties: Renan, Recital, Lemhi, Chinese 166, Synthetic, Opata, Frontana, Inia 66, Fukuokomughi, Oligoculm, Br18, Chinese Spring. Primer sequences will be delivered to the project partners.
INRA (Clermont) has been provided with genomic DNA from the cvs Chinese 166 and Lemhi (JIC).
More genes will be selected on the basis of Activity 3A (wheat genes validation by QRT-PCR).

Activity 4C) Wheat non-host genes mapping in wheat (EMBRAPA, Clermont, CYMMIT, JIC)
Genetic mapping for the polymorphic genes from Activity 4B will be performed at INRA in Renan x Recital and Synthetic x Opata populations.

Activity 4D) Genotyping of a wheat mapping population phenotyped for non-host resistance to M. grisea (EMBRAPA, Clermont)
The population selected for resistance-genes mapping purposes is one from EMBRAPA partner and consists in a doubled haploid population derived from the cross between lines BR18 and Frontana. The INRA (Clermont) partner has received the DNAs from the parents as well as for the population (EMBRAPA). The screening process for polymorphism detection started in February. Overall, 538 SSRs have been tested on the two parents and 263 (49%) have sown polymorphism. Among these, a set of 145 has been chosen in order to evenly cover the whole genome (Table 6). Between four and nine markers have been selected for each chromosome, the chromosomes from homoeologous 7 being the best covered (25 markers) while the homoeologous group 4 is the worst represented with 17 markers. Similar numbers of markers were selected for each of the three genomes, respectively 54, 47 and 44 for the A, B and D genomes. The genotyping of the population for the elaboration of the map will start in June 2007 (as part of Activity 5).

Table 6. Results of the polymorphism screening test on the BR18 and Frontana parents.

Genome		Chromosomes							Total
		1	2	3	4	5	6	7	
A	Polymorphic	12	10	18	13	6	13	20	92
	Selected	8	7	9	7	6	8	9	54
B	Polymorphic	12	13	13	6	9	10	25	88
	Selected	6	9	6	5	6	6	9	47
D	Polymorphic	7	12	11	7	13	16	17	83
	Selected	4	6	7	5	8	7	7	44
Total	Polymorphic	34	42	48	30	31	44	68	263
	Selected	18	22	22	17	20	21	25	145

4E) Non-host phenotyping of a wheat mapping population (EMBRAPA)

This activity is in progress at EMBRAPA.

4F) Haplotyping of rice and wheat germplasm collections for allelic diversity (CIMMYT, EMBRAPA, Agropolis)
A list of 50 wheat germplasms from Cymmit and 50 germplasms from EMBRAPA has been established. CIMMYT is transferring seeds to EMBRAPA. This set of lines included lines with varying levels of resistance to leaf and yellow rust. The material was planted under field conditions and disease infections were done with the help of the spreader rows which is the standard practice in CIMMYT's wheat breeding nurseries. DNAs from wheat Frontana, Inia66, Fukokumughi, Oligoculm, Opata-85 and W7984 (Synthetic) were sent to INRA Clermont.
For rice, a list of 25 rice genotypes, selected from the MiniGB available at Agropolis, has been established. Seeds will be transferred to EMBRAPA.

Activity 5) Enhance NARS capacity in functional genomics and molecular marker technology
Ana Lidia Bonato is presently doing a short-term training at Agropolis (Mid-March-End of May 2007). She is participating in the genotyping and phenotyping of rice insertion lines. She is also involved in QRT-PCR validation of non-host genes.
She will do a short-term training period at INRA-Clermont (June-August 2007) to participate in activities 4.

Tangible outputs delivered so far

- 488 rice IR64 potential non-host genes
- 492 rice Nipponbare potential non-host genes
- 38 wheat non-host genes
- 24 rice non-host genes validated by QRT-PCR
- 23 rice genes established as defense genes
- Homozygous rice mutant lines for 4 non-host genes and one defense gene
- 145 SSRs polymorphic between wheat BR18 and Frontana

Deviations from the workplan
No major deviation.

Data availability
a. Wheat transcriptome data will be available after QRT-PCR validation.
Wheat SSRs will be made available to the community.
b. Rice microarray data will be transferred as soon as the templates for" Gene expression data" are available.

Competitive Project #15: Determination of a common genetic basis for tissue growth rate under water-limited conditions across plant organs and genomes

Principal Investigator:
François Tardieu, INRA
Collaborators:
C. Welcker (breeder, Co-PI), INRA
F. Tardieu (ecophysiologist, PI), INRA
O. Turc (ecophysiologist), INRA

B. Parent (PhD), INRA
G. Davenport (computational biology, co-PI), CIMMYT
Y. Xu, (maize molecular breeder), CIMMYT
J.L Araus (maize physiologist), CIMMYT
M. Reynolds (wheat physiologist), CIMMYT
C. Bencivenni (PhD), CIMMYT
R. Serraj (rice physiologist, Co-PI), IRRI
J. Bennett (molecular biologist), IRRI
J. Cairns (physiologist), IRRI
R. Bruskiewich (bioinformatics), IRRI
R. Mauleon, IRRI
A. Hund (geneticist, Co-PI), ETH
P. Stamp (physiologist), ETH
M. Liedgens, ETH
N. Pa-In (PhD), ETH
P. Lessard, Biogemma
Peter Langridge (geneticist, co-PI), ACPFG
T. Schnurbusch, ACPFG
U. Baumann, ACPFG
A. Schreiber (Bioinformatics), ACPFG
B.M. Prasanna, ICAR
J.Gethi, KARI

Mid-year report

I Rice activities

1.1. Genetic analyses *(IRRI)*

QTLs associated with leaf morphogenesis (under drought stress) and expansion (well-watered conditions) were identified in previous field experiment using the Vandana/ Moroberekan advanced backcross population. To confirm these results and compare them with simultaneous phenotyping under water-deficit and well-watered conditions, two field experiments have been conducted in the dry season (2007) using both the Vandana/ Moroberekan BC population and Apo/IR72 RIL population. A progressive water deficit was applied during vegetative stage by withholding irrigation; leaf biomass, number of leaves, specific leaf area and biomass partitioning were measured destructively throughout the stress period. In addition to the destructive plant sampling, non-destructive measurements of leaf elongation rates have been carried out on single leaves under well-watered and drought conditions in the two populations. Experiments are still ongoing and data being analysed for the identification of putative QTLs for LER under drought and G x E analysis.

1.2. Modelling plant growth responses to water deficit *(INRA and IRRI)*

1.2.1 The bases of the model were established for two genotypes (INRA)

- The time course of the elongation rate of leaves on each position of the stem was recorded at a 15 minute time definition in two genotypes (Azucena and Nipponbare). This experiment was analysed jointly with 3 other experiments in the growth chamber and the greenhouse. A short plateau of leaf elongation rate (3 days) was observed, then the elongation rate declined for 3-4 d. The duration of each phase was stable in thermal time. The maximum leaf elongation rate increased with leaf position on the stem until leaf 7, while leaves 7-8-9 had similar leaf elongation rates.
- The response to temperature was clearly non-linear in the range 13-27°C, but was common to several experiments in the greenhouse and the growth chamber. This result casts doubt on the current calculation of thermal time, based on the existence of a linear relationhip. It may explain inconsistent data in the literature about the threshold temperature for calculation of thermal time. An alternative calculation has been proposed.

- A special protocol was designed for studying the effect of evaporative demand in spite of the non-stable leaf elongation rate. A common response was found for 3 experiments, and was common to genotypes Nipponbare and Azucena.
- The same protocol was used for studying the effect of predawn leaf water potential. Again, a common response was observed for 3 experiments.

1.2.2 A series of genotypes is now under study (INRA and IRRI)

- A series of experiments in the greenhouse and in the growth chamber has been carried out *(INRA)* with contrasting genotypes combining the subspecies (*Indica, Japonica, Glaberima*) and the lowland/upland character (APO, Vandana, Moroberekan, IR64, Azucena, CG14, Nipponbare). The objective was to identify differences in the maintenances of growth rate, photosynthesis, stomatal conductance and water flux under water deficit or high evaporative demand. The first series of experiments has just been completed. A second series aims to study in more detail the differences in expansion and cell cycle between genotypes.
- Another greenhouse experiment has been carried out *(IRRI)* with a subset of contrasting genotypes (APO, Vandana, Moroberekan, IR64, IR72) to examine the effect of a larger range of water deficit intensities. Four levels of water deficits were applied (FTSW 0.8, 0.6, 0.4 and 0.2) in addition to a well-watered control. Target thresholds of soil moisture were reached after 7 days and plants were maintained at these levels for a further 7 days until re-watering. Leaf morphogenesis and leaf elongation were measured daily on all plants during drought stress and during a recovery period, to establish the relationship of leaf elongation rate to the fraction of transpirable soil water (FTSW). Relations between FTSW and predawn leaf water potential are established to allow common analysis of the series of experiments in IRRI and INRA.

2.3. Large scale genomic approaches for candidate gene identification *(IRRI)*

An experiment has been designed for transcript profiling of leaf elongation zone under controlled water deficits, to identify putative candidate genes associated with LER under drought. Vandana, Moroberekan and two Vandana/Moroberekan BC_3 lines with ±QTL on chromosome 4 (and over 80% isogenic) are grown under controlled conditions in the greenhouse. Two treatments with different levels of water deficits an same stress duration will be applied. Water deficit will be initiated when the 4^{th} leaf begins to emerge. The aim is to reach target thresholds of soil moisture at the same time so plants are sampled at the same age. An Affymetric array will be used to compare gene expression between stressed and control plants at different soil moisture levels. Candidate genes will be selected from microarray analysis and QTL studies and tested using QPCR to provide an extensive study of a set of genes showing differential expression in contrasting genotypes for growth maintenance under water stress.

2.4. Focussed genomic approaches for candidate gene identification *(IRRI)*

A preliminary QPCR study of selected candidate genes in five contrasting accessions (Apo, IR64, IR72 and IR71525-19-1-1) showed large changes in expression levels of candidate genes related to leaf elongation. Plants were subjected to a progressive water deficits and the elongating zone of the leaf sampled at three levels of stress (FTSW: 0.6, 0.4 and 0.1.). Control pots were simultaneously sampled. Four expansins and five other candidate genes related to drought response were selected from RT-PCR results for QPCR. Significant differences in the expression patterns of several expansins were detected between cultivars and soil moisture levels. This work was conducted with Dr. Dongcheng Liu from the Chinese Academy of Sciences and analysis is still underway. Dr Liu will rejoin this project later this year on a GCP fellowship to initiate a comparative study of drought response kinetics in leaves and roots, simultaneously comparing growth responses and gene expression patterns in leaves and roots under drought. Extending the study of growth response in rice to roots will further contribute to the comparison of the genetic basis of drought responses of tissue expansion across organs and species.

II Maize activities.

2.1. Genetic analyses *(CIMMYT, KARI, IARI, ETH, INRA)*

2.1.1 Field studies

The greenhouse studies of the response of leaf growth to water deficit are now finished (*Welcker et al. 2007, J. Exp Bot 58, 339*-). The main work is now to analyse jointly field and greenhouse experiments, in order to relate the QTLs of responses of leaf growth to soil water deficit and evaporative demand to QTLs of leaf area and yield in the field.

- Three experiments have already been carried in CIMMYT out with the 220 lines of the mapping population P1 P2, the same as in the greenhouse experiment. Results are under analysis at INRA.
- Three new experiments will take place in June and October *(CIMMYT, KARI, IARI)*, with the same mapping population, in dry areas of Kenya and India, in addition to an experiment in the wet season in Mexico, aimed at obtaining leaves growing in low evaporative demand. The whole set of experiments will be studied (i) with a classical QTL analysis on leaf area, yield components and final yield, (ii) by combining phenotypic data with a model under development (see "model").

2.1.2. QTLs of root growth and response to water stress.

Experiments with the same P1 P2 mapping population have been carried out in 2005 - 2006 (see annual report 2006). The data of the multiple experiments with root growth under favourable and stressing conditions are under analysis.

2.2. Modeling plant growth responses to water deficit *(INRA)*

2.2.1 Whole plant modeling (INRA, U. Queensland, in the project Whole Plant Modeling)

This activity belongs to the commissioned project Whole Plant Modelling. It allows calculation of leaf area, from meteorological data and from QTLs of response of leaf growth to temperature, evaporative demand and water deficit. A minimum set of data on leaf development are required for the considered genotypes. In this way, the P1xP2 population studied both in greenhouse and field studies can be modelled, with a model tested in the field experiments. This will be the base for "virtual genotypes experiments" because the plants carrying any combination of alleles of P1 or P2 (virtual or existing) can be tested in any climatic scenario. Outputs of simulations will be compared to field experiments, and will serve to extrapolate the results of the latters.

2.2.2 The work on silk growth is continuing (INRA)

The main determinants of silk growth have been established in 2005-2006. This study showed, overall, that growths of silk and leaves share common mechanisms of control under water deficit, with the predominance of hydraulic and developmental processes, and a likely low contribution of carbon metabolism. Common response curves could be drawn for the responses of silk and leaf growths to water deficit, opening the way to a joint model.

current period was dedicated to the comparison of 8 lines of the P1P2 population. It was shown that the ranks of maintenances of growth under water deficit were similar for silks and leaves.

2.3. Large scale genomic approaches for candidate gene identification *(CIMMYT)*

Samples of roots, leaves and silks have been collected in the experiments described above. Microrrays analysis will be performed in order to identify groups of genes whose the expression pattern is correlated with the plant growth maintenance. Gene expression patterns will be compared between water deficit and control treatments, in the growing zones of three organs (leaf, silk and root), along 4 contrasting selected lines (P1, P2, RIL151 and RIL245).

mRNA from tissues of treated and control plants, grown in controlled conditions (greenhouse and growth chamber), have to be extracted, labeled and hybridized to microarrays.

The mRNA extraction should be done by the end of June 2007 and microarrays results by the end July 2007.

The array design will include technical replicates (dye swap, same sample twice and with different dye) and biological replicates (experiment repetitions in well-watered and stress conditions).

The experiment design has been drawn according to show a comparison between genotypes in a same treatment and a comparison between treatments for a same genotype (see attached).
The total of slides in this particular case would be 48, which include 96 tissue samples. Further, we would be interested in the comparison inside a same genotype for different tissues.

III Wheat activities

3.1 Genetic analyses *(ACPFG, CIMMYT)*

The wheat work has focussed on three lines; the drought-tolerant lines Excalibur and RAC875, and the intolerant line Kukri.

Mapping populations of 250 and 350 doubled haploids have been generated for the crosses RAC875/Kukri and Excalibur/Kukri. Molecular maps have been constructed for both populations. All lines were grown at three sites in Southern Australia over the 2006 season and experienced severe drought stress. At the Bolloroo site average yields were below 0.5t/ha.

During the 2006/2007 season the lines were also grown at Obregon, Mexico under drought and well watered conditions. The key tissue expansion trait being measured at these sites was peduncle length. The Obregon data from the RAC875/Kukri population have been collected but not yet analysed. All lines have been provided to ICARDA and are currently being multiplied for field trials in the 2007/2008 season at Tel Hadya and Breda.

To close gaps in the Kukri/RAC875 linkage map, 72 SSR markers were added to the map. The computer programme RECORD was used to find best marker orders from 21 linkage groups. The linkage groups were sorted by graphical genotyping in MS Excel. The ordering of markers with RECORD was repeated twice for each linkage group. Between each two marker orders, singletons and other potential errors in the marker segregation data were identified by visual inspection of graphical genotypes.

The phenotypic data from the field were analysed for their spatial variation, and preliminary results of QTL analysis show that flowering time is a major driver for many important traits such as yield in Kukri/RAC875 population; flowering time spanned about 5 weeks. In order to adjust data which were influenced by flowering time, two approaches were applied. Firstly, divide the population in two sub population and perform QTL analyses on separate sub-population. Secondly, adjust data for heading time by fitting the best regression model.

3.3. Large scale genomic approaches for candidate gene identification *(ACPFG)*

3.3.1. RNA extraction

All tissue for the wheat drought stress has been harvested and homogenized ready for RNA extraction and purification.We have established the best method for extracting RNA from the different tissues of interest (head, leaf, stem), the extractions are underway and for the leaf samples will be completed within the next 2-3 weeks. We are expecting results that will determine the best DNase treatment method later this week.

The last month has been spent finalising the ordering the first batch of 150 microarray slides and required equipment and optimising protocols with aim to begin the actual microarray work as scheduled for mid June, when the microarray lab at La Trobe University in Melbourne becomes available for our use.

3.3.2 Comparative Genomics

EST Clustering : The initial step involves finding the right parameters for EST clustering. This is especially important for wheat. We have used the ESTs generated from 3 barley cultivars (Morex, Haruna Nijo, Barke) as a model for wheat ESTs. We are performing several rounds of EST clustering with varying clustering stringencies in order to decide on the appropriate parameters to be used for the wheat sequences.

b) Identification of orthologs : This aspect is required to be able to determine which genes are true orthogs between wheat, rice and maize. The pipeline for the identification of orthologs is underway. A tool for evaluating orthologs has been developed and tested from comparing wheat and barley.

IV Training and capacity building

- Dr Somayanda Impa from the UAS Bangalore, India joined the project as a postdoctoral research fellow in December 2006. Students from the University of Bangalore and Bangladesh are also being recruited to work in the project. Dr Dongcheng Liu from the Institute of Biology and Developmental Biology, Chinese Academy of Sciences, China joined the project in February 2007 for 3 weeks to contribute in gene expression studies and the real-time PCR analysis. Ranee Christine Mabesa from the University of the Philippines also joined the project in mid January and is currently working towards her MS thesis.
- Dr Gethi (KARI) has applied for a training grant the SP5, with a positive answer so far.
- A thesis of a Irakian scientist has been accepted in INRA, the thesis of a Thai scientist is nearly finished.

Tangible outputs delivered
QTLs of the P1P2 population for leaf growth and ASI
QTLs of leaf length in maize, wheat and rice for well watered and water deficient plants
Models of silk and leaf growth for maize
Elements of models for rice leaf growth

Deviations from the workplan
The workplan has been re-established in December 2006. There is no deviation from this revised plan.

Competitive Project #16: Isolation and characterisation of aluminum tolerance genes in the cereals: An integrated functional genomic, molecular genetic and physiological analysis

Principal Investigator:
Leon V. Kochian, U.S. Plant, Soil and Nutrition Laboratory, USDA/ARS and Cornell University
Collaborators:
Jurandir Magalhaes, EMBRAPA Maize and Sorghum
Claudia Guimarães, EMBRAPA Maize and Sorghum
Vera Alves, EMBRAPA Maize and Sorghum
Newton Carneiro, EMBRAPA Maize and Sorghum
Robert Schaffert, EMBRAPA Maize and Sorghum
Sandra Brammer, EMBRAPA Wheat
Pericles Neves, EMBRAPA Rice and Beans
Rosangela Bevitori, EMBRAPA Rice and Beans
Samuel Gudu, Moi University
Owen Hoekenga, U.S. Plant, Soil and Nutrition Laboratory, USDA/ARS and Cornell University
Ed Buckler, U.S. Plant, Soil and Nutrition Laboratory, USDA/ARS and Cornell University

Mid-year report

We have made considerable progress, particularly on sorghum and maize Al tolerance, but also are now making progress on Al tolerance in rice and other cereal species. Because of the 1 page limit, we have summarised progress in bullet format, and can provide more details about any specific bullet upon request.

Sorghum Al Tolerance:

- Verified that gene responsible for the major sorghum Al tolerance locus, *Alt$_{SB}$*, confers Al tolerance. This is based on analysis of homozygous transgenic T3 Arabidopsis lines expressing *Alt$_{SB}$*, which exhibited large increases in Al tolerance.
- These same lines also showed a large increase in Al-activated root citrate exudation, verifying the gene, which is a member of the MATE family of membrane transporters, is a citric acid efflux transporter.
- Transgenic T1 lines of Bobwhite wheat expressing *Alt$_{SB}$* were significantly more Al tolerant than wild type Bobwhite wheat lines, showing the possible utility of using this gene for improving cereal Al tolerance in general.
- Developed a peptide antibody for the Alt$_{SB}$ and verified it specifically cross-reacts with Alt$_{SB}$. Have used this antibody to show that the Al induction of Al tolerance over a 6 day exposure of sorghum roots to Al correlates with a concommitant increase in Alt$_{SB}$ protein (and *Alt$_{SB}$* transcript abundance).
- We have stably transformed an Al sensitive sorghum genotype with the wheat Al tolerance gene, *ALMT1*. Based on preliminary analysis of T1 sorghum lines (which are still segregating for the transgene), we see a large increase in sorghum Al tolerance based on expression of the *ALMT1* transgene.
- Produced T1 transgenic progeny of the Al sensitive sorghum line, CMSXS102B, transformed with the wheat Al tolerance gene, *ALMT-1*, under the control of the ubiquitin promoter. The T1 progeny segregated in a 3 (presence of the transgene) to 1 (absence of the transgene) ratio and homozygous T2 lines fixed for a single copy of the transgene were generated.
- Assessed the Al tolerance of the T1 sorghum progeny expressing the wheat *ALMT1* via determination of inhibition of root growth in nutrient solution containing {27} μM Al^{3+} and observed a bimodal frequency distribution that fit a 3 (Al tolerant) to 1 (Al sensitive) segregation ratio ($P[\chi 2 \geq 0.098]=0.32$) (see slide #3 is attached Powerpoint file). Al tolerant progeny were dramatically more tolerant than CMSXS102B and the sensitive progeny, and the segregation for Al tolerance and sensitivity correlated perfectly with the segregation for presence and absence of the transgene.
- In Caniato et al. (2007) we reported the Al tolerance results for 3 near-isogenic lines in which the *Alt$_{SB}$* allele from different Al tolerant sources were introgressed into the same Al sensitive genetic background. We have now expanded our collection to a total of 9 NILs including those carrying the elite alleles from SC566 and SC283. This will help us confirm the identity of the best *Alt$_{SB}$* alleles in our set.
- Using the sorghum panel described in Caniato et al. (2007), which harbor an allelic series at *Alt$_{SB}$*, we found that root growth inhibition in nutrient solution, *Alt$_{SB}$* expression and organic acid exudation are all highly correlated. This indicates that differences in gene expression constitute the basis of allelic variation at *Alt$_{SB}$* by modulating rates of citrate release into the rhizosphere.
- Initiated the development of near-isogenic hybrids that will be used for detailed studies on the regulation of Alt$_{SB}$.

Maize Al Tolerance:

- Mapped 9 maize MATE genes that are homologs of sorghum *Alt$_{SB}$* in the IBM (North American) population of maize RILs, and identified one of the maize MATEs as a putative maize Al tolerance gene via joint association-linkage analysis.

- Mapped 3 maize MATE genes that are homologs of sorghum Alt_{SB} in the Cateto Al 236/67 x L53 RIL population (South American).
- Identified a second maize Alt_{SB} homolog (MATE gene) as a putative Al tolerance gene via: 1) gene expression profiling (maize microarray); 2) mapping to a major Al tolerance QTL previously identified on the EMBRAPA (South American) maize RIL population; and 3) verified the expression data (increased expression under both +/- Al conditions in the tolerant parent) via RT-PCR analysis.
- Mapped 15 members of the ALMT (Al-activated malate transporter) family in maize on the IBM population. Subsequent association analysis of the maize ALMT family members identified one maize ALMT as a putative maize Al tolerance gene.
- Mapped 6 members of the ALMT (Al-activated malate transporter) family in the the Cateto Al 236/67 x L53 RIL population (South American). One of the members mapped on chromosome 8 in the vicinity of a newly identified Al tolerance QTL.
- Characterised the transport properties (via electrophysiology) of the maize ALMT gene that is most similar in sequence to the wheat Al tolerance gene, *ALMT1*. Found that despite its sequence similarity to wheat *ALMT1*, the maize homolog is not involved in Al tolerance. We found, via electrophysiological analysis of the maize *ALMT* expressed in *Xenopus* oocytes, that unlike wheat ALMT1, the maize ALMT does not transport organic acids (malate or citrate) and is not activated by Al. The maize ALMT is an anion transporter that moves mineral anions (NO_3^-, Cl^-, SO_4^{2-}) into and out of root cells.
- Phenotyped the entire 5000 member maize Nested Associated Mapping (NAM) panel for Al tolerance in hydroponic culture. The NAM panel, developed by Dr. Buckler's lab as part of the NSF funded maize diversity project, consists of 25 linkage mapping populations from carefully chosen parents that together capture roughly 80% of the common SNP diversity in maize. Each of the populations was constructed in a reference design using B73 as a common mother; each population has 200 RILs that share a common SNP-based molecular genetic map. Together, the 5000 RILs represent a meta-RI population that is a very powerful resource that unifies QTL mapping across the maize community, as well as a resource for which almost every trait will show substantial variation in multiple populations. We are now waiting for the subsequent genotyping of the NAM panel to analyse the NAM panel for novel Al tolerance QTL.

Al Tolerance in Other Cereals:

- Have developed new digital imaging tools for phenotyping rice Al tolerance and am currently using these tools to conduct QTL mapping of rice Al tolerance.
- A recurrent selection population in rice has been selected for uniform flowering time and is now ready for association studies for Al tolerance. This population is now being genotyped with 48 SSRs evenly distributed across the 12 rice chromosomes (4 SSRs per chromosome) to assess population structure and relatedness.
- A wheat RIL population, Toropi x Anahuac, is being multiplied for Al tolerance phenotyping in the field and in nutrient solution.
- We have initiated degenerate primer design for the Alt_{SB} gene and have assembled a Triticeae panel to test those markers.
- A set of 44 sorghum accessions including Al tolerant inbred lines, near-isogenic lines carrying different Alt_{SB} alleles, and lines in which Al tolerance and striga resistance have been pyramided have been sent to Dr. Sam Gudu (Moi University) for field testing in Kenya. Included are A (male sterile) and R (fertility restorer lines), which can be used to generate Al tolerant hybrids. This material is now in transit to Kenya.

Tangible outputs delivered

- Identification of a major sorghum Al tolerance gene
- Elite Al tolerant sorghum hybrids developed from the breeding programme

- Phenotypic (Al tolerance) data for Buckler Maize Association Panel
- Phenotypic (Al tolerance) data for CIRAD sorghum Association Panel
- Phenotypic (Al tolerance) data for the nested association mapping (NAM) panel of maize inbreds
- Identification of 6 preliminary candidate genes for maize Al tolerance via association analysis and verification of two of these by linkage analysis
- Maize microarray data for root tips under +/-Al conditions
- Phenotypic (Al tolerance) data for Buckler Maize Association Panel
- Manuscripts in Press, Submitted, or in Preparation:

Magalhaes JM, Liu J, Guimares CT, Lana UGP, Alves VM, Wang Y-H, Schaffert RE, Hoekenga OA, Shaff JE, Pineros MA, Klein PE, and LV Kochian (2007) A member of the multidrug and toxic compound extrusion 'MATE' family is a major gene that confers aluminum tolerance in sorghum. Nature Genetics (In Press).

Pineros MA, Cancado GMA, Maron LG, Lyi SM, Menossi M, Kochian LV (2007) Not all ALMT1-type transporters mediate aluminum-activated organic acid responses: The case of *ZmALMT1*. Plant Journal (under review by journal).

Lana UGP, Alves VMC, Guimaraes CT, Liu J, Kochian LV, Schaffert RE, Shaff, JE, Magalhaes JV. Factors regulating Alt_{SB} expression are key components of Al tolerance in sorghum (in preparation).

Alves VMC, Magalhaes JV, Guimaraes CT, Shaff JE, Kochian LV. Genetic dissection of Al tolerance mechanisms in a maize recombinant inbred line population (in preparation).

Hoekenga OA, Mason PA, Shaff JE, Buckler ES, Kochian LV. Interdisciplinary analysis of aluminum tolerance responses in the Intermated B73 x Mo17 recombinant inbred population of maize. (In preparation).

Lyi SM, Krill AM, den Bakker M, Doyle JF, Buckler ES, Kochian LV, Hoekenga OA. Joint linkage-association analysis of aluminum tolerance in maize: dissection of a membrane transporter family and identification of a single tolerance gene. (in preparation)

Deviations from the workplan

The subcontracts for EMBRAPA Maize and Sorghum, EMBRAPA Wheat, EMBRAPA Rice and Bean, and Moi University were all finalized and started in September of 2006. Thus we have asked for and granted a second year long no-cost extension. This report covers progress for the first 8 months of the first year of funding for these subcontracts, plus the second year of research for the Kochian lab progress.

Data availability

1. Data set for maize root tip gene expression profiling in response to Al toxicity for an Al tolerant (Cateto 100-6) and Al sensitive (L53) maize inbred. These data were generated using the maize long oligonucleotide arrays generated by the NSFPG-funded Maize Oligonucleotide Array Project (NSF DBI #0321663; www.maizearray.org). These arrays contain 70-mer oligonucleotides representing >30,000 identifiable unique maize genes. Our experiment was part of the beta testing of the maize oligonucleotide arrays prior to their public release. In this project, gene expression was profiled in a time-course experiment involving Al treatment at 0 (control), 2, 6, and 24 hours. Maize root tip samples were collected from an Al tolerant maize genotype (C100-6) and an Al sensitive genotype (L53) at each specific time point and were contrasted with the previous and subsequent time points. Additionally, the two genotypes were contrasted directly for each time point. It should be noted that the entire gene expression data set for this experiment is available at the Zea mays Microarray Gene Expression Database (ZEAMAGE) on the Maize Oligonucleotide Array Project website

(www.maizearray.org). It is also interesting to note that the maize root tip gene expression data we deposited in the ZEAMAGE data base makes up approximately 38% of all the gene expression data contained there.

2. Phenotypic data for Al tolerance determined by measuring root growth in the presence and absence of Al for the EMBRAPA sorghum association panel of 47 sorghum inbred lines.

3. Phenotypic data which is root growth in the presence and absence of Al for the 288 maize inbred line association panel assembled by Dr. Ed Buckler.

4. Phenotypic data which is root growth in the presence and absence of Al for the 5000 RILs in the Nested Associated Mapping (NAM) panel of maize RILs. The NAM panel, developed by Dr. Buckler's lab as part of the NSF funded maize diversity project, consists of 25 linkage mapping populations from carefully chosen parents that together capture roughly 80% of the common SNP diversity in maize. Each of the populations were constructed in a reference design using B73 as a common mother; each population has 200 RILs that share a common SNP-based molecular genetic map. Together, the 5000 RILs represent a meta-RI population that is a very powerful resource that unifies QTL mapping across the maize community, as well as a resource for which almost every trait will show substantial variation in multiple populations.

5 .Map locations on the maize genetic/physical map for maize homologs of the *ALMT* family that harbors wheat and Arabidopsis Al tolerance genes, and also maize members of the Alt_{SB}-like family that in sorghum includes a major sorghum Al tolerance gene. This information will be made available upon acceptance of manuscripts on these topics for publication.

Subprogramme 3: Trait capture for crop improvement

Competitive Project #3: Identifying the physiological and genetic traits that make cassava one of the most drought-tolerant crops

Principal Investigator:
Alfredo Augusto Cunha Alves, EMBRAPA
Collaborators:
Martin Fregene (Co-PI), CIAT
Morag Ferguson (Co-PI), IITA
Tim Setter (Co-PI), Cornell University
Hernán Ceballos, CIAT
Geoffrey Mkamilo, ARI
Edward Kanju, IITA
Cecil Osei, SARI
Antonio Souza, EMBRAPA/CNPMF
Miguel Angel Dita Rodríguez, EMBRAPA/CNPMF
Alineaurea Silva, EMBRAPA/CPATSA

Mid-year report

Activities accomplished:

1. Micropropagation and hardening of drought tolerant cassava contrasting varieties at EMBRAPA/CNPMF to produce planting material (stakes) for the experiments.
2. Harvest of the field screening trial, planted in May/2006 in Petrolina (Brazil), in which 60 genotypes were tested under irrigation and water deficit condition. The data evaluated are in process of final analysis.
3. Intermediate harvests in the field experiment planted in Dec/2006 in Petrolina (Brazil), with 24 drought tolerant contrasting cassava varieties under control and water deficit. Seven blocks were planted, with five plants/plot under randomised completed block design. Three blocks were harvested: 1) After 3 months under irrigation (Control 0); 2) After 4,5 months under irrigation (Control 1); and 3) After 3 months under irrigation followed by 1,5 month without irrigation. The following data were collected: plant height and branch lenght, leaf area of the top fully expanded leaf, leaf retention, leaf conductance, leaf temperature, air temperature, air relative humidity, radiation, soil moisture at 20 and 40 cm, samples of leaf disks (for ABA and carbohydrates), samples of stems (for carbohydrates), fresh weights of shoots and storage roots, and number of storage roots.
4. Establishment (Jan/2007) of a field experiment in Petrolina (Brazil), with 10 drought tolerant contrasting cassava varieties planted in a randomised completed block design, using five plants/plot, three plots/block, and four blocks (1 and 2 for water deficit and 3 and 4 for controls) with a total of 60 plants/variety, spaced 1.5 x 1.0 m. Three months after planting, the artificial irrigation was interrupted in block 1 and 2 and the first growth parameters were measured in all blocks (plant establishment, plant height and branch lenght, leaf area of the top fully expanded leaf, and leaf retention).
5. Establishment of a field trial at Hombolo, site in Dodoma (Tanzania) in Dec/2006, using the same cassava contrasting varieties as last experiment. In 2007 cropping season, data on foliar symptoms of major cassava diseases (cassava mosaic disease and cassavabrown streak disease) and insect pest (cassava green mites and cassava mealybug) were collected at three months after planting whereas data on plant establishment was collected at one month after planting.

6. Final harvest from previous studies of 15 cassava genotypes in two field locations in Colombia and in a root-containerized study at CIAT. Samples of leaf disks and stem plugs from these studies, which were sampled periodically during their growth last year, are currently being analysed for sugars, starch, and ABA (scheduled completion is August 2007).
7. A follow up experiment was started (Jan/2007) at CIAT, Colombia, with 45 cassava genotypes. Twenty-one stakes of each genotype were plant in 3 kg plastic bag for sprout phase to assure uniformity between and within genotypes. After one month, the best plants were selected and randomly assigned to experimental pots (well watered and water stress treatments) that will be grown under field condition and evaluated for physiological and growth traits to assess their drought tolerance, and traits underlying drought tolerance.
8. Six mapping populations, 3F1s and 3S1s, have been developed (at CIAT, Colombia) for drought tolerance in cassava. The populations are being established in vitro and also micro-propagated in preparation for shipment to partners in Brazil, Ghana, Tanzania, and Nigeria. A total of 200 genotypes and 5-10 plants per genotype will be shipped to partners.
9. Ten in-vitro plantlets each of twenty-two cassava varieties with stay-green characteristics have been sent from IITA-Headquarters to Tanzania for clonal evaluation against Tanzanian local varieties. These plantlets are being hardened at Mikocheni Agriculture Research Institute, before transplanting for multiplication purposes. The same genotypes are ready for shipment to Ghana. Unfortunately the collaborator in Ghana, Cecil Osei, is on sabbatical leave, and has not handed the project over to anyone in his absence. We have therefore been unable to get an permit to import the plantlets. A further 32 genotypes are being propagated at IITA-Ibadan, and will be transferred to Tanzania (and Ghana) within the next two months.
10. Application of Dr. Geoffrey Mkamilo (collaborator from ARI, Tanzania) to a GCP Travel Grant. He was selected to spend 1 month in Brazil, Aug/2007, participating in hands-on training experiences on cassava phenotyping at EMBRAPA/CNPMF, supervised by Dr. Alfredo Alves.
11. A Kenyan PhD student, Rosemary Mutegi, on a DAAD Scholarship has been appointed by IITA to work on the project.
12. A training workshop on cassava phenotyping methods for African partners is being organised in Tanzania, in collaboration with all project partners. The training is scheduled to take place in July 2007.
13. Equipments (porometer and soil moisture meter) have been sourced for African NARS partners

Tangible outputs delivered

- Identification of drought tolerant contrasting cassava varieties
- Production of plant material of the drought tolerant contrasting varieties via micropropagation and hardening in the field
- Establishment and harvest of field trials to evaluate several drought tolerant contrasting cassava varieties submitted to different water status treatments in Brazil, Colombia, and Tanzania.
- Establishment of crossing blocks and development of mapping populations for drought tolerance in cassava.
- Accomplishement of parental screen on the parents of IITA mapping populations for drought tolerance in cassava

Deviations from the workplan

The following constraints have been faced in this project:

- Delay in receiving and establishing in vitro plants imported from Colombia (CIAT) to Brazil, Tanzania, and Ghana

- Importing cassava germplasm into USA is a problem not yet solved, although performing studies at CIAT has circumvented this problem.
- Damages occured in the imported *in vitro* plants during shipment from Colombia (CIAT) to Brazil (EMBRAPA)\
- Delay for acquisition of equipments for evaluation of physiological parameters in the field (mainly for African partners)
- Lack of planting material (stakes) of some selected contrasting genotypes to be used in the field trials.
- Problems to micropropagate some genotypes due to the varietal differences in the genotype-media interaction.
- The occurrence of a strong flooding in Colombia during Jun/2006 caused irreversible damage in the crossing block at CIAT and in the field trials planted in Colombia. A new crossing block was established in Palmira (Jul/2006), in which the flowering of the parentals has not been as expected. So, the workplans of the EMBRAPA, African NARS (inTanzania and Ghana), and IITA, have been delayed as they have not yet received the segregating populations developed at CIAT.
- Delay in training African partners in the evaluation of physiological traits related to drought tolerance. The training of African partners on phenotyping has been scheduled for Jul/2007.

These constraints have caused significant delay of the original workplan and most of them can be minimised by an extension of the project's timetable (one year no-cost extension).

Data availability
The project has generated data of field and screenhouse evaluation of agronomic and physiological traits in several constrasting cassava genotypes under water deficit and control conditions, as well as data from analysis of SSR markers. These data have not been published into local or public GCP database. The data can be posted into a GCP database as soon as we know the appropriate database for this kind of data, the procedures to organise them, and how to feed and update the database.

Competitive Project #5: Unlocking the genetic diversity in peanut's wild relatives with genomic and genetic tools

Principal Investigator:
José Valls, EMBRAPA
Co-Principal Investigators:
David Bertioli (Co-PI), Universidade Católica de Brasília
Wellington Martins (Collaborator), Universidade Católica de Goiás
Ousmane Ndoye (Co-PI), Centre d'étude régional pour l'amélioration de l'adaptation á la sécheresse
Vincent Vadez (Co-PI), ICRISAT
Udaya Kumar (Collaborator), University of Agricultural Science
Angelique d'Hont (Co-PI), CIRAD
Guillermo Seijo (Co-PI), Insituto Botánica del Nordeste
Jens Stougaard (Co-PI), University of Aarhus
Charles Simpson (Consultant), Texas Agricultural Experiment Station

Mid-year report
BAC library construction and validation (CIRAD/EMBRAPA/UCB)
The most probable wild ancestors of cultivated peanut, *A. duranensis* and *A. ipäensis* with genome types AA and were used to construct two BAC libraries, one for each of the diploid species. The libraries are respectively 7.4 and 5.3 genome equivalents with low organelle contamination and average insert sizes of

106 kb. Both libraries have been used for the isolation of clones containing genetically mapped legume anchor markers, and resistance gene analogues.

Improvement of maps for the AA and BB genomes of *Arachis* (EMBRAPA/UCB/Aarhus)
The published map of the AA genome of *Arachis* (Moretzsohn et al 2005) has been further improved. The methods of analysing and representing synteny have been refined and the main affinities of the ten *Arachis* linkage groups with *Lotus* and *Medicago* are now clear.
The map for the BB genome of *Arachis* (*A.ipaensis* and *A.magna*) now has about 130 mapped markers, and is almost ready for publication. Clear synteny between three linkage groups of the AA and BB maps are visible.
578 SSRs have been screened in the AABB population parentals, (*A. ipaensis* x *A. duranensis*)4x and *A. hypogaea*. 167 markers (28.9%) are polymorphic. Of these 106 have been genotyped giving 74 markers in the AA genome and 32 markers in the BB genome. Marker distortion in this population is particularly low at only 8.5% (X2 5%).
In Brazil, with a view to the construction of RILs, F3 seeds for the AA, BB and AABB populations have been planted. F4 seed has now been harvested.

Identification of QTLs (EMBRAPA/UCB)
With a view to the identification of QTLs, F3 families from the AABB population ((*A. ipaensis* x *A. duranensis*)4x x *A. hypogaea*) have been phenotyped for SCMR leaf area index.

Clarifying the genomic relationships of cultivated and wild Arachis IBONE/EMBRAPA/UCB
A paper "Genomic relationships between the cultivated peanut (*Arachis hypogaea* – Leguminosae) and its close relatives revealed by double GISH." has been accepted for publication in the American Journal of Botany. The paper describes further evidence that the extant wild species most similar to the ancestors of cultivated peanut are *A. duranensis* and *A. ipaensis*. It also shows that in *Arachis,* divergence in genome repetitive element content accompanied, or has helped to drive, speciation.

An initial charactersiation of drought response in wild, synthetic and cultivated germplasm (EMBRAPA/ICRISAT/UCB)
Bioassays for component traits of drought tolerance have been done. 14 wild accessions, two synthetic and, as controls, three cultivated *Arachis,* were used. The results reveal a large variation of response to a progressive exposure to water deficit among different species and among different accessions of the same species. In general, the wild accessions evaluated have a conservative behavior: transpiration decreases dramatically as soon as water availability decreases in soil. On the other hand, the cultivated peanut varieties tested have a more opportunistic behavior.

The behavior of synthetic amphidiploids is dramatically different from the wild parentals, and is similar to cultivated peanut. Three wild parentals of the synthetics are available, *A. duranensis* V14167, *A. ipaënsis* KG30076 and *A. gregoryi* V6389. These accessions all have conservative behavior showing a steady reduction in transpiration at FTSW values below c.0.8. In contrast, both synthetics have opportunistic behavior, continuing high levels of transpiration until FTSW values of 0.3-0.2 have been reached. The synthetics thus have a similar transpiration profiles to cultivated peanut. This is a very significant result, because it shows that some drought-tolerance assays on diploid wilds may have limited predictive powers as to the behavior of synthetic amphidiploids. This can be attributed to the effects of polyploidy.

Selection of new lines with wild genes in breeding programmes (EMBRAPA/CERAAS/ISRA)
Three BC1 populations derived from synthetic and cultivated crosses have been made in Senegal, and five more are under development. In Brazil, two backcrossed populations are under development.

Tangible outputs delivered

- A greater understanding of the genomic affinities of A. ipaensis, A. duranensis, A. villosa, A. cardenasii and A. williamsii with cultivated peanut. Further support for the first two mentioned species being ancestral to cultivated peanut.
- An improved understanding of the relationships of the BB genome species, leading to a greater number of potential BB genome donors being identified.
- Three new amphidiploids developed.(A. hoehnei x A. simpsonii)4x, (A. batizocoi x A. cardenasii) 4x, (A. vallsii x A. williamsii) 4x
- Knowledge of drought response of different wild species and accessions. (Which ones are conservative, which ones opportunistic). The knowledge that synthetic amphidiploids are distinct in their transpiration profiles from their wild ancestors.
- Knowledge of TE characteristics of cultivated x cultivated RILs. Some single marker associations for Transpiration, TE and SPAD.
- New knowledge of variability for TE in 440 genotypes of cultivated peanut (including ICRISAT minicore, and selected breeding lines)
- 6,264 ESTs analysed (paper published) and sequences deposited in Genbank.
- Softwares, TROLL module for Staden, primer design programme and GeMprospector publicly available.
- 107 new characterised microsatellite markers made available for public use.
- 459 candidate legume anchor loci primer pairs made available on-line.
- A publicly available map for the AA genome of Arachis (BB map will be made available in due course). Identification of syntenic regions of AA and BB genomes.
- Candidate gene positions on the AA genome map
- Syntenic regions of Arachis and model legumes identified, greatly increasing information content of AA and BB maps.
- QTLs for disease resistance that co-segregate with candidate genes identified
- Repetitive element fragments isolated that are “specific” for the AA and BB genomes of Arachis.
- A tandem repeat specific in the IGS region of Arachis isolated.
- A tandem repeat present in the heterochromatic regions of Arachis isolated.
- Almost complete characterisation of one LTR AA genome “specific” element.
- Greater knowledge of the Arachis genome. (all results will be published in due course)
- Optimised extraction protocol for HMW Arachis DNA
- Large insert BAC libraries for the AA and BB genomes of Arachis with respectively 7.4 and 5.3 genome equivalents, and low organellar DNA contamination.
- Amphidiploids transferred to breeding programmes in Brazil, Senegal and ICRISAT-India
- Three BC1 populations, each of more than 300 plants in Senegal. Five other populations in development.
- Four BC4 families selected for leaf spot and rust resistance incorporated into EMBRAPA’s breeding programme for peanut.

Deviations from the workplan

The transfer of the amphidiploids to partners in SENEGAL and ICRISAT-India took longer than anticipated. This has delayed work.

Multiplication of wild *Arachis* seed took longer than anticipated, and the drought assays use more seeds than initially anticipated; this delayed the evaluation of wild *Arachis* for drought tolerance, but this is work now running smoothly.

Screening of polymorphic markers in the AABB mapping population, derived from (*A. ipaensis* x *A. duranensis*)^{4x} and *A. hypogaea,* has revealed very low levels of polymorphism on the BB genome. This will reduce the usefulness of the map that is derived from this cross.

Data availability
Much map and marker data have already been published, and softwares have been made publicly available:

Proite K, Leal-Bertioli SCM, Bertioli DJ, Moretzsohn MC, da Silva FR, Martins NF, Guimarães PM (2007) ESTs from a wild *Arachis* species for gene discovery and marker development BMC Plant Biology, **7**:7 doi:10.1186/1471-2229-7-7 *Open Access Journal*
All marker information at:
http://www.biomedcentral.com/content/supplementary/1471-2229-7-7-S1.xls
EST sequences have been submitted to Genbank. (numbers EH041934- EH048197)

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L.A (2006) General pipeline for the development of anchor markers for comparative genomics in plants BMC Genomics, 7:207 *Open Access Journal*
Table of anchor markers available at:
http://cgi-www.daimi.au.dk/cgi-chili/GeneticMarkers/table

Fredslund J, Madsen LH, Hougaard BK, Sandal N, Stougaard J, Bertioli D & Schauser L (2006) GeMprospector - Online Design of Cross-Species Genetic Marker Candidates in Legumes and Grasses. Nucleic Acids Research **34** (Web Server issue): W670-W675 *Open Access Journal*
The web programme GeMprospector available at:
http://cgi-www.daimi.au.dk/cgi-chili/GeMprospector/main

Moretzsohn M.C, Leoi L, Proite K, Guimarães P.M, Leal-Bertioli S.C.M, Gimenes M.A, Martins W.S, Grattapaglia D, Bertioli D.J. (2005) Microsatellite based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theoretical and Applied Genetics. **111**:1060-1071.
Genetic map info in paper. All marker information at::
http://dx.doi.org/10.1007/s00122-005-0028-x

Martins W, de Sousa D, Proite K, Guimarães P, Moretzsohn M, Bertioli DJ (2006) New softwares for automated microsatellite marker development. Nucleic Acids Research, **34** (Web Server issue): E31.
Open Access Journal
Software available at:
http://finder.sourceforge.net/
(follow the link to resources)

Other data will be published in due course.

Competitive Project #6: Marker development and marker-assisted selection for*striga* resistance in cowpea

Principal Investigator:
Satoru Muranaka, IITA
Collaborators:
Christian Fatokun, IITA
Adebola Raji, IITA
Boukar Ousmane, IITA
Dong-Jin. Kim, IITA

Michael Timko, University of Virginia
Ndiaga Cisse, CERAA
Moctar Wade, CNRA

Mid-year report

The present project seeks to develop a MAS strategy for cowpea that will allow rapid breeding of *Striga* resistant varieties in West and Central African countries.

Development of molecular markers linked to race specific *Striga* resistance genes in cowpea

A SCARs AGG/CTT 200B that shows polymorphic fragments of ~150 and 300 bp that were linked to susceptible phenotype for *SG5* have been developed. Pot trials were completed to score the RIL population derived from IT84S-2246 X Tvu-14676 against seed collected at Sous station de Koporo campaigne for *SG2* resistance. *SG2* resistance is not effectively marked by 61R or by MahSE2. We are in the process of testing both AFLP and SSR based markers for linkage to *SG2* resistance.

Screening of cowpea genotypes in *Striga* 'hotspots' in West Africa

The data of Striga hotspot trial in 2005 and 2006 were compiled and some cowpea lines were selected as potential parent lines for further MAS for Striga resistance in cowpea. As well as well known Striga resistant variety B301 and TVu14676, IT98K205-8 was selected as multiple Striga race resistance line (resistance to SG1, 2, 3 and 5) with beneficial agronomic traits, such as extra-early maturity (60 days cowpea), high-grain yield and disease resistance. Also, IT98K-216-44, IT81D-994 and IT 98K-503-1 were identified resistant cowpea lines to SG4z, Striga race in Zakpota, Benin Rep which multi-Striga resistant cowpea lines B301, TVu14676 and IT98K205-8 can not overcome. For 2007 trial, we selected 20 cowpea lines based on past 2 year trials to validate promising Striga resistant parents to each region.

Testing of markers and development of MAS protocols

Two F_3 populations (IT97K-499-35 x Dan Ila, IT97K-205-8 x IT95K-238-3) have been advanced to F_4 generations in IITA screen house in IIbadan during this dry season. The leaf samples of individual F_3 populations were taken for further MAS test with the molecular markers for *SG3* resistance, 61 R and MahSe2 developed in UVA. The Striga resistance of individual F_4 lines to *SG3* will be tested and the MAS efficiency of each marker at F_3 generation will be analysed.

Development of user-friendly tool kit for MAS

Dr. Kim joined in the project since Jan 2007 to develop user-friendly tool kit for NARS breeders to apply easily the MAS strategy. The 61 R and MahSe2 markers are converting into a user-friendly marker kit format which is composed of PCR tubes containing lyophilised forms of all the reagents including the PCR buffer, Taq polymerase, dNTPS, and primers, except the DNA template. Application of these kits will be used for the training of NARS partners.

Diversity analysis of Striga races in Nigeria (SG3) and Senegal (SG6)

To identify the diversity of Striga race existing in Nigeria and Senegal, pot experiments have been conducting in IITA Kano, Nigeria and CNRA, Senegal. From January to May, it is not an adequate season to grow cowpea even in the screen house due to too cool and too hot weather and short day length. During this period, *Striga* seeds collected in 2006 from 12 sites of Senegal cowpea belt were prepared for 2007 trials. The reference set of 60 cowpea lines was selected and will be used for diversity analysis of the Striga seeds in 2007 rainy season.

Identification of QTL for drought tolerance

At IITA, a set of RILs from the cross Dan Ila x TVu 7778 comprising 120 individuals along with the parents have been planted in Minjibir farm, Kano late February for phenotyping for drought tolerance. Continuous drought treatment was imposed from three weeks after planting and a number of traits (e.g.

transpiration rate, SPAD values, RWC, etc.) were measured in the field. During Mid-late May, the matured plants will be harvested to determine the productivity reduction rate of each individual line under drought condition for QTL analysis. The relation between each physiological trait measured in the field and the productivity reduction rate will be also analysed.
At CERAAS, the seeds of RILs from the cross DanIla x TVU7778 were planted in April and have been multiplying under irrigated field condition in order to be used for drought phenotyping trial during the coming rainy season.

Capacity-building and training activities
The training of Mr. Francois Badiane a PhD candidate from the University of Dakar and one student from Senegal in this summer have been discussed. We have also discussed the possibility of Mr. Lucky O. Omoigui from IITA to come to UVA. Lucky will come from September/Ocotber 2007 through early 2008.

Tangible outputs delivered
SCARs MahSe 2 marker to *SG1* and *SG3* resistance
SCARs AGG/CTT 200B to *SG5* resistance
Reference set of 60 cowpea lines for genetic diversity analysis
Potential parents selected in the field conditions for further Striga resistance breeding
Four populations for MAS test of *SG3* resistance
Phenotyping data of drought tolerance of Dan Ila x TVu7778

Deviations from the workplan
Due to some logistical problems at CERAAS, our partner Dr. Ndiaga Cisse could not carry out his assigned activity in Bambey, Senegal during year 2006. The major activities of CERAAS under the project were to conduct phenotyping of drought tolerance on three RILs developed in CERAAS and IITA, and to deliver the necessary plant materials to the project partners for genotyping work. This problem is causing serious delay of both phenotyping and genotyping works to identify QTL for drought tolerance. After a series of discussion in late 2006, CERAAS agreed on further support for Dr. N. Cisse's activity in 2007. Though Dr. Cisse was planning to conduct a drought resistance phenotyping work with the RIL population in 2007 dry season, the trial was postponed till late rainy season 2007 due to lack of RIL population seeds for the trial.
To cover the activity of CERAAS under the project, IITA has started phenotyping trial for drought resistance in Kano, Nigeria. Dr. Cisse will start his field evaluation of RIL population in late rainy season 2007.

Data availability
1. List of *Striga* resistant cowpea genotypes evaluated in 2005 and 2006 *Striga* hotspot trial was presented as "Identification of cowpea lines with resistance to 6 races of *Striga gesnerioides* in West Africa" in the Annual meeting of Japanese society for Tropic Agriculture held in April 1st (Tokyo Agriculture University, Japan) and the proceeding was posted in Japanese Journal of Tropical Agriculture 51 extra issue 1: 73-74

2. Timko M.P, Ehlers J.D, Roberts P.A. (2007) Cowpea. In: Genome Mapping and Molecular Breeding in Plants, Volume 3, Pulses, Sugar and Tuber Crops, C. Kole (Ed.), Springer Verlag, Berlin Heidelberg. pp. 49-67.
3. Chen X, Laudeman T.W, Rushton P.J,. Spraggins T.A, Timko, M.P. (2007) CGKB: an annotation knowledge base for cowpea (*Vigna unguiculata* L.) methylation filtered genomic genespace sequences. BMC Bioinformatics 8: 129 doi:10.1186/1471-2105-8-129
4. Timko MP, Gowda BS, Ouedraogo J, Ousmane, B. (2007) Molecular markers for analysis of resistance to *Striga gesnerioides* in cowpea. In. Integrating New Technologies for Striga Control:

Towards Ending the Witch-hunt. G. Ejeta and J. Gressell (Eds.) World Scientific Publishing Co. Pte Ltd, Singapore, pp. In Press
5. Timko M.P, Singh B.B (2007) Cowpea, a multifunctional legume. In: Genomics of Tropical Crop Plants, Paul H. Moore and Ray Ming (eds) Springer-Press. pp. In Press

Competitive Project #9: Development of low-cost technologies for pyramiding useful genes from wild relatives of cassava into elite progenitors

Principal Investigator:
Anthony Bellotti, CIAT
Collaborators:
Martin Fregene, CIAT
Alfredo Alves, EMBRAPA-CNMPF
Emmanuel Okogbenin, CIAT-NRCRI
Elizabeth Okai, CRI
Robert Kawuki, NAARI

Mid-year report

1. *Delayed Post Harvest Deterioration (PPD)*

Three backcross populations (BC_1), from controlled crossing of the PPD resistant genotype CW 429-1 to two elite genotypes (MTAI 8 and SM 909-25) and open pollination, were developed to identify molecular markers associated with delayed PPD for efficient transfer into cassava gene pools. The BC_1 populations were established *in vitro* from embryo axes, micro-propagated, and 4-10 plants per genotype established in the field. The BC_1 populations were harvested in March 2007 and evaluated for PPD at the 7th and 14th day after harvest.

In the BC_1 family B1PD280, derived from open-pollinated (OP) seeds from CW429-1, 23% of individuals showed delayed PPD trait at 14 days after harvest (DAH). In the BC_1 families B1PD 284, cross of CW 429-1 and MTAI 8, and B1PD289, cross of CW 429-1 and SM 909-25, 14.3% and 5.8% respectively of genotypes revealed resistance to PPD. Frequency distributions of delayed PPD in these families are shown in Appendix 1 and 2. The evaluations will be repeated for a second year in a replicated trial of 8 plants per plot and 3 replications.

BC1 genotypes with delayed PPD were selected and crossed with CMD resistant genotypes to combine the two traits and thus facilitate the testing and use of PPD delayed lines in breeding programmes by NARs partners in Africa. Currently a total of about 6000 crosses have been made and fruits are being harvested for embryo rescue to facilitate shipping of *in vitro* plants to project partners.

2. *Cassava Green mites (CGM)*

Good levels of resistance to CGM have been found in ***M. esculenta sub spp flabellifolia*** and its inter-specific hybrids. Crosses have been made to identify markers for the trait and to transfer CGM resistance to cassava. Higher levels have been reported in other wild *Manihot* species and a study was conducted to assess resistance. Field evaluations were carried out for several crop cycles (3 or more years) and genotypes that consistently showed a low damage rating were selected for further study under controlled conditions. Cassava leaves taken off plants in the screen house were placed on water-saturated cotton in plastic petri-dishes and one female mite (***M. tanajoa***) placed on each leaf lobe. There were 30 repetitions for each genotype.

Female mites are allowed to feed and oviposit for three days (72 hours) on the selected genotypes and the number of eggs oviposited counted and compared with oviposition on the susceptible control (CMC-40).Ten separate experiments involving 35 genotypes from inter-specific crosses and 8 wild *Manihot*

species were carried out. The inter-specific hybrids CW 235-72, CW 257-10 and CW 259-43 from *M. esculenta sub spp flabellifolia* showed a reduction of 51 to 49% in oviposition compared to the susceptible check, CMC-40 (Annex 5). Evaluation of several wild *Manihot* species revealed reduction in oviposiition of 45-69% (Appendix 5-9).

3. *Whiteflies*

Recent studies indicate that high levels of whitefly (*A. socialis*) resistance may exist in accessions of wild *Manihot* species such as *M. esculenta sub spp flabellifolia*. Genetic mapping of resistance to whiteflies is ongoing; another study to evaluate additional accessions of *M. esculenta sub spp flabellifolia* namely: MFla 61, MFla 52, MFla 33, MFla 19, MFla 21, MFla 25, MFla 75, MFla 15 for the ovipositional preference (or non-preference) of *A. socialis* together with the commercial control CMC-40 was conducted. Field and screen house plants were evaluated for free choice and non-free choice of ovipositional preference with adult whiteflies obtained from the CIAT colony. In general *M. esculenta sub spp flabellifolia* accessions had very low population levels (1.5 to 2.0 on a 1 to 6 scale). Leaf damage symptoms were absent (1.0 on a 1 to 6 scale). Previous studies have shown that whitefly oviposition is a reliable indication of resistance/susceptibility in cassava accessions. Results of free choice oviposition tests and non-free choice oviposition tests are presented in Appendix 3 and 4 respectively.

4. *Neo-tropical germplasm resistant to cassava mosaic disease (CMD) for Africa*

Improved Neo-tropical cassava germplasm bred for resistance to CMD at CIAT have been introduced to project partners in Nigeria (NRCRI), Ghana (CRI), and Uganda (NARO) and are in different stages of evaluation. A new set of 507 genotypes bred for resistance to CMD and for high protein/beta carotene content at CIAT is being prepared for introduction into Nigeria as tissue culture plantlets. The introductions will be evaluated for the target nutritional traits at NRCRI and in selected sites representative of the agro-ecologies where cassava is grown. Selections amongst the introductions with the desired traits will be sent for on-farm trials and to a crossing block for crosses to local Nigerian varieties to deploy widely the high protein and beta-carotene traits. Non-selected genotypes will be transferred to the regular breeding scheme.

Tangible outputs delivered

- The delayed PPD trait reported in a wild *Manihot* species and confirmed in a F1 progeny has been recovered in BC_1 derivatives
- Three BC_1 families evaluated for delayed PPD at 7 and 14 days after planting showed between 5 and 23% of genotypes resistant to PPD
- Several BC_1 genotypes showing excellent delayed PPD have been crossed to CMD resistant lines to generate BC_2 progenies and seeds are being harvested in preparation for establishment in vitro and shipment to partners in Africa
- Evaluation of several wild *Manihot* species and inter-specific hybrids for resistance to cassava green mites revealed reduction in oviposition of between 45 and 69%
- Discovery of high levels of resistance to whiteflies in several accessions of *M. esculenta sub spp flabellifolia* suggesting the trait is widely spread in the wild species gene pool
- Development of a new set of CMD resistant materials with enhanced nutritional quality for shipment to partners in Africa

Deviations from the workplan

There have been no deviations from the Workplan during this period of the project

Data availability

Data sets produced in the last six months include:

- PPD evaluations of BC_1 families at 7 and 14 days after planting
- Genetic crossed between PPD resistant BC_1 genotypes and CMD resistant parental lines

- Field and screen house evaluations from cassava green mites and whiteflies in many wild *Manihot* species and inter-specific hybrids

This data is being prepared at CIAT for submission to the GCP database.

Competitive Project #12: Drought tolerant rice cultivars for North China and South/Southeast Asia by highly efficient pyramiding of QTLs from diverse origins

Principal Investigator:
Zhi-Kang Li, Institute of Crop Sciences, CAAS
Collaborators:
Ze-Tian Hua, Rice Research Institute, Liaonin Academy of Agricultural Sciences
Zheng-Jin Xu, Shenyang Agricultural University
Co-Principal Participators:
Yongming Gao, Institute of Crop Sciences, CAAS
Arvind Kumar, IRRI

Mid-year report

1. Research activities and progresses at ICS of CAAS and Chinese collaborators

A total of 197 DT pyramiding lines derived from 7 DT pyramiding populations were progeny tested under stress and non-stress conditions during the dryseason of the of 2006-2007 in the Hainan province. Genotyping on these pyramiding DT lines were conducted with 53-111 differentiating SSR markers. Based on the analysis on genotype of 2 DT pyramiding crosses, a total of 22 DT QTLs from CDR22 and Yue-Xiang-Zhan, and 30 from Zhong413 and Teqing were pyramiding into C418. The progeny testing showed that a small portion of the pyramiding lines performed better than the parental DT ILs under drought. The detail analysis on genotype and phenotype data is in progress.

Table 1. Progeny test and genotyping for DT pyramiding lines derived from 7 crosses between DT introgression lines from different BC populations

Cross	Recipient	Donor for IL1 (female parent)	Donor for IL2 (male parent)	Individuals selected	No. of polymorphic SSR markers for genotyping
051F15	C418	Yue-Xiang-Zhan	C71	9	53
051F26	C418	Zhong413	Teqing	34	91
051F28	C418	CDR22	Yue-Xiang-Zhan	32	56
051F43	C418	Nipponbare	Yue-Xiang-Zhan	22	60
051F53	C418	Teqing	Yue-Xiang-Zhan	13	64
051F58	C418	Teqing	Q5	31	111
051F65	C418	Yue-Xiang-Zhan	Teqing	56	79

2. Research activities at IRRI

2.1 Developing DT IR64 lines by 2nd round of Designed QTL Pyramiding (DQP)

Genotyping of the selected F_3 lines of 2nd round designed QTL pyramiding with differentiating SSR markers are in progress and will be completed before June of 2007. The progeny test of the selected DQP lines of 2nd round pyramiding were conducted in lowland, upland and normal conditions, the data collection will be finished before June 2007.

2.2 Transfer DT QTLs into Smarna

New backcross populations are developed by crossing and backcrossing Swarna (the recurrent parent), an elite indica variety in India with some 1st round pyramiding DT lines in the IR64 background as donors (Table 2). The BC_2F_2 populations were planted in lowland stress and normal conditions in the dryseason of 2006-07 in IRRI and selection in the BC_1F_2 populations for improved DT and yield potential are being made.

Table 2. Transfer DT QTLs into Smarna

BC1F2 population	No. of plants in lowland drought stress	No. of plants in nonstress
Swarna*2 x DK23	~450	~200
Swarna*2 x DK36	~450	~200
S.Mahsuri*2 x DK23	~450	~200
S.Mahsuri*2 x DK36	~450	~200
S.Mahsuri*2 x DK371	~450	~200

Tangible outputs delivered

A new hybrid cultivar, Liaoyou 5224, derived from the cross between a selected DT IL, C5224 and a CMS line - Liao5216A, has been approved to be released to farmers in the Liaoning province based on its superior performance in both DT and yield potential in the provincial multi-location yield trails of Liaonin in 2006. This cultivar has been recommended to be tested in the national multilocalational yield trails of North China. Another new hybrid cultivar, Liaoyou 5249 derived from the cross between an IL (C5249) and Liao99A, will be tested in the Liaoning provincial multilocalational yield trails this year. A high yield DT japonica IL, HR95 is recommended to be tested in the provincial yield trial in Ningxia of Northwest of China.

Deviations from the workplan

Dr. Arvind Kumar, a new plant breeder for the rainfed ecosystem has joined IRRI to replace Dr. Atlin as the co-P.I. of the project since Sept. of 2006 and the research activities of IRRI have been fully resumed since then.

Data availability

All data will be rearranged as the format of IRIS and will be deposited in the GCP database before the end of 2007.

Subprogramme 4: Bioinformatics and crop information systems

Competitive Project #4: An eco-physiological–statistical framework for the analysis of GxE and QTLxE as occurring in abiotic stress trials, with applications to the CIMMYT drought stress programmes in tropical maize and bread wheat

Principal Investigator:
Fred van Eeuwijk, WUR
Collaborators:
Jean-Marcel Ribaut (Co-PI), GCP Director
Matthew Reynolds (Co-PI), CIMMYT
Scott Chapman (Co-PI), CSIRO
José Crossa, CIMMYT
Mateo Vargas, Universidad Autónoma Chapingo
Marco Bink, WUR

Mid-year report

Maize

According to the plan, the tasks during this period included: a) the continuation of the work on the methodology of multi-trait multi-environment QTL mapping (Task 5) and b) the development and preparation of a training course on GxE and QTLxE.

With respect to the first item, during this period a manuscript was submitted and is presently under reviewing process. The manuscript develops methodology based on mixed models to perform QTL mapping using multiple traits and multiple environments simultaneously. The approach allows investigating issues related to: a) the causes underlying GxE (QTLxE), b) the causes of genetic correlations between traits (pleiotropic and linked QTLs), and c) the causes of changing genetic correlations between environments. A second publication will appear in June 2007 in the book "Scale and Complexity in Plant Systems Research".

The second major task during this period consisted on the preparation of a training course on GxE and QTLxE. Two courses in Latin American have been scheduled for the next month of July. The first course will be delivered in Uruguay (a two-day course) and the second course will be delivered in Brazil. In collaboration with the corresponding local organisers, the necessary arrangements for the logistics and further organizational details were done. The course in Uruguay will take place at the Instituto Nacional de Investigaciones Agropecuarias, INIA, in Colonia. The local organiser is Sergio Ceretta, from INIA. The four-day course in Brazil will be held at the University of São Paulo, Piraciba. The local organiser in São Paulo is Antonio Augusto Franco García, from the department of Genetics of the University. The target audience consists of researchers and students. Facilities have been arranged for around 40 students per course. In addition to the logistic arrangements, we have been progressing in the compilation of materials and exercises to be used during the training. It is the intention to include a reasonable amount of practical work as a way to demonstrate the methodology used. The necessary arrangements with respect to the software the be used have been taken (contacts with VSN International as we are using the free software Genstat Discovery Edition).

Wheat

There are three tasks currently in progress in this component of the project:

- Task 4: single trait multi-environment eco-physiological modeling
- Task 5: multi-trait multi-environment eco-physiological modeling

- Task 6: development and presentation of course material

Task 4: This task is currently in progress for two different datasets. The wheat population analysed in this project was grown in 9 trials in Australia between 2002 and 2006; and in 5 trials under heat, drought or irrigated conditions at CIMMYT, Obregon, Mexico between 2004 and 2006.

For the Australian dataset, the 2006 trial data became available in December 2006. This was an irrigated trial and provides a contrast with the dryland environments experienced by the other trials in the dataset. A draft of a paper presenting the single trait mult-environment QTLxE analysis for yield, anthesis and height is in progress and will be submitted for review in June 2007.

The Mexican dataset was measured and collated by a Master's student, Suzuky Pinto, supervised by Matthew Reynolds (CIMMYT). Since Suzuky's visit to Australia in 2006, GxE and preliminary QTL analyses (QTL Cartographer analyses) have been performed on a number of traits including phenology, grain size, canopy temperature and spectral reflectance. These latter two traits required significant input from the wheat post-doctorate as the repeated measures nature of these traits makes their analysis more complicated than traits measured only once, such as, yield. In fact, the analyses are very similar to the multi-trait analyses discussed in Task 5. QTLxE modelling is underway for correlated traits, yield and canopy temperature and the results available for presentation.

In a recent trip to CIMMYT, Mexico (Mar-Apr 2007) Ky Mathews was able to complete a number of these analyses with Suzuky Pinto. Also, in a meeting between Scott Chapman, Matthew Reynolds, Suzuky Pinto and Ky Mathews a plan for the presentation of these results from these analyses was outlined. Papers for submission of this work will not be completed until September 2007.

Task 5: This task is in progress for the Mexican dataset as described below. It is planned for Ky Mathews to travel to The Netherlands in June 2007, and the multi-trait analyses of the Australian dataset performed then in collaboration with Marcos Malosetti.
The analysis of a repeated measures trait, such as canopy temperature, is very similar in structure to a multi-trait analysis. Canopy temperature is measured on the same plot on different days through the crop cycle, and frequently twice on the same day. Thus, each measurement time can be thought of as a separate trait and the analysis of all canopy temperature measurements across the crop cycle as a multi-trait analysis where the between measurement time correlations are accounted for. An analysis has been performed for this trait in the Mexican dataset, the results are currently being interpreted.

Task 6: A key objective of Ky Mathews' visit to The Netherlands in June 2007 is to assist in the development of the course material for the courses in July 2007, reported on in the Maize section above. Examples are available from the wheat datasets to use in the courses, and this will be discussed in June.

Tangible outputs delivered

- Malosetti, M.; J.-M. Ribaut; M. Vargas; J. Crossa; F. A. van Eeuwijk. A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stressed trials in maize (*Zea Mays L.*). Submitted to Euphytica, under revision.
- Malosetti, M.; J.-M. Ribaut; M. Vargas; J. Crossa; F. A. van Eeuwijk. 2007. Multi-trait multi-environment QTL modelling for drought stress adaptation in maize. In: Scale and Complexity in Plant Systems Research. Spiertz, J.H.J.; Struik, P.C.; Laar, H.H. van (Eds.). 332 p. Springer, ISBN: 978-1-4020-5905-6

Deviations from the workplan
Task 5 originally considered multi-cross QTL mapping. The data on wheat and maize actually did not allow the fitting of multi-cross models. For that reason, the emphasis has been changed from multi-cross to multi-trait modeling.

Data availability
Not applicable to this project

Subprogramme 5: Capacity-building and enabling delivery

There are currenly no Competitive Projects under Subprogramme 5

COMMISSIONED GRANTS

Subprogramme 1: Genetic diversity of global genetic resources

2005-01a: Completing genotyping of composite germplasm set of barley

Principal Investigators:
J. Valkoun, ICARDA
S. Grando, ICARDA
M. Baum, ICARDA
Collaborating Scientists:
Zhang Jing, CAAS

Mid-year report
The Generation Challenge Programme barley composite set consists of the following germplasm: ***H. vulgare*** **ssp.** ***spontaneum*** **(15%)** The barley wild progenitor is represented by 445 accessions, 65 % of this total are original accessions from GRU collection missions. Hyper-arid, arid and semi-arid collection sites represent 1%, 20% and 63% of the total, respectively. The set originate from 20 countries and collection sites belong to 58 ecological clusters. **Landraces (65%):**Landraces are a major part of the set with 1935 accessions. A significant part (20 %) is original material collected by ICARDA. Hyper-arid, arid and semi-arid collection sites are present with 3%, 33% and 43%, respectively. The landrace set originates from 85 countries and 78% is of CWANA origin. Collection sites belong to 255 ecological clusters. **Improved germplasm (20%):** This category includes cultivars, unfinnished breeders' materials and genetic stocks, which represent 13%, 6% and 1% of the CP barley set, respectively.

Analysis of 3000 accessions of barley with the 50 SSR has been carried out. A high level of genetic variability was observed. The number of alleles detected by the SSR markers varied (from 7 alleles for HVHVA1 to 33 alleles for SCSSR3907) and between accessions. The frequency of alleles varied also between SSR loci. (Figure 1).

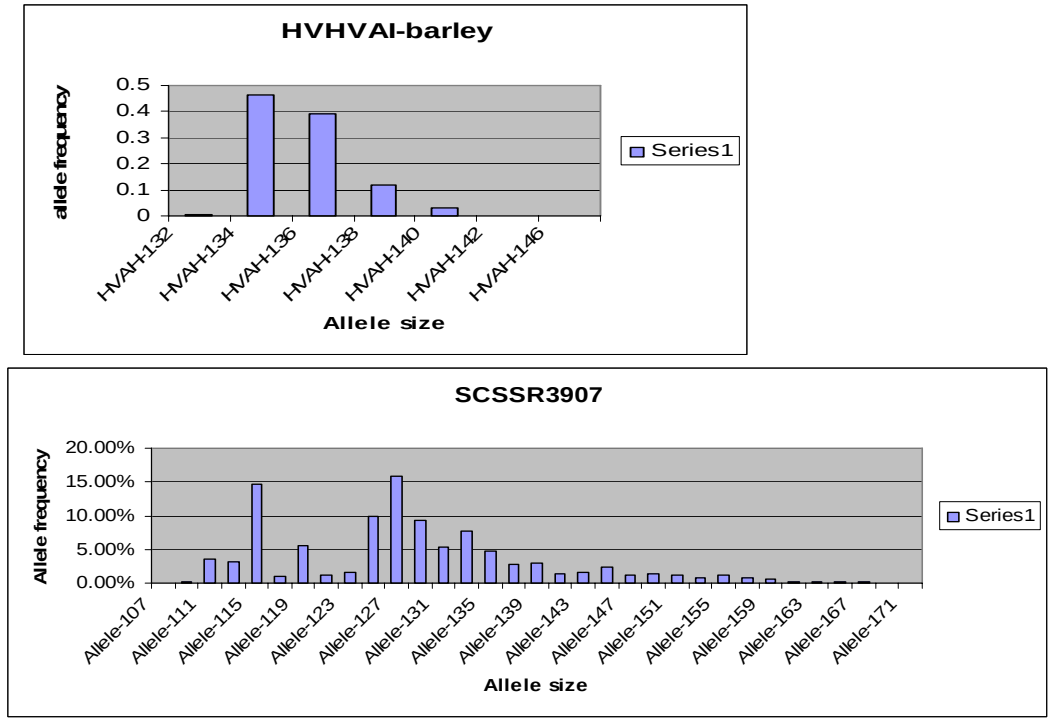


Figure 1. Differences in allelic frequency at different SSR loci in barley

The analysis of data with structure 2.1 (Prichard et al., 2000) allowed the detection of structured genetic variability. The germplasm from different regions was characterised by different allele frequencies. The distribution of genetic variability was divided according to regions. Some regions were very rich in alleles and possessed high levels of genetic diversity. Specific patterns were detected in the different regions demonstrating the presence of region specific alleles. The analysis of the data with DARwin5 (Version 5.0.130) allowed the illustration of genetic dissimilarity between accessions collected from the 85 countries (Fig. 2).

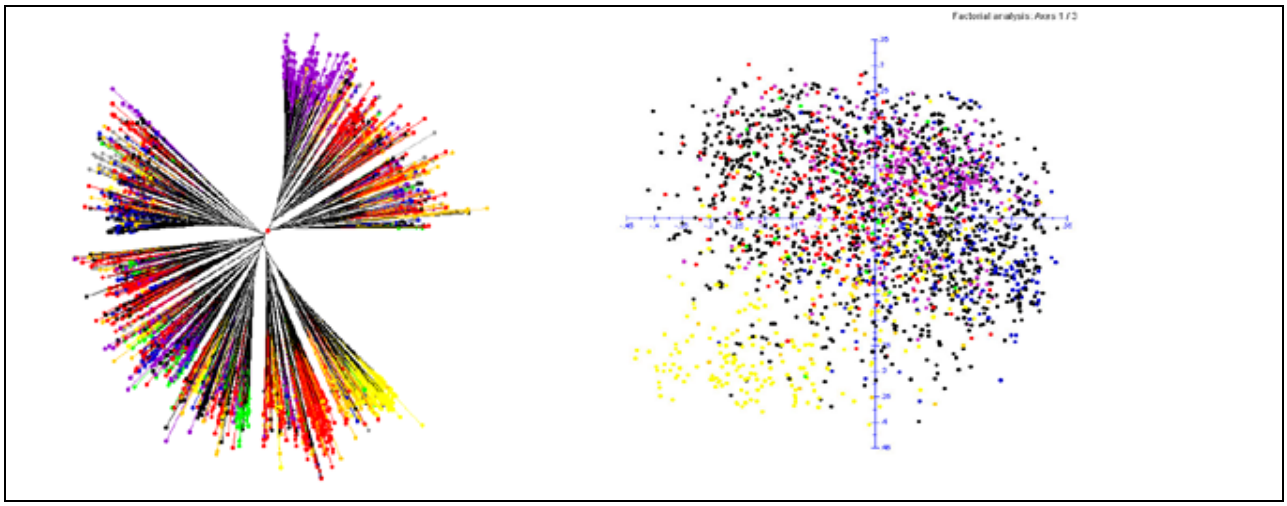


Fig. 2. Illustration of genetic dissimilarity between accessions from different regions (Darwin5).

Based on the genetic diversity of SSR markers, a core collection was developed.

Tangible outputs delivered

- 2692 barley accessions were analysed with 15 SSR primers at ICARDA.
- The composite collection has been planted for seed increase and seeds have been collected. Within one month all the remaining accessions will be harvested and the full collection of 2692 accessions will be available with the Genetic resource Section at ICARDA.
- A core collection has been established (about 300 lines) representing 10% of the whole collection. DNA is being used in the ADOC project.

Deviations from the workplan

- The analysis of barley collection with 50 SSR and EST-SSR derived from barley was to be done by ICARDA and CAAS. ICARDA had to genotype 15 SSR, while CAAS had to genotype 35 SSR.
- ICARDA has genotyped the collection with 15 SSRs and has posted the data in the central registry.
- CAAS analysed the barley collection with the 35 primer pairs and data was sent to ICARDA. We have made several suggestions how to improve the data quality and how to standardize the data by using known and common alleles. However, as CAAS did not respond to these suggestions we are unable to combine the data sets.
- After the Data Quality workshop held in Philippine, 8% of the collection was selected to be reanalysed with the same 15 primer pairs to estimate the error values. This work is still ongoing. Amplifications with the selected primers were made and we are in the stage of scoring the results. These results will be provided for the final report which will include 15 SSR on 2692 accessions.

Data availability

Genotyping data on 2692 with 15 SSR carried out by ICARDA is available locally at ICARDA in the GCP data base format. We also have deposited the data in the central register.

2005-01b: Completing genotyping of composite germplasm set of wheat

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborators:
Nachit Miloudi, ICARDA

Mid-year report
CIMMYT has sent ICARDA the primers for genotyping the 450 landraces which was their responsibility. No additional work has been done since the last reporting period.

Tangible outputs delivered
450 wheat landraces have been characterised and analysed with 26 SSR markers
Deviations from the workplan
ICARDA will be given yet another chance to complete the data; however, with whatever money remains in the budget (maybe 20% was not distributed yet from the GCP) CIMMYT will look to having some genotyping done at one of the genotyping hubs, for example, at AGROPOLIS. If there is no more budget, and ICARDA does not submit useable data, the PI is at her wit's end and doesn't know what else to try.

Data availability
Not ready yet

2005-01c: Completing genotyping of composite germplasm set of sorghum

Principal Investigator:
C.T Hash, ICRISAT
Collaborators:
Claire Billot, Agropolis
Ramu Punna, ICRISAT
Hari D Upadhyaya, ICRISAT
Jean-Francois Rami, Agropolis
Laeticia Gardes, Agropolis
Roman Rivalian, Agropolis
Monique Deu, Agropolis
Yu Li, CAAS
Tianyu Wang, CAAS
Ping Lu, CAAS

Mid-year report
During the extended reporting period genotyping of additional entries (including 60 wild accessions and over 300 sources of resistance to biotic stresses and/or tolerance to abiotic stresses) was completed at ICRISAT and Agropolis, as it became clear that the 250 accessions originally contributed by CAAS would not be made publicly available. The data sets (43 primer pairs x 3384 accessions) were debugged and analysed. A small number of accessions dropped due to missing data and a small number of markers dropped for which data appeared to be unreliable. Analysis of the population structure of the GCP sorghum composite germplasm set was undertaken with this slightly reduced data set (41 SSRs x 3367 accessions), and work to identify a reference subset was performed. The GCP sorghum SSR kit was completed and the data on this was made available and subsequently updated on the CIRAD website. The GCP sorghum composite germplasm collection SSR data set was submitted to the GCP Central Registry.

Tangible outputs delivered

- Set of 104 publicly available sorghum SSR marker primer pairs, detecting loci well-distributed across the 10 sorghum linkage groups, were assessed across labs at Agropolis, ICRISAT and CAAS for their ability to generate reproducible results using 48 diverse sorghum DNA samples (completed in 2004)
- Additional sorghum SSR primer pairs were developed, assessed for their ability to detect polymorphism, and mapped at Agropolis and ICRISAT (completed 2005)
- The GCP SSR marker kit for sorghum was developed at Agropolis, made available on line at http://sat.CIRAD.fr/sat/sorghum_SSR_kit/, and updated (completed 2006)
- Marker data set (41 SSRs x 3384 sorghum entries) for the GCP composite germplasm set for sorghum has been generated, debugged, and submitted to the GCP database (completed 2007)
- An invited presentation was made on the SSR-based diversity analysis of the GCP sorghum composite germplasm set at the GCP ARM in Sao Paulo, Brazil in September 2006
- An invited presentation on SSR-based diversity analysis of the GCP sorghum composite germplasm set was presented in the Allele Mining Workshop at Plant & Animal Genome XV in January 2007
- An invited presentation on SSR- and DArT-based diversity analysis of the GCP sorghum composite germplasm set was presented in the Generation Challenge Programme Workshop at Plant & Animal Genome XV in January 2007

Deviations from the workplan

This portion of the sorghum genotyping project was originally intended to have been completed by the end of 2005. Delays in generation of marker data of acceptable quality, and changes in the composition of the set of sorghum germplasm included in the study (due to lack of global public access to seed of the 250 entries contributed by CAAS), resulted in marker data generation continuing into mid-2006. A major hard disk crash at Agropolis delayed completion of the initial analysis and debugging of the combined raw marker data set from Agropolis and ICRISAT. Following the return of Claire Billot from maternity leave in early 2007, this initial analysis has been refined, the marker data set (allelic variants for 41 SSR primer pairs x 3367 sorghum entries) has been submitted to the GCP database, and a reference set of 383 diverse sorghum accessions have been identified for further phenotypic and genotypic evaluation (included as a first round in the ADOC project).

Although CAAS has received approximately 33% of the funds allocated for this project, it has yet been able to contribute sorghum SSR marker data that are considered sufficiently reliable by the Agropolis and ICRISAT team members. In the first year of this project (2004), it was found that the marker data set generated at CAAS could not reliably be merged with data sets generated at Agropolis and ICRISAT. For 2005, CAAS was requested to repeat its SSR marker data generation for the 670 sorghum accessions genotyped in 2004 and to include the agreed panel of standard entries in each run. For 2005 CAAS was also assigned a small portion of the total sorghum marker data generation (7 of 50 loci)—but agreed to initiate this only after demonstrating that it could produce a satisfactory data set for the initial subset of 670 accessions. The markers assigned to CAAS are in bins covered by markers generated by Agropolis or ICRISAT, so failure to complete this has not seriously affected genome coverage offered by the reduced number of markers (39, excluding data for two SSR primer pairs that appear to detected duplicate loci and two other for which allele calls are not considered sufficiently reliable) that we have ultimately been able to use for diversity analysis of the GCP sorghum composite collection. Besides over 3000 accessions were genotyped and some 138 000 Datapoints are released to the GCP repository (92% of the committed work).

Data availability

- The GCP sorghum composite germplasm set SSR marker data has been uploaded onto the GCP –Bioinformatics General Registry. The updated file (as of 28th Mar 2007) is available at

http://gcpcr.grinfo.net/index.php?app=datasets&inc=file_details&file_id=668 and contains fingerprinting data for nearly 3400 sorghum accessions with 39 SSR markers. A second version of the file was submitted on 15th May 2007 and will also be available at the same address.

- Information on the GCP sorghum SSR kit is available at http://sat.cirad.fr/sat/sorghum_SSR_kit/ and contains SSR primer sequence information, experimental protocol as well as details on expected allele sizes (based on sequencing) of PCR products for 10 diverse control accessions used to produce 3 control pools.
- Allelic information for 24 SSR markers generated at ICRISAT for the GCP sorghum composite germplasm set is available on our local ICRIS database.

2005-01g: Genotyping of composite germplasm set, Tier 1, rice

Principal Investigator:
Kenneth McNally, IRRI
Collaborators:
N. Ruaraidh Sackville Hamilton, IRRI
Claire Billot, CIRAD
Brigitte Courtois, CIRAD
Cesar Martinez, CIAT
Matthias Lorieux, CIAT
Claudio Brondani, EMBRAPA
Long-zhi Han, CAAS
Marie-Noelle Ndjiondjop, WARDA
Collaborator of WARDA: Susan McCouch, Cornell University

Mid-year report

Genotyping on 30 SSRs was completed during October 2006 for 30 SSRs. All genotyping was performed on LiCor 4300s with positive controls. Genotyping at the other sites (CIRAD, CIAT, EMBRAPA, and WARDA/Cornell was completed with data delivered between January and April, 2006. For alleles, the overall range of missing data was from 0.3 to 64% missing whereas for SSRs it was from 0 to 34%. Dropping those accessions (73) with the most missing data and the 3 SSRs with the highest levels of missing data, results in a data set of 2684 accessions by 47 SSRs where 93.6% of the alleles are called. Initial data analyses were accomplished during August-November 2006, on partial (39) or complete (50) SSR data. Analysis was done using DarWIN 5 and Structure 2.2. These analyses indicated that *O. sativa* grouped into 5 or more sections corresponding to indica, aus, aromatic, tropical and temperate japonicas. Depending on the setting of K, more or less putative admixture was observed. For the *O. glaberrima*, 6 different types were identified. Two of these were closely affiliated with *O. sativa*, while the other 4 appear to have partitioned by either their collection time and/or geographic origins. *O. barthii* was included in the analysis and the 5 accessions grouped among 2 of the *glaberrrima* groups. For the other 44 AA genome relatives, *O. glumaepatula* and *meridionalis* formed distinct groups while *O. nivara* and *rufipogon* grouped into 4 overlapping sections. Trees can be found in the attached powerpoint file. The new software for structure analysis of inbreeding species "InStruct" has recently become available from Carlos Bustamante (Cornell). Applying InStruct using the deviance informance criterion for best fit to a range of groups indicated that for the O. sativa subset, K=5 was the best fit. This corresponds nicely to the expectation of 5 subgroups.

This data set is version 1. Some peculiarities were noted, such as a deep split in the indica group and duplicate entries with different scores; to verify these results a careful re-evaluation of the IRRI data is underway. Also, 10 of the 30 SSRs assigned to IRRI were found to be less reliable and very difficult to score. The rescoring is being undertaken using the same approach as that of Claire Billot. The scoring seems to likely change for 2-20% of the calls usually by one step, with more differences among the 10 less reliable markers.

Furthermore, as a result of the SP 4 Data Quality workshop, obtaining better estimates of the quality of the dataset would be useful. Hence, we are re-genotyping the first 48 samples and ~1/10 of the remainder (8 from each of 30 plates). This re-genotyping is underway, and these results will be incorporated into data version 2.

Tangible outputs delivered
A version 1.0 of the genotyping data is avaialbel for 50 SSRs on the GCP composite collection of 2757 accessions. 3 SSRs will be dropped due to high levels of missing data. Dataset version 1.0 has been shared with IRRI (N.R.S. Hamilton), CIRAD (B. Courtois), and Wageningen (T. Van Hintum) with the understanding that a revision is underway.

A preliminary data analysis on this dataset has been accomplished using clustering (DarWIN5) and Structure2. Groupings were observed similar to those from work at Cornell by Garris et al (2005) for *O. sativa* and Semon et al (2005) for O. glaberrima (published in *Genetics* 2005). A representative subset of 275 O. sativa accessions was chosen for the ADOC and Haploryza projects for accessions with variety group identification based on prior analyses. The initial results from Structure2 (K=10, 8 and 2) were used to identify 157 additional accessions showing putative admixture for the Haploryza project.

Deviations from the workplan
While the genotyping was completed at IRRI in October 2006, analysis of the data indicated that allele calls needed to be verified. Furthermore, to improve quality assessment, re-genotyping of 1/10 of the accessions is being undertaken for the 30 SSRs assigned to IRRI. Both the rescoring and retyping are well underway. This will result in an improved version 2 of the dataset with higher confidence.

Data availability
A version 1.0 of the dataset will be submitted to the GCP repository in July 2007. As soon as version e version 2.0 is available this will replace version 1.0. Currently, data are stored in our local IRIS database at IRRI and in a private view of the TropGenes database at CIRAD.

2005-01j: Genotyping of composite germplasm set, Tier 1, common bean

Principal Investigator:
Matthew Blair, CIAT
Collaborators:
Maria Jose Peloso, EMBRAPA
Rosana Brondani, EMBRAPA
Shumin Wang, CAAS
Teresa Avila, CFP
Gloria Santana, CORPOICA
Sandra Lorigados, INCA
Steve Kresovich, Cornell University
Sharon Mitchell, Cornell University

Final project report
Executive summary:
Cultivated common bean germplasm is especially diverse due to the parallel domestication of two genepools in the Mesoamerican and Andean centres of diversity and due to the introgression between these gene pools and between cultivated germplasm and wild common bean populations. The two genepools can be morphologically distinguished into various races, however the association of these phenotypic differences with genetic structure have not been clear. In the GCP genotyping project for common bean we tried to address this through a thorough analysis of international and national germplasm collections representing wide genetic variability from both primary and secondary centres of

diversity. We were able to select a reference set from the analysis of the CIAT core collection with a newly developed marker kit made up of both genomic and gene-based microsatellites. The reference set represents the genetic structure found for the species which is divisible into the known genepools and two predominant racial groups per genepool. The division between the Mesoamerica race and the Durango-Jalisco group was very evident in the Mesoamerican genepool while the Andean genepool showed somewhat less diversity overall and a continuum between the Nueva Granada and Peru races. The Chile race could not be distinguished within the Andean genepool but there was some support for a Guatemala race within the Mesoamerican genepool. Introgression between the genepools was evident as was probable introgression between cultivated and wild common beans although this could not be assessed as thoroughly since wild accessions were not evaluated. In conclusion, this study has shown that common bean has very significant populations structure that could help guide the construction of genetic crosses that maximise diversity as well as serving as a basis for future association studies.

2005-01k: Unlocking genetic diversity in crops for the resource-poor (IITA) Phase I: genotyping of cowpea global collection

Principal Investigator:
Sarah Hearne, IITA
Collaborators:
Morag Ferguson, IITA
Jorge Franco Duran, Universidad de la República

Mid-year report

Primary data analysis of the global cowpea genotyping data with UPGMA using Cavalli Sforza-Edwards distance (CED) indicated nine potential groups of cowpea. This grouping was utilised to define a reference set (20% of the core set) of cowpea germplasm representative of the core set in terms of diversity/distance indices. The methodology of Franco *et al* (2006) was utilised to define the reference set. This methodology determines the sample size to be drawn from each cluster in proportion to the genetic distance or diversity within the cluster; we tested CED (genetic distance) and the number of effective alleles (EA; diversity). Simulations were run which selected at random the defined number of individuals from the core collection. The genetic diversity indices of these simulated reference sets was then determined and compared with the core collection. We found that CED gave more uniform results maximising key diversity and distance indices (see table below) and in general gave higher values than using EA as a selection criteria. We have outlined a reference set of germplasm using the top performing simulation from the CED selection. We have evaluated the UPGMA clustering of cowpea looking at patterns of clustering with reference to passport and where present characterisation data of the individuals within the clusters. No geographical basis of clustering has been determined. Nor has any basis of clustering been determined to date based on other characteristics, though these are still being assessed. In order to re-assess the clustering of cowpea we are utilising the model-based clustering method for multilocus data described by Pritchard *et al* (2000) and contained within the programme Structure 2.0 (Daniel Faluch, Matthew Stephens, Jonathan Pritchard, (March 2002), code by D. Flauch and J. Pritchard). Preliminary data generated using the admixture model, k from 1 to 35, 10,000 burnin, 50,000 MCMC and 4 iterations indicates that a larger number of clusters may be optimal. We are re-addressing the UPGMA classification in light of this finding and will investigate whether some re-definition of the reference set is required or desirable.

Tangible outputs delivered

We have defined a core set of cowpea germplasm form the ~15,000 accessions held by the IITA genebank. This core set has been genotyped using 16 SSR markers. A reference set of germplasm (20% of core) which is representative of the core in terms of genotypic, passport and characterisation data has been defined. In addition a further 100 accessions has been genotyped using 10 individuals per accession to estimate intra-accession variation within the accessions held by the genebank.

Deviations from the workplan
The project proposed to adopt 50 SSR markers into the genotyping efforts for cowpea. However, 50 suitable SSR markers could not be identified (all available SSRs screened) and the genotyping has been limited to 16 SSRs. To address the issue of marker limitation in a separate GCP funded initiative we have generated >40,000 EST sequences from varying cowpea genetic backgrounds to facilitate the search for EST-based markers. This effort has culminated in the identification of a further 5172 potential markers (3226 putative SNPs and 1805 SSRs) which once validation is completed can be adopted to expand the genotyping effort to 50 markers.

Data availability
All the genotyping data for cowpea is available from Sarah Hearne and is being uploaded on to the GCP database

2005-01l: Genotyping of composite germplasm set, Tier 1, cassava

Principal Investigator:
Martin Fregene, CIAT
Collaborators:
Paula Hurtado, CIAT
Morag Ferguson, IITA
Sarah Hearne, IITA
Alfredo Alves, EMBRAPA
Carmen de Vicente, GCP

Mid-year report
Genotyping of the global cassava genetic resource, over 2500 cassava accessions, with 30 SSR markers was conducted by CIAT and IITA; CIAT evaluated the collections with 22 SSR markers and IITA with 8 markers. Six other SSR markers to be analysed by IITA were eliminated from the study due to reports of duplicated loci in some accessions. In an earlier pilot study to select the set of 36 markers, CIAT evaluated 450 cassava accessions and found no duplicated loci.

In 2006, two datasets were generated for statistical analysis. The first data sets corresponds to 2494 genotypes evaluated at CIAT using 22 SSR markers, this data sub-set is 94.2% complete on average. The second dataset consists of 2575 genotypes evaluated at IITA using 8 SSR markers (it has between 50-100% of missing data per genotype). During 2007, additional work at IITA was carried out to reduce the missing data to about 10.4% average per genotype.

Data from 22 accessions evaluated at CIAT and IITA with the same marker sets were compared to detect the level of concurrence between silver stained PAGE (CIAT) and automated fluorescent capillary method (IITA).Agreement of SSR marker allele scores (homozygous or heterozygous alone) of 62% was observed between the two data sets. The large differences, 38%, between the two methods for visualisation of SSR markers is worrisome and led to CIAT sending the same set of 22 accessions and 8 markers to a third-party lab for automated fluorescent capillary analysis. This analysis has now been conducted and results are being checked with those from IITA.

Tangible outputs delivered

1) Completion of IITA data set from 8 out of the 14 SSR markers initially proposed for genetic characterisation (Table 1)

Table 1. Status of SSR marker data for the evaluation of global cassava resources from IITA in 2006 and 2007

SSR	Missing data in 2006 (%)	Missing data in 2007 (%)
SSRY9	41	9
SSRY102	18	11
SSRY135	26	6.4
SSRY147	36	8.9
SSRY148	32	8.3
SSRY161	57	14.4
SSRY181	26	13.3
SSRY182	24	12.1
Average	**32.5**	**10.4**

2) Comparison between scores generated by silver stained PAGE (CIAT) and fluorescent capillary method in 22 accessions with 8 SSR markers (Table 2).

Table 2. Concurrence of scores of 8 SSR marker in 22 cassava accessions based on silver stained gels and automated fluorescent capillary

SSR	% OF COINCIDENCE	CORRELATION
SSRY9	75.0	0.6
SSRY102	50.0	0.3
SSRY135	70.6	0.3
SSRY147	64.7	0.3
SSRY148	66.7	0.3
SSRY161	56.3	0.3
SSRY181	56.3	0.2
SSRY182	61.5	0.03
Average	**62.6**	**0.3**

Deviations from the workplan

- A third evaluation by a third party lab of 22 accessions simultaneously evaluated by IITA and CIAT is proposed as a mean to detect the source of differences between silver staining and fluorescent capillarity scores.
- Once the differences between scores have been detected, IITA and CIAT data set will be assembled and a new round of diversity analysis will be done using the complete data set (2400 accessions evaluated with 30 SSR) to conclude characterisation of the global genetic resources in cassava.

Data availability

SSR marker data of 30 SSR markers analysed in 2400 genotypes have been produced and will be stored in the web site of the Molecular diversity network (MOLCAS); it will be transferred to the GCP web site

for storage upon completion of analysis. Other forms of data, for example the new cluster analysis that will be conducted after validation of SSR, marker scoring will also be submitted to the GCP data base.

2005-01n: Genotyping of composite germplasm set, Tier 1, *Musa*

Principal Investigator:
Nicolas Roux, Bioversity International
Collaborators:
Isabelle Hippolyte, CIRAD in collaboration with:
C. Billot, CIRAD
V. Pomies, CIRAD
P. Cubry, CIRAD
A. Weber, CIRAD
L. Gardes, CIRAD
K. Tomekpe, CIRAD
Maria Kolesnikova-Allen, IITA

Update on progress
Since SSR marker profiles on *Musa* accessions are difficult to read, (see Appendix 1, slides 9-10) a first subset of 93 samples representing the already known diversity has been analysed on 26 SSR markers with Darwin software (Perrier et al. 2003).

These markers were either already published (13) and developed from A genotypes or specifically developed by CIRAD on BB accession Pisang Klutuk Wulung (Appendix 1, slides 2-3), chosen for amplification on A genome only. Analysis shows a good differentiation of subgroups of *Musa* (Cavendish, plantains, *Banksii*, etc). The markers were developed on A and B genomes and were all used in previous analysis (diversity mapping), particularly on A genome. The results indicate that none of them is specific to B genome and as a consequence, the first split of the diversity tree does not demonstrate the difference between the two genomes (see Appendix 1, slide 1).

From these 26 SSR markers, 12 were chosen for their good amplification quality. Results compared to those obtained with the 26 SSR markers exhibit the same pattern (Appendix 1, slide 1); for this reason, they were used to genotype the global set of 549 accessions.

Regarding the entire set of 549 accessions, 53 accessions could not be included in the first analysis because of missing data in pairing. Giant Cavendish was also removed because of unclear results regarding DNA extract quality. After checking the diversity tree with taxonomists in CIRAD, Paliama and Pisang lilin accessions were also discarded; is was clear that they were not the correct accessions (both triploid profiles instead of diploids). Some ambiguous results are still present in the diversity tree, perhaps from errors in collecting or extracting plant samples, or contamination during analysis, which should be discussed. The ploïdy and/or genome content of some accessions given with the samples should be also checked (Appendix 1, slides 4-8).

At IITA, the set of 27 SSR markers (four of them identical to CIRAD set) was used for genotyping the global set of 549 accessions of *Musa*. Raw data were collected using capillary-based electrophoresis on a semi-automated sequencer (ABI3100), and analysed using GeneMapper v.3.7 software. The resulting marker profiles proved to be fairly complex and inconsistent with existing information (such as ploïdy levels) for number of accessions.

For the initial analysis, a set of 366 accessions was selected from the global set, after exclusion of accessions with more than 50% missing values. The analysis was performed with set of 11 SSR markers

based on their good amplification quality. TreeMaker v. 1.0.1 software ([IV 2000] S.Piry, INRA – http://www.ensam.inra.fr) was used to create a neighbour-joining tree (see Appendix 2).

The resulting grouping of the accessions showed some similar patterns that were similar to the accession distributions obtained by the CIRAD team, although there were some differences as well. Generation of data for accessions with missing values is continuing; doubtful accessions are being cross-checked before the final statistical analysis is performed.

Tangible outputs delivered
Diversity analysis of 493 accessions of *Musa* (CIRAD)
Diversity analysis of 366 accessions of *Musa* spp. (IITA)

References
Perrier X, Flori A, Bonnot F, (2003) Data analysis methods. In: Hamon, P., Seguin, M., Perrier X, Glaszmann J.C (*eds.*) Genetic diversity of cultivated tropical plants. Enfield, Science Publishers, Montpellier. pp 43–76

2005-02: Supporting distribution of reference germplasm (IITA)

Principal Investigator:
Dominique Dumet, IITA
Collaborators:
Odu Babajide, IITA-Ibadan

Mid-year report
Since the last report, 720 accessions of cowpea have been planted in screen house in order to be indexed and to possibly produce virus free seeds. Out of these, 145 accessions are part of the cowpea mini core collection. So far in 2007, we received 5 requests for the minicore collection from GCP partners located in the USA (3), Senegal (1), and Nigeria (1). Four of these requests are still pending due to the delay in receiving import permission and/or signed copies of the standard material transfer agreement (SMTA).

In addition, 92 accessions of yam and 150 accessions of *musa* are at various stages of *in vitro* multiplication, acclimatization and indexing.
Tangible outputs delivered
Germplasm is stored under optimal conservation conditions and germplasm distribution is facilitated.

Deviations from the workplan
Due to the delay in genotyping, the formation of cowpea reference sets has been delayed, and therefore cleaning, multiplication and distribution has been delayed.

Data availability
Not applicable to this project

2005-02: Support and distribution of reference germplasm, *Musa* (Bioversity International)

Principal Investigator:
Nicolas Roux, Bioversity International
Collaborators:
Isabelle Hippolyte, CIRAD
Mathieu Rouard, Bioversity International
Elizabeth Arnaud, Bioversity International
Jaroslav Dolezel, IEB

Mid-year report
The 'mini-core' collection developed in the framework of the GCP project 2005-01n: 'Genotyping of composite germplasm set, Tier 1, *Musa*' comprises 48 accessions to which three other accessions were added to better represent the *Musa* diversity. The collection currently comprises 52 *Musa* accessions, which are listed at http://www.musagenomics.org/index.php?id=137. It was assembled using the GCP criteria regarding genetic diversity representation, duplication at Bioversity International Transit Centre (ITC), use in breeding programme and the FAO designation. For each of the accessions, ten extractions – each of 3 g of frozen leaves – were made with the Matab method (Risterucci et al. 2000); about 1 g of DNA was obtained per accession.

The DNA samples were deposited at the *Musa* Genome Resource Centre (MGRC) based at IEB in Olomouc, Czech Republic where they were stored for further distribution to the *Musa* research community upon request. The MGRC operates on a cost-recovery basis and the recipients are expected to cover packing and shipment costs. In order to sufficiently ensure that researchers benefit from this service, a maximum of 1 µg of DNA per accession is provided under a simplified MTA recently developed in collaboration with the Bioversity's Policy Research and Support Unit. The availability of DNA set was recently announced on the web site of the Global Musa Genomics Consortium: http://www.musagenomics.org/index.php?id=137.

Tangible outputs delivered
DNA from the reference set of *Musa* accessions (48 accessions, 1 mg/accession) was deposited and stored at MGRC for further distribution (1 µg per shipment per accession). An MTA was developed to acknowledge the GCP and other institutions involved in the work.

Deviations from the workplan
DNA of four accessions is missing because either the plants no longer exist in the field (2/4) or because there were not enough leaves available at the moment of extraction (2/4). This problem should be resolved, however, in the coming months.

Data availability
All Information is accessible on line: http://www.musagenomics.org/index.php?id=137.

2005-03d: Molecular characterisation of tier 2 (orphan) crops – Yam

Principal Investigator:
M.Kolesnikova-Allen, IITA
Collaborators:
J. Obidiegwu, IITA
R.Asiedu, IITA

Mid-year report
All generated on semi-automated sequencer AB13100 raw data were assembled and genotyped using the Genemapper v 3.7 (Applied Biosystems, USA). Some of the missing data and those samples that did not give reliable microsatellite peak profile were subjected to re-run in all affected markers. Fresh PCR was done for those cases when the first PCR failed. With the re-runs all missing samples were brought to less than 5% across all markers. All assembled raw data is set for genetic analysis using the genetic software Arlequin v 3.1, SAS, NTSYS and DARwin v 5.0.148 (CIRAD, France).

The first analysis results shown clear separation between species in the core collection. The allelic data generated during the analysis of raw data using Genemapper v 3.7 were converted to binary data. The presence (1) or absence (0) of allele was scored for each for each genotype across SSR markers used for

the study. Missing data accounted for less than 5% of the entire data set. Pairwise distance (similarity) matrices were computed using sequential, hierarchical and nested (SAHN) clustering option of the NTYSY-pc version 2.02j software package. The software generated dendrograms from TREE programme which grouped test genotypes on the basis of Nei's genetic distance using unweighted pair group method with arithmetic average (UPGMA) cluster analysis. The structure of the genetic diversity within population was further analysed by principal component (PCA) using the statistical analytical software (SAS) v.9. The eigenvectors for the first three principal components displayed a grouping of genotypes in a 3-dimensinal plot (Annex 1).The groupings were similar to that generated by the dendrograms. The result showed distinct separation of the six species at inter and intraspecific levels.

Tangible outputs delivered

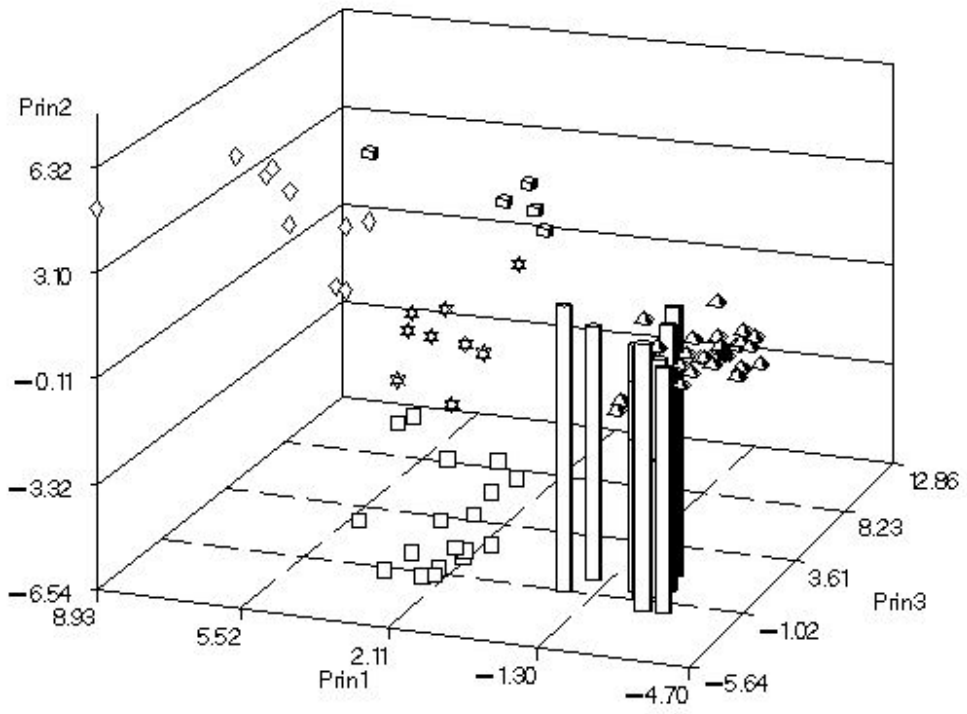
- Screening the full core set of 354 yam accessions with SSR markers is completed
- Missing data amount in full set is brought to less than 5% across all markers
- The genotyping of the full dataset is completed
- Preliminary diversity analysis is complete

Deviations from the workplan

Slight delay in data generation due to the full occupancy of the ABI3100 took place, due to which the data analysis is currently on its initial stages. Given NCE till October 15th 2007 would allow the analysis to be completed and verified.

Data Availability

The data is available and ready for submission to GCP repository. The scientists within SP4 in charge of data management in Central GCP repository will be contacted and the latest agreed template filled with their assistance. The submission would be completed before 15 October 2007.



Annex 1: Three- dimensional representation of genetic relationship between yam genotypes as revealed by principal component analysis of SSR markers

D. bulbifera- star
D. dumentorum- diamond
D. esculenta- cube
D. alata- square
D. cayenensis- pillar
D. rotundata- pyramid

2005-03e: Molecular characterisation of tier 2 (orphan) crops – Lentil

Principal Investigators:
Bonnie Furman, ICARDA
Michael Baum, ICARDA
Collaborators:
Christian Jung, ICARDA
Universität Kiel, ICARDA
Aladdin Hamwieh, ICARDA

Mid-year report
Analysis of 30 SSR markers was undertaken for the 1000 accessions of the global composite collection of lentil. A total of 6 SSR markers were deleted from the project due to inconsistent results (multiple peaks, etc.). Of the remaining 24 markers, data have been completed for 960 accessions (Tables 1). As discussed in the recent GCP data quality workshop at IRRI, we will run an additional 5 – 10 percent of

the samples to estimate genotyping errors. A no cost extension has been approved until October 2007. A final report will be available at that time.

Tangible outputs delivered

- Analysis of 24 SSR markers completed for 96% of the accessions.
- Initial data analysis for heterozygosity and polymorphic information content (PIC) was completed.

Deviations from the workplan
Completion of the work has been delayed due to the need to estimate genotyping errors.

Data availability
We have deposited genotyping in the central registry.

2005-03f: Molecular characterisation of groundnut (*Arachis hypogea* L.) composite collection

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborators:
R Bhattacharjee, ICRISAT
D Hoisington, ICRISAT
S Chandra, ICRISAT
RK Varshney, ICRISAT
JFM Valls, EMBRAPA
MC Moretzsohn, EMBRAPA
S Leal-Bertioli, EMBRAPA
P Guimarães, EMBRAPA
D Bertioli, UCB

Mid-year report
The groundnut composite collection, consisting of 1000 accessions (850 accessions from ICRISAT and 150 from EMBRAPA), was established using the available phenotypic characterization and evaluation data (Table 1). Composite collection consists of 184 mini core accessions, landraces, breeding lines, genetic stocks, wild *Arachis*, and four controls. In addition, ICRISAT and EMBRAPA added another 62 and 32 diverse accessions, respectively; totaling 1094 accessions for genotyping under this project. At ICRISAT, 912 accessions were genotyped using 21 SSR markers (10 markers contributed from ICRISAT and 11 from EMBRAPA). A fluorescent-based multiplex genotyping system was then used to generate four multiplexes that were used to fingerprint 912 accessions. The amplified PCR products were separated by capillary electrophoresis in an automated system using ABI 3700. SSR fragment sizes were called to two decimal places using the Genotyper v3.7 software. Because of the tetraploid nature of groundnut crop more than two peaks of almost equal size or 80% of the highest peak were observed (Fig. 1). Sometimes peaks of equal sizes were observed at two different positions. All peaks, which were of equal height or 80% of the highest peak, were recorded in the raw data file. A filter programme was then developed to remove all peaks that were less than 80% in peak height as compared to the highest peak and the data was reduced to two-peak situation. Final binning on the corrected data was carried out using the Allelobin software for calling the alleles.

At EMBRAPA, DNA extracted from 182 accessions (102 wild *Arachis* and 80 cultivated) representing the most diverse South American materials including those from Argentina and new collections from Santa Catarina and Xingu Native reserve in Brazil. ICRISAT supplied 48 DNA samples of the controls for use as standards to enable a comparison of diversity studies done in Brazil and India. Genoyping is

underway using 21 fluorescently labelled primer pairs that data on 182 accessions from EMBRAPA is still awaited.

Using 21 SSR data on 912 accessions, we detected 506 alleles (ranging from 6 to 47) with an average of 24.1 alleles per SSR locus (Table 2), mean PIC value of 0.871 (ranging from 0.568 – 0.971) and gene diversity 0.882 (ranging from 0.628 – 0.971) (Table 3). Among the three botanical types, *hypogaea* were more genetically diverse (based on range between minimum and maximum PIC values) than *fastigiata* and wild types. Gene diversity ranged between 0.464 and 0.960 in *hypogaea*, 0.613 and 0.960 in *fastigiata*, and 0.595 and 0.968 in wild types. Accessions from South America region revealed high gene diversity (0.910) while those from Oceania revealed low gene diversity (0.571). Detection included 391 rare and 115 common alleles at 5% and 266 rare and 240 common alleles at 1% levels. Principal coordinate analysis delineated the accessions in two major clusters (Fig. 2). The *hypogaea* and *fastigiata* types formed distinct clusters; however, a number of *fastigiata* types also grouped with *hypogaea* types probably due to common geographical origin. All wild *Arachis* accessions grouped together but within *hypogaea* cluster. .

Tangible outputs delivered

- Genotyping on 912 accessions and 21 SSR markers completed
- Data analysed and information on allele size and gene diversity available

Deviations from the workplan
None. Additional 94 accessions (62 from ICRISAT and 32 from EMBRAPA) were included along with those identified for groundnut composite collection (1000 accessions)

EMBRAPA data is still awaited

Data availability
Full data set, once EMBRAPA completes the genotyping of 182 accessions using 21 SSRs, will be provided to GCP.

2005-04: Assesing DarTs as a genome-wide scanning technology

Principal Investigator:
Andrzej Kilian, DArT P/L
Carmen de Vicente, GCP
Jean-Christophe Glaszmann, CIRAD-Agropolis
Associate Investigators:
Eric Huttner, DArT P/L
Peter Wenzl, DArT P/L
Ange-Marie Risterucci, Agropolis-CIRAD
Collaborators:
Ken McNally, IRRI, rice genotyping group
Claire Billot, Agropolis-CIRAD, sorghum genotyping group
Michael Baum, ICARDA, barley genotyping group
M Fregene, CIAT, cassava genotyping group
Nicolas Roux, BIOVERSITY INTERNATIONAL-INIBAP, Musa genotyping group
Patricia Lebrun, Agropolis-CIRAD, coconut genotyping group
Chandrika Perera, Coconut Research Institute, Sri Lanka
Prapit Wongtiem, Rayong Field Research Station, Thailand

Mid-year report

The project objectives are

- Validate existing DArT arrays for barley, sorghum and rice
- Develop and validate a high-density array for cassava
- Develop and validate medium-density arrays for coconut and banana
- Compare genetic-diversity patterns revealed with DArT and SSR
- Train GCP scientists in DArT to facilitate technology transfer

Barley has been changed to wheat due to insufficient SSR data on barley.

Since the previous report, no additional DArT data has been produced.
Several analyses using partial data sets have been performed and yielded conclusive positive conclusions on the compatibility between DArT results and those obtained with others.

Final SSR datasets have been released for all crops.
The currently available datasets can be summarised as follows:

Crop	**SSR Nb ind x Nb mk**	**DArT Nb ind x Nb mk**	**Collaborator**
Wheat	94 x 73	94 x 714	F. Balfourier
Rice	90 x 47	90 x 519	B. Courtois
Sorghum	192 x 41	192 x 684	S. Bouchet
Coconut	192 x 21	223 x 337	P. Lebrun
Cassava	134 x 36	124 x 424	P. Hurtado
Musa	184 x 22	184 x 836	A.M. Risterucci

Final formal comparative analysis is in progress by each collaborator, including simulations on comparative genetic structure revealed by different marker systems (SSR, DArT, RFLP when available) in relation to the number of markers.

Tangible outputs delivered

- about 150,000 data points were generated using DArT PL's sorghum, rice and wheat arrays, for a cost of $ 10500.
- 12 DArT arrays were developed by GCP trainees at DArT PL within approximately 5 months.
- three visitors were trained .
- over 600,000 data points were generated by the trainees. Over 500 new markers were identified for coconut, over 1000 for cassava and 1500 markers for banana.

Deviations from the workplan

The final analysis of the data, including comparison with SSR has been delayed by delivery of final SSR datasets for several species.
A no cost extension has been requested up to October 15, 2007 date at which a final report will be provided.

2005-05: Assesing Ecotilling as a methodology for targetted genotyping and SNP discovery

Principal Investigators:

Kenneth McNally, IRRI
Claire Billot, Agropolis-CIRAD (Co-PI)

Collaborators:
N. Ruaraidh Sackville Hamilton, IRRI
M. Deu, CIRAD
I. Hippolyte, CIRAD
F.-C. Baurens, CIRAD
J.-F. Rami, CIRAD

Mid-year report (submitted October 2006, revised report to be submitted shortly)

1. Our revised manuscript describing the agarose-based ecotilling procedure was accepted for publication in Molecular Breeding in August 2006: Raghavan C, Naredo MEB, Wang H, Atienza G, Liu B, Qiu ***F, McNally KL, Leung H (2006)*** Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping. Mol Breed, in press.
2. For rice, ecotilling using the agarose based procedure has been accomplished on 10 candidate gene loci with contrasts to both indica (IR 64) and japonica (Nippon-bare) for the complete IRRI minicore of 1536 O. sativa accessions. Most loci show frequent indica/japonica mismatches. The coding region for one of the TPP loci is conserved between indica and japonica types with only infrequent (rare) mismatches detected. A total of 384 accessions representing the different mis-match types were selected for ERF3, MAPK_7a, BZIP, and ADF_2a; and their amplicons were sequenced obtaining reads for both strands. The sequence data is currently being analysed.

Primer Name	Candidate Gene	Chr	Mis-matches detected on 1536 sativa accessions	
			vs. IR64	vs. NB
DREB2	AP2 domain TF	1	6	6
ERF3	ethylene responsive factor TF	1	7	4
ADF_2a	actin depolymerizing factor	2	7	6
ADF_2b			6	6
BZIP	bzip DNA binding protein	1	2	4
EXT	extensin	10	4	6
SUC	sucrose synthase	7	8	5
TPP_2a	trehalose 6-P phosphatase	2	7	7
TPP_2b			2	4
MAPK_7a	mitogen activated protein kinase	7	5	5

Ecotilling has also been accomplished on a panel of 95 diverse wild species accessions representing 23 Oryza species using the LiCor-based system for 11 primer pairs at 8 candidate gene loci. For 5 candidate genes, the agarose based method was compared to the LiCor. Within the AA genome species, the agarose based method can easily distinguish mismatches. Outside of the AA genome, banding patterns become complex and the technique is similar to fingerprinting by AFLP or RAPD. Additionally, an extended panel of 95 AA genome accessions has been screened using the agarose-based method with 5 primer pairs. Species-specific mismatches have been detected.

Ecotilling on a set of 190 O. glaberrima samples from the GCP composite collection and WARDA nominations (in collaboration with Dr. Marie-Noelle Ndjiondjop) was initiated in January 2006, using 21 primer pairs representing 12 drought candidate genes against two contrasts, Nipponbare and the O. glaberrima variety CG14 (IRGC 96717). Presently, 10 primer pairs have been completed on the O. glaberrima panel.

A manuscript on the use of ecotilling outside of Oryza sativa is being prepared.

3. For ecotilling on sorghum, new candidate genes and primers were designed since the rice primers failed to amplify from sorghum. Consequently, a pipeline was developed to identify the sorghum regions:

1. Blast rice primers onto rice pseudomolecule, release 3 (Primerblaster).
2. Align predicted PCR product with Orygene db and take into account longer full length sequence.
3. Compare rice sequence with assembled (SAMI) or non assembled (Orion Genomics) sorghum methyl filtrated sequences, and align the sequences (Multipipmaker).
4. Choose pertinent sorghum sequences to cover the same region as amplified in rice.
5. Design sorghum primers from the sorghum sequences.

From the 12 primer pairs designed from sorghum, 8 amplicons could be produced. Additional candidate genes from stay green QTL regions have also been chosen. The activity of cloned Cel I was compared to the crude extract (CJE) with similar results obtained. The LiCor and agarose based ecotilling at the AEC1 locus were compared yielding congruent mismatch patterns. For the trehalose phosphatase gene, 6 haplotypes have been identified in a panel of 24 sorghum accessions. Haplotypes identified at the RAV2 locus, for the most part, grouped into similar clusters as those identified by SSR analysis. Sequence verification of sorghum mismatches will be undertaken in the next few months.

4. Progress on ecotilling Musa has been limited due to the difficulty in identifying candidate gene targets with complete enough sequence for primer design. However, ecotilling has been achievable on Cacao and Coffee, highly heterozygous species.Extension of ecotilling to polyploid species and simulation with DNA pooling still needs to be explored.

Tangible outputs delivered

1. A publication on the agarose-based detection method has been accepted for publication in Molecular Breeding.
2. Ecotilling on agarose accomplished for 11 loci on 1536 O. sativa germplasm accessions with contrasts to both indica and japonica, an extended AA genome panel of 95 accessions for 5 loci, and a panel of 190 O. glaberrima accessions at 10 loci.
3. Sequences for 384 rice accessions at 4 candidate genes have been obtained to verify SNP calls.
4. Sorghum ecotilling on agarose has been achieved with similar results to those obtained on the LiCor. Haplotypes at a number of candidate gene loci have been identified.
5. Ecotilling in the highly heterozygous species Cacao and Coffee has been achieved.

Deviations from the work plan

While work on rice ecotilling is almost complete, the sequence analysis of sequences needs to be finished, and these results compared to mismatch patterns. Additional data (gels) have been obtained but the scoring is not complete. Further candidate genes need to be screened on sorghum, and amplicons chosen and sequenced to convert representative mismatch patterns to SNP identities. Extension of ecotilling to polyploids should be tested.

Hence, we are requesting a no-cost extension until May 15, 2007 in order to complete these tasks and prepare the final report.

2005-06: Supporting emergence or reference drought tolerance phenotyping centres

Principal Investigator:
Reinaldo Lúcio Gomide, EMBRAPA Maize and Sorghum
Collaborators:
Paulo E. P. Albuquerque, EMBRAPA Maize and Sorghum
João Herbert Moreira Viana, EMBRAPA Maize and Sorghum
Reinaldo Lúcio Gomide, EMBRAPA Maize and Sorghum
Camilo de Lelis Teixeira Andrade, EMBRAPA Maize and Sorghum
Elto Eugenio Gomes e Gama, EMBRAPA Maize and Sorghum
Sidney Netto Parentoni, EMBRAPA Maize and Sorghum
Frederico O. Machado Durães, EMBRAPA Maize and Sorghum
Fredolino Giacomini dos Santos, EMBRAPA Maize and Sorghum
Paulo César Magalhães, EMBRAPA Maize and Sorghum
Jurandir Vieira Magalhães, EMBRAPA Maize and Sorghum
Cláudia Teixeira Guimarães, EMBRAPA Maize and Sorghum
Newton Portilho Carneiro, EMBRAPA Maize and Sorghum
Flávio Dessaune Tardin, EMBRAPA Maize and Sorghum
Lauro José Moreira Guimarães, EMBRAPA Maize and Sorghum
Adelmo Resende da Silva, EMBRAPA Maize and Sorghum
Cleber Morais Guimarães, EMBRAPA Rice and Bean
Orlando Peixoto de Morais, EMBRAPA Rice and Bean
Beatriz da Silveira Pinheiro, EMBRAPA Rice and Bean
Maria José Del Peloso, EMBRAPA Rice and Bean
Natoniel Franklin de Melo, EMBRAPA Tropical Semi-Arid
Luiz Balbino Morgado, EMBRAPA Tropical Semi-Arid
Edson Alves Bastos, EMBRAPA Mid-North Agric.
Francisco Rodrigues, EMBRAPA Mid-North Agric.
Freire Filho, EMBRAPA Mid-North Agric.
Walter Quadros Ribeiro Jr., EMBRAPA Wheat
Maria da Glória Trindade, EMBRAPA Wheat
Ana Christina S. Albuquerque, EMBRAPA Wheat

Mid-year report

Introduction

The development of drought tolerant varieties for economical importance crops represents a major challenge for the 21st. century, considering that agriculture growth will be limited by world water availability. A first step to be taken in this direction is to select germplasm adapted to water stress conditions through appropriated screening techniques and defined protocols, techniques, and methods. Thus, the great challenge is the identification and characterisation of drought tolerant genitors to provide material to be used in genetic breeding programme focused on regions historically known as prone to water deficit during crop growing season. The improvement of drought tolerance relies on the manipulation of the traits that limit yield per each specie and their accurate phenotyping under the prevailing field environment conditions being target. This issue is particularly crucial for the breeding programme and identification of QTLs for traits categorised as adaptive as compared to constitutive traits. On this purpose it is necessary to have an infrastructure to allow plant exposure to water deficit pressure to be used for the evaluation of genotypes and characterization of plant physiological responses to these stress conditions.

The objectives of this project are to develop and make useful phenotypic evaluation protocols for cereals (maize, sorghum, rice, and wheat) and legume crops (common bean and cowpea), as well as to establish

some Phenotyping Centre of Excellence and Reference for Drought Tolerance Studies composed of phenotyping central laboratories, including controlled environment field and greenhouse and a training unit for researchers and research assistants, with facilities and well defined dry season periods to assure total soil moisture control during the drought phenotyping field experiments. In fact, the project seeks to establish a scientific and service network, like a model to drought tolerance phenotyping of cereals and legumes, including national and international genotypes.

Precision site-specific experimental areas for drought tolerance phenotyping in Brazil regions
The site-specific experimental (SSE) areas already were selected, installed, and characterised for cereals and legumes genotypes drought tolerance and adaptation phenotyping investigation to water stress, according to their geographical coordinates (determined by means of a differential global position system -DGPS), climatic condition (characterised with weather data series of at least 15 years with hydrological soil water balance), soil physical (texture, structure, porosity, apparent and real density, soil moisture retention curves, water infiltration rate) and chemical (fertility, organic matter, micro nutrients) properties. Some soil and environment attributes variability were determined on basis in methods and techniques of precision agriculture, DGPS, and geographical information system (GIS), which are being used in the data acquisition, storage, treatment, analysis and visualisation in each site. The spatial variability of soil physical (texture, structure, macro and micro porosity, apparent and real density, soil moisture retention curves, water infiltration rate) and chemical properties (organic matter, fertility, some micro nutrients) was evaluated by means of topographic survey and division of the sites areas in grids of 25 m x 25 m, utilizing a accurate survey laser total station Topcon Hiper and DGPS, on SAD-69 datum and UTM projection (south zone 23, 48 W a 42 W). Samples of the referred soil properties were collected in the grid intersections. Soil properties contour maps were obtained by interpolation with geostatistic adjusted models (krigagem). These maps were used to divide the sites-specific areas in uniform zones for abiotic water stresses studies in cereals and vegetables genotypes. In each selected site area, the water table was deep in order to avoid soil water capillarity effect in the genotypes root systems, and it was identified high and low points in order to avoid drainage problems.

The technical coordinators of the SSE areas are: Sete Lagoas-MG, Reinaldo L. Gomide, and Janaúba-MG, Paulo E. P. de Albuquerque from EMBRAPA Maize and Sorghum, Santo Antônio de Goiás-GO, and Porangatu-GO, Cleber M. Guimarães from EMBRAPA Rice and Bean, Teresina-PI, and Parnaíba-PI, Edson A. Bastos from EMBRAPA Middle-North, Planaltina-DF, and Passo Fundo-RS, Walter Q. Ribeiro Júnior from EMBRAPA Wheat/Savannah, and Petrolina-PE, Luiz B. Morgado from EMBRAPA Tropic Semi-Arid. The SSE areas are the following:

Site-Specific Experimental Areas for Drought Tolerance Phenotyping investigation in Brazil Regions (GCP SP1 DPN Project # 6)						
Brazil - UF	**Site-Specific / Local**	**Longitude (W)**	**Latitude (S)**	**Altitude (m)**	**Crop Species**	**Genetic Material[(1)]/ Phenotyping Strategies**
MG	Sete Lagoas [(2)]	-44.2467	-19.4658	731	maize, sorghum	Access and elite/ Preliminary
MG	Janaúba, North MG [(3)]	-43.3089	-15.8025	533	maize, sorghum	Access/ Preliminary
GO	Planaltina, DF [(3)]	-47.6142	-15.4528	944	wheat	Access/ Preliminary
GO	Santo Antônio de Goiás, GO [(2)]	-49.1711	-16.2811	823	rice, common bean, wheat	Access, elite, segregation/ Prelim.
GO	Porangatu, Nort GO [(3)]	-49.1486	-13.4408	396	rice	Access/ Preliminary
PI	Teresina, PI [(3)]	-42.8019	-5.0892	72	Maize, sorghum cowpea	Access/ Preliminary
PI	Parnaíba, PI [(3)]	-41.7767	-2.9047	5	Maize, sorghum	Access/ Preliminary
PE	Petrolina, PE [(3)]	-40.5008	-9.3986	376	sorghum, maize, cowpea	Access/Preliminary
[(1)] Cereals and Legumes Supporting Breeding Programmes at EMBRAPA, [(2)] Sites of Excellence, [(3)] Sites of Reference						

Protocols, methods and techniques for controlling and monitoring water stress in cereals and legumes genotypes for drought tolerance phenotyping

The protocols, methods and techniques for controlling and monitoring water stress in plants and soils are consensus technical-scientific documents developed and adopted by the EMBRAPA's researchers team to meet standardization procedures and practices needs to identify, characterise and select some new cereals and vegetables drought tolerant genotypes within the scope of the contrasting SSE areas in Brazil. Reference precision site-specific experimental data and information utilised in these documents are related and described, principally regarding agricultural field and laboratory equipment and sensors, structures and irrigation facilities used to control, to measure and to characterise the plants and soil water stress. Parameters such as surface climatic conditions, irrigation water application, soil water status, and plants water status and crop evapotranspiration are precisely controlled and managed. The standardization procedures and practices are taking into account the design, installation, calibration, evaluation, measurement, registration, storage, and transference of data in each site-specific.

The irrigation systems that are being used in the SSE areas are: conventional sprinkler (low to medium service pressure), localised (drip or trickle), and continuously moving straight lateral or linear-move system. The irrigation systems were tested and evaluated for water distribution uniformity (flow rate/discharge) or applied water depths by means of defining and controlling water pressure, flow rate, radius of throw, and emitters or sprinklers spacing. The water depths applied in the irrigations are being measured in collectors or catch cans in each genotype field plot, which are being placed (layout) transversally to the crop rows. The uniformity of distribution of the water in the irrigated plot was set to be equal or greater than 95 % (Christiansen Uniformity Coefficient or CUC = 95%). Some hydrometers are being coupled to the irrigation systems main lines. Every where, irrigation water application rate is

being set to be lower than basic soil saturated water infiltration rate in order to avoid surface runoff, which is not allowed in the SSE areas.

Irrigation management are being carried out through computation of reference evapotranspiration (ETo) and crop evapotranspiration (ETc), using both class A pan, modified Penman-Monteith equation methods, and also the crop (kc) and pan (kp) coefficients. The ETc is being determined by multiplying ETo for each genotype crop coefficient (Kc). Irrigation management strategy and irrigation timing criteria establishment are being performed based on a spread sheet (Excell) for ETo and ETc computation and a monitoring of soil water content in different depths. The irrigation is uniform after planting, germinating, and stand formation with 100% replacement of the soil water availability (SWA) and ETc - non water stressed situation. The water stress treatments are being obtained with different replacement level of the ETc and SWA, generating different application of water depths in the plots at pre-defined crop growth phases, defined for each genotype, according to breeders and physiologists indication in order to establish the water stress.

In each site-specific experimental area climatic condition was characterised and hydrological water balance (Thornthwaite & Mather) was determined with 15 to 50 years data series from standard weather stations (Brazilian National Institute of Meteorology – INMET). A standard procedure was established to calibrate and install the equipments and sensors of automatic weather stations in order to register automatically the microclimatic surface parameters locally for drought tolerance phenotyping purposes. Climatic surface data are being registered by means of automatic weather station, configured to measure temperature and relative humidity of the air, global solar radiation (net radiation in some), precipitation, speed and direction of the wind, class A pan water evaporation (in some) with half to one hour intervals.

Soil water content, in different soil layers, is being monitored by gravimetric method (the protocol consists of locating the soil moisture site, decide upon the sampling frequency and strategy, and assemble the necessary materials, collecting soil samples with an auger, weighing, drying in oven overnight and reweighing soil samples), equipments and sensors. The gypsum blocks sensors are being used for continuous monitoring of soil water content. These sensors are being calibrated by means of the gravimetric method – taking measurements of the electrodes (inside the porous blocks) electrical resistance (electrodes involved in known soil reservoir and embedded in water until saturation) against its water content by weighing. The wetter is a porous block, than the lower is the resistance measured across two embedded electrodes. This type of sensor is suited to various irrigation applications mainly with soil water stress condition. These sensors are being left in field to automatically monitor continuously soil moisture, allowing many replicates. The time domain reflectometers (TDR) equipment combine the knowledge of the waves signal propagation velocities in the presence of water in the soil medium, which affects the speed of these electromagnetic waves (slows them down slightly). The accuracy of TDR measurements depends on precise measurement of time and precise calibration with the relative volumetric content of water around the probe. The soil sampling and sensors installation to register soil water content are being made in at least four soil depths (15, 30, 50, and 80 cm).

Genetic material and traits per crop specie under evaluation for dought tolerance

Greenhouse experiments in Sete Lagoas-MG site: In the reference site of Sete Lagoas-MG, maize and sorghum genotypes experiments were carried out under greenhouse conditions, using two maize inbred lines (L29.1.1 and L14.1.1) and two sorghum inbred lines (SC 283 and BR007B) (drought tolerant and sensitive) and different water regimes (water stressed and no-water stressed pots), evaluating several soil and plant traits, aiming to evaluate maize and sorghum drought tolerance performance and extraction of RNA in Maize and DNA in Sorghum. The maize was sowed in August 15 2006, the water stress treatment was initiated in October 19 2006, and the inbred line L29.1.1 RNA extraction occurred in

October 25 2006. The sorghum was sowed in August 15 2006, the water stress treatment was initiated in October 25 2006, and the inbred line SC 283 DNA extraction occurred in November 01 2006.

GCP SP1 DPNetwork (#6) and WPModelling (#7) collaborative projects: During the dry season of 2007, from May until October, six maize and four sorghum selected genotypes were sowed in the Sete Lagoas and Janaúba-MG sites, for attending the trials of GCP DPNetwork and WPModelling collaborative projects. The maize plots were formed by 7 rows of 5,5 m with 0,8 m and 0,2 m between rows and plants along the row, respectively. The sorghum plots were formed by 12 rows of 6,0 m with 0,7 m and 0,2 m between rows and plants along the row, respectively. The statistical design in both experiments was random blocks with four replications. Both experiments are being irrigated by a 12 m x 18 m spacing sprinkler irrigation system. Two water regime treatments (with and without water stress) were applied in the pre-flowering and flowering crop growth stage for maize and sorghum, respectively. Also five rice selected genotypes were sowed in the Porangatu-GO site, following the same field procedures of last year experiment (2006).

In accordance with was planned in the GCP DPNetwork Project, in the Reference Site of Janaúba, MG, until the present data, there were performed the followings 2005/2006 activities including maize genotypes:

- Data were collected (Oct 2005) for several traits in the experiment where 36 inbred lines were evaluated for drought tolerance in the pre- and post-flowering of plant stage. The data are under statistical analysis.
- The experiments with 221 progenies derived from two synthetics under evaluation for drought tolerance were harvested in Nov 2005. Data will be collected for several traits and will be analysed statistically.
- EMBRAPA- Maize "Core Collection" – Two sets (Caatinga and Cerrado) of maize materials were planted, with and without water stress, using sprinkle irrigation. A lattice design 5 x 5 with 2 replications were used, and water stress was applied in the pre- and post-flowering time. Data are been collected for several plant traits (e.g., plant and ear height, data of male and female flowering, ear number and grain yield) and environment conditions.

Maize genotypes activities

F3 maize progenies evaluation experiment: One hundred (100) F3 maize progenies are being evaluated again in two experiments at Janaúba-MG and Teresina-PI sites. The Janaúba trials were planted on May 22 and 23 2007. The Teresina experiments will be sow on July 2007. These progenies came from a F1 contrasting cross between two inbred lines tolerant and non tolerant to drought. In each site, the experiments consisted of 100 maize F3 progenies including the two parental, using a lattice 10 x 10 design with three replications, which were planted in two blocks, where the two water regime treatments (with and without water stress) were applied in the pre-flowering crop growth stage. A dripping and a sprinkler irrigation system are being used in Janaúba-MG and Teresina-PI sites, respectively, for better water application control. The data are being collected. In Janaúba site, the last year experiments were harvested on October, 2006, and, in Teresina site, the last year experiments were harvested on Febrary, 2007, and all the data for several traits were collected for both sites. They have been described in detail in the last year report. The complete analyses of the results will be presented on the end of this year report. Some of results of these experiments are being utilised to identify the presence of news QTL for drought tolerance in a suitable contrasting cross for drought tolerance through the use of molecular markers, which is providing an important advance in genetic studies, generating new interest in their application in maize breeding programmes. The development of molecular markers has provided tools to assess important genes identifications in maize inbred lines.

.

Maize single crosses evaluation experiment: Single cross hybrids (SCH) are being evaluated again to enhance drought tolerance in inbred lines on 2007. In the Janaúba-MG site, there were set up trials to

evaluate maize SCH for drought tolerance. The two experiments consisted of 176 SCH tested last year + 29 new SCH + five commercials hybrids as testers: BRS 1010, DKB 390, DKB 359, AG 9010 and P30F90. A dripping irrigation system is being used for better water application control. The experiments were planted on May 23 and 24 2007, under two water regime treatments, with and without water stress, in the pre-flowering crop growth stage. The last year experiments (2006) were harvested at the middle of December 2006 and all the collected data for several traits are described in the last year report. The complete analyses of the results will be presented on the end of 2007 report. The development and selection of inbred lines are rather costly and time consuming tasks. Extensive trials are required to evaluate hybrids performance because of the large number of crosses produced. Hence, the identification of superior crosses for specific condition of growth would be important in predicting inbred line performance and to assign them to their specificity.

Maize progenies derived from two Synthetic: Synthetic drought tolerance (DT) has been selected during 4 years for N and P uptake. BAG-SE 029: selected for P use efficiency; BAG-BA 183: selected for P use efficiency; Synthetic multiple tolerance: 12 lines with tolerance to acid soils and Al toxicity and P use efficiency 4; BR 106: most OPV planted in Brazil, released by EMBRAPA Maize and Sorghum; BRS Sertanejo: variety develop for the Brazilian Northeast region with some tolerance to high temperature; BRS 1010- best single cross hybrid released by EMBRAPA Maize and Sorghum. In Janaúba-MG site, four types of experiments were installed in June/2007. Two of them were planted in areas with high (>20 ppm) and low P (3 ppm) levels. The others two experiments were planted in normal fertile soils. A randomised complete block design with 4 replications was applied. In addition, we have imposed controlled periods of water stress in these experiments, which were initiated by the pre-flowering time. These experiments will be harvested by middle December, 2007, and the data collected are being processed.

Maize inbred lines morphological root system evaluation: This work was carried out in 2006 with the objective of and growth characteristics of root and plant canopy of inbred lines selected for drought tolerance. Two inbred lines tolerant (L1 and L3) and two sensitive (L2 and L4) to drought were seeded, manually at space 0,20 m among plants and 0,20 m among lines, in seedling beds (0.8 m^2 - four lines of 1.2 linear m being five plants per linear m) with different levels of phosphorus, low (4 mg. dm^{-3}) and high (20 mg. dm^{-3}) at Sete Lagoas, MG site. The experimental design used was randomized complete blocks, with six replications. There were three evaluations of root system morphology by using the digital images system, WinRhizo Pro 2007a (Regent Instruments Inc.) and as well as evaluations of characteristics of growth roots and plant canopy at 14, 21, 28 (days after seeding). It was observed significant differences for root morphological attributes root and canopy growth. In general the inbred lines considered tolerant to drought showed root system different from inbred lines sensitive to drought. Those tolerant resulted in a larger root length, surface area, volume, and greater contribution of roots with diameter less than 0.5 mm in plants grown in the condition of low phosphorus availability.

Sorghum genotypes activities

Sorghum activities for DT Phenotyping for each site-specific experimental (Semester 1, 2007)				
Sites	Available Genotypes		Sowing date	Harvested date
	Preliminary	Advanced		
Janaúba, MG	49	100	June/2007	October/2007
Teresina, PI	49		August/2007	December/2007
Petrolina, PE	36 to 49		September/2007	January/2008
Note: Genotypes: divided in early, medium, and late cycle				

Sorghum inbred lines experiments in Janaúba-MG and Teresina-PI
Two experiments will be carried out again in both Janaúba and Teresina sites, in 2007, including 49 sorghum inbred lines using lattice 7 x 7, 3 replicates, experimental parcel with two rows 5 m (useful parcel with central 2 rows 4m), and plant arrangements with 0.50m spacing and 200 thousands plants/ha population, under two water regimes (water stress in the pre-flowering growth stage and plenty irrigation whole cycle). In Janaúba-MG, two others experiments will be carried out, with similar design, aiming the evaluation of 100 sorghum progenies to cross two contrasting inbred lines for drought tolerance (SC 283 and BR 007). The results are to identify isogenics inbred lines (RILs) aiming molecular markers studies. All those experiments are scheduled to be harvested by middle of December 2007, and the data collection already have initiated at Janauba site.

Sorghum genotypes experiments, in Petrolina-PE
Three experiments will be carried out again in Petrolina, PE, with pre- and pos-flowering water stresses, and plenty irrigation water regimes, utilizing 36 to 49 sorghum genotypes, that will be divided in early, medium, and late cycles, utilizing an experimental design with a random blocks with 4 replications (parcels of 2 rows: 0.5 m spaced and 9 m long). The following variables will be measured: flowering, plant height, leaf enrolment, waxy, leaf death percent (at flowering, dough grain, and physiological maturity), panicle weight, grain weight, grain number/panicle (at pre-flowering stress), 100-grain weight, and grain weight/panicle weight relation. The experiments will be sowed in September 2007 and the results will be compared between both experiments with and without water stress.

Rice genotypes activities

Rice genotypes evaluation at greenhouse in Santo Antônio de Goiás-GO: The experiment is being conducted in green house, using a split plot experimental design, with three replications. The levels of water regime, with and without stress, constitute the main plot and the genotypes form the sub plots. Water stress is being applied during the reproductive rice growth stage, and the genotypes are planted in PVC tubes (24,4 cm of diameter and 80 cm of height) filled with soil column. Eighty (80) genotypes are being evaluated. Well watered, soil moisture content is being kept at an adequate level up to panicle emergence growth stage with an soil water potential of –0,035 MPa at 15 cm soil depth. From this growth stage, two moisture treatments are being applied: 1) No water stress; and, 2) water stress up to the harvesting, with the application of 50% of the water evapotranspirated in the treatment 1. Traits that are being evaluated: grain yield, dry matter of shoot, percentage of spikelets sterility, number of grain per panicles, tiller fertility, plant height, and 100 seed weight. In addition, the treatments canopy temperatures are being measured, the drought susceptibility indexes are being determined, and the root density at harvest is being evaluated at 20 cm intervals down to 80 cm soil depth.

Core collection experiment at Porangatu-GO site: The experiment is being carried out with sixty four (64) genotypes from the EMBRAPA Rice and Bean Core Collection, using a random blocks design in split plot, with four repetitions. The water regime treatments constitute the main plots and the 64 rice genotypes are sowed in the split plots. Initially all plots will be well irrigated from planting date throughout 30 days after emergence with the soil water potential higher than - 0,035 MPa, at 15 cm soil depth (field capacity). Afterwards, the water treatments will be applied: 1) well irrigated 2) water stress. In the first treatment, irrigations will be controlled with tensiometers. The second treatment is receiving ½ of the water depth applied at the treatment 1. The following traits are being evaluated: yield, spikelets sterility, number of grain per panicles, tillers sterility and plants height. In addition, the treatments canopy temperatures are being measured, and the drought susceptibility indexes are being determined.

Wheat genotypes activities

Wheat genotypes evaluation at Planaltina-DF site in *out season (Safrinha)* planting period: During 2006 "safrinha" or outseason period, 120 genotypes were tested in preliminary drought phenotyping screening, in 3 planting periods and 2 replications. All tested genotypes until now were inferior to drought tolerance than Aliança (control genotype). Water status is being evaluated by measuring natural precipitation rate, and soil water content (gravimetric method and water retention curve). Considering water content in soil compared with water retention curve, for the 3 periods of planting, the water stress is being applied during beginning of tillering period and also during wheat grain filling. Twenty five (25) genotypes (with higher yield than the experimental average yields in the previous years) is being growed again in 2007 "safrinha" period in Planaltina-DF sites. This 3 planting periods started at the end of February and are separated by about 10 days intervals. The main criteria for DT selection is yield under water stress. As the secondary criteria, it is being used earlyness and initial vigour. A third drought preliminary "safrinha" field experiment is being carried out in 2007, with 78 new entries belonging to 21st semi-arid wheat SN from CIMMYT. Previous experiments results showed that some genotypes were unable to finish their cycle due lack of water (shortage). Data are being collected and processed.

Line source system field experiment in the winter period at Planaltina-DF site: Similarly to 2005 and 2006, in the winter 2007 at Planaltina-DF, when no natural precipitation happened, a "line source" experiment with several water levels under irrigation were planted only with 10 genotypes selected during two previous year with high and low drought tolerance. The experiment now are in the tillering stage (july). The water stress in several levels will be applied from tillering stage. A total of 600 mm will be applied in a 120 days of the wheat cycle in the best treatment and differences in water levels will be measured at the right and left side of the system. Previous results showed that the best drought stress genotypes (Aliança and E21) in safrinha, also performed best in the line source system during winter period. This means that selection for drought tolerance during winter (line source) can be useful for selection in out-season period. In the line source experiment control and some tested (outstanding) genotypes during previous years were included. This year we included also BR18 and Frontana with the purpose to compare them as parents for drought tolerance, because we produced a mapping population with those two genotypes for other purposes. Outstanding genotypes tested were BH1146, Brilhante (new cultivar with good performance under water stress), BRS 234, BRS208, PF0220337, two mapping genotypes (BR18 and Frontana), and 3 control genotypes (Aliança as a drought tolerante control), one sensitive PF020062 and one genotype adapted to irrigated system considered sensitive to water stress (BRS264) are being tested. Data is being collected. This year in addition to yield and components under stress, we will include root measurements. This year also physiologic measurements will be included such as proline contents, Aluminum tolerance with hematoxiline and photossintesis. During begginning of grain filling RNA will be collected from sensitive and tolerante genotypes under presence and absence of water stress, for a molecular work with functional genomics.

Wheat genotypes evaluation at Passo Fundo-RS site: In vitro experiments will be carried out under controlled conditions, with few genotypes, in order to maximise the efficiency of selection applying osmotic stress, simulating drought stress. A Master thesis was developed with this specific objective.

Cowpea genotypes activities

Cowpea genotypes at Teresina-PI and Petrolina-PE sites

Three experiments will be carried out again at Teresina, PI, and Petrolina, PE, sites, with pre- and pos-flowering water stresses, and plenty irrigation water regimes application, utilizing 30 cowpea genotypes, that will be divided in early ((IT 86D-386, IT 86D-394, Epace 1, IT 86D-1010, CB 3, BR 12–Canindé, UCR 2-1, UCR A-31, Manaus, BR 17 – Gurguéia), medium (IT 82D-699, IT 85F-2264, IT 85D-3850-2,

TE 86-75-56E, TE 90 1694F, TE 90-180-5F, UCR 1-12-13, Ipean V-69, Balinha, Marataoã), and late (IT 86D 627, TE 90 177-5F, TE 90 179-14F, TE 90 180-6F, TE 90 180-15F, BR 14 – Mulato, Vita 3, T – 28, Canapu, IPA 206) cycles, utilizing an experimental design with a random blocks with 4 replications (parcels of 4 rows: 0,45 m spaced and 9 m long). The experiment will be sowed on September 2007, and the following variables will be measured: flowering, plant height, leaf enrolment, waxy, leaf death percent (at flowering, dough grain, and physiological maturity), pod weight, grain weight, grain number/pod (at pre-flowering stress), 100-grain weight, and grain weight/pod weight relation. The experiments will be sowed in January 2008 and the results will be compared between both experiments with and without water stress.

Common bean genotypes activities

Common bean genotypes evaluations at Porangatu and Santo Antônio de Goiás-GO drought phenotyping sites: The following experimental designs are being used: a tripe lattice 11 x 11 and 8 x 8 in Santo Antônio de Goiás-GO and Porangatu-GO, respectively, with 4 replications. The levels of water regime, with and without stress, constitute the main plot and the genotypes form the sub plots. Water stress is being applied during the reproductive crop growth stage. Initially all plots are being well irrigated from planting date throughout vegetative period with the soil water potential higher than - 0,035 MPa, at 15 cm soil depth (field capacity). Afterwards, the water treatments are being applied: 1) well irrigated 2) water stress. In the first treatment, irrigations are being controlled with tensiometers. The second treatment is receiving ½ of the water depth applied at the treatment 1. The following traits are being evaluated: flowering, plant height, leaf enrolment, waxy, leaf death percent (at flowering, dough grain, and physiological maturity), pod weight, grain weight, grain number/pod (at pre-flowering stress), 100-grain weight, and grain weight/pod weight relation, pod sterility. In addition, the treatments canopy temperatures are being measured, and the drought susceptibility indexes are being determined. It is being evaluated 975 lines in F1.3 and 407 lines in F1: 4.

Tangible outputs delivered

- Two (2) EMBRAPA´s Phenotyping Centre of Excellence for Drought Tolerance sites (**Sete Lagoas-MG** at EMBRAPA Maize and Sorghum and **Santo Antonio de Goiás-GO** at EMBRAPA Rice and Bean)
- Five (5) EMBRAPA´s Phenotyping Sites of Reference (**Janaúba-MG** at EMBRAPA Maize and Sorghum Gorutuba Experimental Station, **Porangatu-GO** under EMBRAPA Rice and Bean coordination at Rural Agency of Góias Experimental Station, **Teresina-PI** and **Parnaíba-PI** at EMBRAPA Middle North, and **Petrolina-PE** at EMBRAPA Semi-Arid Tropic) (**Note:** The site of Teresina, PI is carrying out experiments in maize, sorghum, rice and cowpea species; and Petrolina, PE with cowpea and sorghum species)
- Climatic conditions (hydrological water balance, temperature and relative humidity of the air, global solar or net radiation, precipitation, speed and direction of the wind, class A pan water evaporation), soil physical (texture, structure, macro and micro porosity, density,EC, water retention) and chemical (fertility, organic matter) properties evaluated and characterised for each environment site
- Soil spatial variability of the site-specific areas determined and evaluated to investigate the genotypes in uniforms zones or plots in the field
- Irrigation system scheme installed and evaluated for each environment site (conventional sprinkler- low pressure and sectorial, linear moving sprinkler, localised –drip or micro-sprinkler)
- Irrigation system water flow rate and water management monitoring devices installed for each environment site (hydrometer, catch can kit, pressure meters, etc.)
- A reasonable number of genotypes (access and selected materials for each crop specie: maize, sorghum, rice, wheat, common bean, cowpea) were preliminary phenotyped and rice and maize selected entries were available under intermediate phenotyping level

- In course, the definition and implementation of data base (climatic, soil and plant data set) and modelling (in partnership with GCP Whole Plant Modelling Project # 7– CIRAD). During dry and wet seasons (2006/2007) were carried out experiments under WW-Well Watered (rain + supplementary irrigation) and WS (Water Stressed, only rain) with maize (04 genotypes), sorghum (05 genotypes) and rice (05 genotypes) in differents site-specific experimental (Sete Lagoas and Janaúba, MG; Goiânia, GO; and, Teresina, PI).

Deviations from the workplan

At the beginning, the project was delayed for 2-3 months, because of the financial resource transference problems until last June 6, 2005, but technically corrected by that time. This situation affects directly the acquisition of equipment and materials to be used in the sites. Thus, the preliminary actions were directed toward the arrangement of genetic material per each crop specie, during the first and second semester of 2005. There were also delaying problems with the importation of equipment, sensors and material, due to work documents in order to avoid paying importation tax, but most of the materials and equipment have already been purchased, since October 2006.

Regarding the partnership between GCP DPNetwork (CYMMIT/EMBRAPA) and GCP WPModelling (CIRAD) Projects, activities were initiated after Nov/2005-Fev/2006 (depending of each Site-Local). However, some legal aspects and procedures about Brazilian quarantine rules are delaying the arrival of new maize material from CIRAD.

Overall, the project is being carried out in time now. There is some time difference among crop seasons in each Brazilian region (Northeast, Southeast, and Centre- West regions).

Data availability

All the data produced are related to microclimatic condition, soil-water status and soil water availability in the soil profile (effective root system region), crop water requirement and water stress, and selected number of genotypes materials for each crop specie studied (maize, sorghum, rice, wheat, common bean, cowpea) for DT phenotyping are being transferred into database for each environment target site.

All the data sets are being posted in the Morph database and will be available for the project team by means of internet access. The data and information, generated in the experiments of the different sites, per different species, are being digitilised in documents and tables of the Microsoft Word, and spreadsheets of the Excel. Later these digitalized data are being inserted in the system called "**Morpho"**, which is a data management tool for ecologists, agronomists, and others researchers scientists. Morpho is a component of the Knowledge Network for Biocomplexity (KNB) and was created to provide an easy-to-use, cross-platform application for accessing and manipulating metadata (e.g. documentation) and data (both locally and on the network). The KNB is an international data repository dedicated to facilitating ecological and environmental research on biocomplexity. It enables the efficient discovery, access, and interpretation of data ranging from individual researcher efforts to highly distributed field stations, research sites, and laboratories. Morpho allows researcher to create metadata, (*i.e. describe their data in a standardized format*), and create a catalog of data and metadata upon which to query, edit and view data collections. In addition, Morpho provides the means to access network servers, in order to query, view and retrieve relevant data. Many types of "data" can be used with Morpho, including data tables and images. Morpho provides the means to access network servers, in order to query, view and retrieve relevant data. In the case of this project a specific server will be used, to be offered by EMBRAPA, for manager the use and the manipulating of the data generated at the different phenotype sites.

2005-07: Whole plant modeling of drought response in cereals

Principal Investigator:
Delphine Luquet, CIRAD
Collaborators:
Michael Dingkuhn, CIRAD
Jean-Claude Combres, CIRAD
Anne Clément-Vidal, CIRAD
Michel Vaksmann, CIRAD
Nourollah Ahmadi, CIRAD
Scott Chapman, CSIRO
Graeme Hammer, UQ
François Tardieu, INRA
Claude Welcker, INRA
Karine Chenu, INRA
Alexandre Bryan Heinemann, EMBRAPA
Elto Gama, EMBRAPA
Reinaldo Gomide, EMBRAPA
Cleber Morais Guimaraes, EMBRAPA
Edson Bastos, EMBRAPA
Alexandre Bryan Heinemann, EMBRAPA
Camilo Andrade, EMBRAPA
Fredolino Giacomini, EMBRAPA

Mid-year report
Component 1:
This component is almost finished; the valorisation of its interesting results is underway. Also during the last 6 months, activities consisted of:

- Preparation of one publication by Alexandre Heinemann (EMBRAPA) on his 18 month post doc (until Dec. 2006) at Cirad on target population of environment (TPE) characterisation for rice and maize in Brazilian Cerrados. Article will be submitted in June 2007 after last discussions required with S. Chapman (CSIRO), D. Luquet & M. Dingkuhn (Cirad) (during A. Heinemann & S. Chapman's visit in France at the end of May 2007).
- Presentation by A. Heinemann of two proceedings in a breeding conference in Sao Paulo state (Brazil)
- Presentation by A. Heinemann of one poster in IRRC in India in October 2006
- Application of component 1 approach (i.e. model based TPE analysis using the same model SARRAH) for sorghum in West Africa, regarding drought and photoperiod constraints (by M. Kouressy and M. Dingkuhn, CIRAD, publication submitted).

Component 2:
This component met (cf. previous reports) several problems in experiment activities, which resulted recently in the necessity of asking for a project extension (until 31/12/2007). This is mainly due to the difficulty to provide enough suitable data for applying, testing and improving model assisted phenotyping approach(es) in studying GXE interactions on three cereals: rice, maize, sorghum, among 5 to 6 experimental EMBRAPA sites and three different cropping seasons. The main steps of this component in the last 6 months are:
2006 dry season experiments totally achieved in February 2007:

- Analysis by A. Heinemann and D. Luquet of 2006 data sets acquired in Janauba (maize, sorghum, EMBRAPA CNPMS, coordinated by C. Andrade), Porangatu (rice, maize, sorghum; EMBRAPA CNPAF; coordinated C. Morais) 2006=> Only maize data in Janauba and of one experiment on

rice in Porangatu will be useful for the project and suitable for model application (and analysis of stress effect).

- Data from Teresina 2006-2007 experiments for rice, maize and sorghum underway (D. Luquet, A. Heinemann)
- Ecomeristem model application to rice: Porangatu data (2006) and Goiania data (2005-2006); by D. Luquet
- Sarrah crop model application to maize in Sete Lagoas (2005-2006, no stress), Janauba (2006) and rice in Porangatu 2006, Sete Lagoas 2005 (no stress); by A. Heinemann.
- 8 day visit of D. Luquet at EMBRAPA CNPMS (maize & sorghum) and CNPAF (rice & bean), 10-18th of Feb 2008:
- Organisation and last setup for 2007 season of experiment in (i.e. in Janauba and Sete Lagoas: maize & sorghum; in Porangatu: rice).
- Work on previous season data (to select what is worth being analysed or not) with C. Andrade, C. Morais and A. Heinemann).
- Seminars in CNPMS and CNPAF(at the university of Goias) to present WPM first results and try to find out students for participating in 2007 WPM experiments and discuss further possibilities of collaborations within the GCP.

Component 3:

This component is already providing a lot of results and tools for assisting process based trait model assisted phenotyping under drought in the GCP. The main results of the last 6 months are:

- Leaf expansion model (Ben-Haj-Salah & Tardieu, 1996, 1997) coupling with APSIM crop model (Wang et al. 2002); (K. Chenu, 18 month post-doc): model coupling is achieved. Leaf expansion of maize plants grown in contrasting field environments was simulated and successfully compared to observations to test model accuracy. A new module was developed in APSIM to account for yield sensitivity near flowering and formalize genetic control similarities of leaf and silk expansion in well-watered and stressed treatments (Welcker et al., 2007).
- Adaptation of Ecomeristem model (simulating cereals phenotypic plasticity) for simulating cereals behavior under drought conditions (M. Dingkuhn & D. Luquet). One publication submitted in May, other one during summer.
- Development of EcotropV4 modeling platform (by L. Tambour) hosting Ecomeristem model (with its new version dealing with drought impact) and a set of genotypic parameter optimisation tools for genetic analyses (finding parameter values to be provided for association studies, QTL analyses…). The platform opens the doors for future model improvements, developments (other crops, biological processes…)
- High throughput application of Ecomeristem for assisting (1) morphogenetic phenotyping of a rice mapping population (IR64 X Azucena) under P deficiency (publication submitted before July, by D. Luquet et N. Ahmadi) and (2) phenotypic plasticity and diversity of 200 sativa rice of core collection (IRRI, collaboration with K. Mc Nally); association with sugar analysis and morphogenetic traits phenotyping: by D. Luquet and A. Clément-Vidal.
- Development and proof of concept of a sorghum photosensitivity model enabling simulating genotypic behavior variability vs. environment: application for exploring genotype behavior among different regions in West Africa and defining ideotypes and compute genotype dependent parameters controlling photosensitivity that can be used for association studies (in sorghum core collection) or QTL detection (by. M. Dingkuhn, M. Kouressy, M. Vaksmann, publication accepted).

Tangible outputs delivered

- **Component 1:** Crop model based TPE (Target Population of Environment) analysis for rice and maize breeding in Brazil, regarding drought constraint. Same work realized for sorghum in West

Africa environments regarding drought and photoperiod**.** Methodology is now operational and can be applied to other regions and various crops;

- **Component 3:** Development and proof of concept of modeling tools for assisting (high throughput) cereals morphogenetic phenotyping and behavior analysis (extrapolation, etc…) under drought (cf. component 3 outputs: work of K. Chenu with F. Tardieu and G. Hammer; and L. Tambour, D. Luquet & M. Dingkuhn); 3 publications underway.
- **Component 3:** Vegetative morphogenetic and biochemical (sugar analysis) phenotyping of 200 sativa rice of the GCP core collection in rainfed conditions (CIRAD (D. Luquet, M. Dingkuhn, A. Clément-Vidal – IRRI, K. Mc Nally); Data analysed and soon available.
- **Component 3**: Development and proof of concept (on West Africa case study) of a sorghum photosensitivity model enabling simulating genotypic behavior variability vs. environment (drought X photoperiod). Potential application: exploring genotype behavior among environment, defining ideotypes, computing genotypic parameters controlling photosensitivity (e.g. for association studies in sorghum core collection or QTL detection): by. M. Dingkuhn, M. Kouressy, M. Vaksmann, publication accepted.

Deviations from the workplan

After a first 6 month extension asked to GCP until 31/12/2007 (to enable the organization of a last dry season experiment in 2007, because of previous experiment failures and missing suitable data for the project), a second extension was recently asked so that the final meeting (that will be hosted by Pioneer, USA) can be held at the beginning of 2008 (in the first half of February), rather than in December 2007. The main reason of this second extension need is that the last 2007 season experiments will be completed in October 2007, and we will then need time to collect data and finalize analyses, for writing WPM final report, abstracts and preparing correctly final meeting…). The second reason is that most of Brazilian and Australian partners of the project are in summer holidays in December/January and won't be able to attend the meeting in this period.

Data availability

As said in the last project report:

"The data sets acquired during the season 1 and 2 (2005-beginning of 2007) are currently gathered, cleaned and processed for statistical analysis and modeling application. The data set were acquired based on the protocol presented in the previous project report (in summary: fine plant phenology and morphogenesis measurement, biomass growth characterization, soil water dynamics, and daily meteorological data). These data should be all deliverable in the second semester of 2007. It is the same for high throughput phenotyping data on 200 sativa rice (component 3). The last season of experiment in component 2 will generate data that will be deliverable only at the end of 2007."

2005-08: Population structure, phenotypic information and association studies in long-generation crops

Principal Investigator:

M Carmen de Vicente, GCP

Collaborators:

Martin Fregene, CIAT
Luc Baudoin, CIRAD
Kodjo Tomekpe, CARBAP
Merideth Bonierbale, CIP
Jean-Louis Noyer, CIRAD

Final project report

Executive summary:

This project is an attempt to test the feasibility of association studies with materials and evaluation data available from regular breeding activities. The condition for using these materials is the existence of significant linkage disequilibrium (LD), correlated with variation, in genetically linked genes/markers. Five vegetatively propagated crops, highly heterozygous, are involved: cassava, potato, yam, *Musa* and coconut. These are scientifically less-endowed crops, with limited availability of advanced genomic tools. The success of the project is based on high quality phenotypic data and data, belonging to broad germplasm representations, or to breeding materials, with a significant level of LD demonstrated by surveys of linked molecular markers.

Progress for potato, cassava and coconut has been significant. In the three cases, phenotypic and genotypic data has been compiled/generated, genetic diversity analyses performed, and determination of LD has been conducted. Progress in *Musa* and yam has been delayed, however. The genotypic characterization of *Musa* swapped the use of SSR for DArT to catch up the time lost. Activities in yam were significantly hurt due to an unexpected mix-up of materials; however, progress has been made here and the results obtained have triggered relevant decisions for further work. The information for *Musa* was last updated 8 months ago with further details to follow.

A data analysis workshop, held in October 2006, was deemed very helpful by the collaborators. It was meant to acquaint researchers with basic theoretical issues around diversity analysis and association studies, and to gain hands-on exposure to data types either gathered in the project or representing it. Additionally, the importance of data quality and experimental design provided an important learning aspect of the activity.

This report is final, after the award of a NCE until May 15th 2007. However, activities are still ongoing in cassava, coconut, potato and yam, and research teams may send final updates when available.

2006-01: Developing strategies for allele mining within large collections

Principal Investigator:

N.R. Sackville Hamilton, IRRI

Collaborators:

Lorieux Mathias, CIAT
H.D. Upadhyaya, ICRISAT
Michael Baum, ICARDA

Mid-year report

Other GCP-funded work at IRRI has shown that one of the two datasets on which this project relies contains a high frequency of errors. Based on progress to date in detecting and correcting the errors, intermediate calculations suggest error rates of up to 20% for some loci in rice. The final corrected rice dataset is not yet available to this project, and therefore substantive progress cannot be made.

Using the available, error-prone rice dataset (therefore generating necessarily only tentative results), a preliminary analysis of genetic variation in 1957 traditional varieties Oryza sativa was undertaken (**step 1** of the project document), followed by a preliminary assessment of possible targets for novelty in terms of the molecular data space (**step 3** of the project document). No attempt has been made to start seeking relationships between molecular, passport and phenotypic data (step 2), because the local relationships are too sensitive to the data errors.

In accordance with the project document **step 1**, genetic distances between accessions were calculated, at least initially, using the $(\delta\mu)^2$ metric, which is designed for microsatellite data [(1)]. Questionable classifications were explored by repeating the analyses in two ways: first with other genetic distance metrics, including Roger's distance dR [(2)], and the simple matching coefficient most commonly used in the GCP; and second by reducing the geographic range of accessions analysed to include only varieties from countries where rice has been grown for more than 2000 years. All methods generated broadly similar classifications, but they differed in detail, and each generated different dubious cases: this typically occurs where the data either are not clustered or form overlapping clusters.

The overall analysis broadly confirms the previous 6-group classification of Glaszmann based on isozymes, although no support was found for the minor isozyme groups 3 and 4, and some indications were found that isozyme groups 1 (indica) and 6 (japonica) could be further subdivided. 7 groups were tentatively identified, falling in two main clades: indica-aus and japonica-basmati (see appendix). These comprise

- Two subgroups of indica (isozyme group 1): a diverse widespread subgroup and a separate subgroup from Madagascar and Sri Lanka
- Aus-type (isozyme group 2), a highly diverse group with a narrow geographic origin in Bangladesh, India and Pakistan
- Basmati-type (isozyme group 3), a group with a narrow geographic origin mainly from India, Pakistan and Iran, but also found in Afghanistan, Azerbaijan, Bangladesh and Nepal
- Three rather uniform subgroups of japonica (isozyme group 6), including two temperate japonica and one tropical japonica

15% of the accessions were not classified into one of the above groups, and there was on average a 15% mismatch between the SSR classification and the isozyme classification. Correspondence was best for the temperate japonicas and worst for the tropical japonicas (see appendix).

Three approaches were taken to identify accessions near which we may find genetic novelty (project **step 3**). Accessions were identified that had the highest distance from:

1. Their nearest neighbours - these may indicate a gap in the collection;
2. Their furthest neighbours - these are extreme genotypes and only novel genotypes will occur beyond them in molecular hyperspace
3. The centroid of molecular hyperspace - these are also extreme genotypes and only novel genotypes will occur beyond them in molecular hyperspace

However, the selected egregious accessions show a mismatch between SSR and microsatellites that is twice as high as for the typical accessions (see appendix). This is because data errors tend to generate outliers. As the algorithms are based on searching for outliers, they tend to select the accessions most affected by data errors. The success of the project depends critically on the reliable identification of

---

1 $(\delta\mu)^2 = (\mu_a - \mu_b)^2$, where μ_a and μ_b are the mean number of repeats found in the alleles of accessions *a* and *b* respectively: Goldstein DB, Linares AR, Cavalli-Sforza LL and Feldman MW 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. PNAS 92: 6723-6727

2 Roger 1972: $d_R = \frac{1}{m}\sum_{i=1}^{m}\sqrt{\frac{1}{2}\sum_{j=1}^{n_i}(p_{ij} - q_{ij})^2}$, where p_{ij} and q_{ij} are allele frequencies of the j[th] allele at the i[th] locus in the two accessions, n_i is the number of alleles at the i[th] locus, m is the number of loci. Reif *et al* 1995 conclude d_R is "appropriate to examine (i) the assembly and validation of core collections and (ii) the uncovering of pedigree relationships among OTUs (operational taxonomic units) such as the detection of essentially derived varieties in plant breeding or the identification of duplicates and collection gaps in seed banks": JC Reif, AE Melchinger and M Frisch 2005. Genetical and Mathematical Properties of Similarity and Dissimilarity Coefficients Applied in Plant Breeding and Seed Bank Management. Crop Science 45:1-7.

accessions that are true genetic outliers, not spurious outliers because of data errors. Therefore it has been vital to delay the start of the project until accurate data are available.

Based on the imminent availability of a corrected rice dataset, recruitment of staff has now been initiated.

At the time of this progress report, correction of the collated dataset is still in progress. Completion of the process of correcting the data is expected within 2 months. It is therefore now possible to initiate the current project; appointment of a PDF to process the corrected data is in progress. Step 1 of the project (analyse genetic variation) has been tentatively undertaken using the initial error-prone rice dataset, to assess the practical significance of the errors for progress in the project.

Tangible outputs delivered

- A draft SSR dataset has been delivered to the SP4 leader as part of SP1 activities at IRRI. This will shortly be replaced with the corrected dataset
- Knowledge of the importance of accurate data for searching for genetic novelty, and the sensitivity of the search to data errors. Using the draft dataset currently available, we have demonstrated that the detection of structure in the genepool is relatively robust: even using a dataset with a high frequency of errors is, we have identified a structure in the rice genepool that is generally consistent with previous studies. However, searching for genetic novelty by focusing on outliers is much more sensitive to data errors in the data, since data errors generate spurious outliers. We need to be confident that the outliers we test are true genetic outliers, not spurious outliers.

Deviations from the workplan

The have been two principal deviations from the workplan. First, the whole project timeline has been shifted, because of a major delay in the initiation of this project. This occurred because of unexpected continuing delays in the availability of data on which it depends. The first phase of the project requires complete, accurate SSR datasets from Subprogramme 1 of the Generation Challenge Programme. In the event, a whole series of problems caused extensive delays in the generation of this prerequisite dataset for rice. These have been reported elsewhere under GCP SP1. In summary:

- The start of the original SSR genotyping was delayed by delays in delivery of data, and then in delivery of germplasm supply from some sources; and eventually the composite collection was redesigned to exclude the nominations from three contributors
- Equipment failure at IRRI further delayed the start of genotyping at IRRI; eventually the equipment manufacturer declared bankruptcy, invoking a further delay while new equipment and supplies were ordered and purchased
- Data received from different laboratories showed evidence of different standards having been used to call allele data from the raw trace files, generating difficulties in collating data; eventually the one contributor's dataset was discarded because it proved impossible to reconcile with the others datasets; and IRRI undertook to repeat the genotyping for most primers in the rejected dataset
- Initial analysis of the final collated dataset still showed indications of problems with inconsistent data recording standards and dubious accuracy; therefore all available raw data files have been re-processed to verify and correct the allele calls.

Second, as a result of the delays, one partner in the original proposal (EMBRAPA) withdrew from the project. IRRI will undertake the work assigned to EMBRAPA.

Revised workplan
Other than the two changes above, the project workplan and milestones are essentially the same but following a different calendar. If we can regard July 2007 as the new start date, we have identical time line and milestones presented in the project document:

	Year 1				Year 2		
Activity	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**
Step A1 Analyse genetic variation							
Step A2 Establish relationships between new and old data							
Step A3 Identify targets for novelty							
Step A4 Select accessions							
Step A5 Genotype selected accessions							
Step A6 Validate objective function							
Step A7 Evaluate progress and efficiency							
Step B1 Analyse genetic variation							
Step B2 Establish relationships between new and old data							
Step B3 Identify targets for novelty							
Step B4 Select accessions							
Step B5 Genotype selected accessions							
Step B6 Validate objective function							
Step B7 Evaluate progress and efficiency							

Milestone	**Completion date**
1. Completed selection of first batch of accessions	End month 6
2. Completed genotyping first batch of accessions	End month 9
3. Completed first iteration	End month 12
4. Completed selection of second batch	End month 15
5. Completed genotyping second batch	End month 18
6. Completed second iteration	End month 21

Data availability
Draft SSR rice dataset previously reported as part of SP1

2006-02: A dataset on allele diversity at orthologous candidate genes in GCP crops (ADOC)

Principal Investigator:
Dominique This, Montpellier SupAgro, Agropolis
Collaborators:
OL1: Brigitte Courtois, Claire Billot, Romain Philippe, Jean François Rami, Dominique This, Agropolis
OL2: Merideth Bonierbale, Reinhart Simon, Roland Schaftleitner, CIP
OL3 : Dave Hoisington, Rajeev Varshney, Spurthi Nayak, Jayashree B, ICRISAT
OL4 : Dominique Brunel, Pierre Mournet, INRA-CNG

CrP1: Rice: Ken McNally, IRRI
CrP2: Barley: Michael Baum, Wafaa Choumane, ICARDA
CrP3: Sorghum: Tom Hash, Rajeev Varshney, Dave Hoisington, ICRISAT
CrP4: Bean: Matthew Blair, CIAT
CrP5:Chickpea: Rajeev Varshney , Hari Upadhyaya, Dave Hoisington, ICRISAT
CrP6: Cassava: Martin Fregene, CIAT
CrP7: Potato: Merideth Bonierbale, Reinhart Simon, Roland Schaftleitner, CIP

Mid-year report

The ADOC project aims to characterise allelic diversity at orthologous loci of candidate genes for drought tolerance in seven GCP crops, working on reference collections of around 300 accessions for each crop, selected by crop partners. Six gene families (ERECTA, DREB, SS, SPS, ASR and INV) were selected, during an initial scientific exchange with **gene specialists and advisers**, as the initial subset of target genes. The definition of consensus and specific primers amplifying those sequences was shared among ortho-labs, in order to complete our information on those gene families and facilitate the allelic sequence project.

Since October 2006, **crop partners** (CrP1 for rice, CrP2 for barley, CrP3 and OL1 for sorghum, CrP4 for bean, CrP5 for chickpea, CrP6 for Cassava and CrP7 for potato) have fulfilled the selection of the reference collection, extracted DNAs and sent them (for most partners) to partner OL4, responsible for allelic sequencing. Those collections have been screened for anonymous markers and field performance in other programmes and could be made available for collaborative phenotyping projects and additional genotyping.

During the same period, **ortho-labs** have improved their knowledge of the gene families and design specific primers for some candidate genes. Here is some update on this task, for each gene family:

Sucrose synthases

Consensus primers were designed by CODEHOP primers for 8 conserved blocks. Cloning was necessary in some cases to retrieve SuSy gene fragments. Five new Susy gene fragments were obtained by this methodology and confirmed by sequencing them. Specific primers were designed for 7 rice loci but two very similar genes were difficult to distinguish. Specific primers were designed for potato and barley SuSy gene fragments, and the primers for the remaining crops should be available in September.

Sucrose phosphate synthases

The S6PP domain of SPS domain is present in SPS and often absent in “SPS-like proteins” in the databases. This domain was proposed by A. Price (Aberdeen, UK) as a putative functional domain involved in root growth genetic determinism, and will be integrated in the analysis. Two primer pairs are generally sufficient to amplify this domain, and there are about five SPS genes containing this domain in each species. Primer design for the S6PP domain has been accomplished for rice, sorghum and potato. For cassava, chickpea and phaseolus, degenerated primers were designed to get new sequences for specific primer design. One new sorghum sequence was identified from the sequence release. Consensus primer design is underway.

Vacuolar invertases

A number of sequences similar to vacuolar invertases have been identified in different databases, some of them being putative fructan biosynthetic enzymes. FBE sequences have been conserved for further analysis because of their evolutionary significance in relation to sugar metabolism. Consensus sequences were defined for dicots and monocots and sequence tests are running. Specific primers were defined for rice, sorghum, barley, potato and common bean and are currently tested.

ABA-Stress–Ripening (ASR) proteins

The ASR gene family contains at least two branches for monocots, one for solanaceae and one for legumes. Specific primers have been designed and tested to amplify 6 genes in rice, 6 in sorghum and 3 in barley. A similar work is underway for dicots. New ASR genes have been identified in chickpea by using specific primers designed on phaseolus sequences. Genetic diversity has been identified already for rice and sorghum and varies among genes. Three ASR genes have been now sequenced on the rice diversity panel and will be analysed in detail. A parallel study is underway in order to determine the copy number of the ASR gene family in the species under consideration.

ERECTA and ERECTA-like proteins

The gene family is divided in two groups, ERECTA and ERECTA like proteins, with a different distribution in monocots and dicots. A method relying on consensus primers was adopted to find ERECTA and ERECTA-like genes on the seven species. Besides the full gene sequences from rice (2 ER, 1 ERL) and sorghum (2ER and 1ERL), sequence contigs were confirmed in barley for one ER (3500 bp) and one ERL (4500 bp), in chickpea for one ER (4300 pb), in bean for one ER (4300 bp), in potato for one ER (3300 bp) and one ER-L (2000 bp), and in cassava for one ER gene (4000 bp). A good overlap exists between the different contigs identified. Currently genes and species primers sets have been defined for all 7 species which will allow sequencing of different genotypes (7-8 per species). Sequence variation in ER and ERL genes has been already identified for 3 rice genes: 1 SNP/182 bp in ER1, 1 SNP/305 bp in ER2 and 1 SNP/275 bp in ER-L. Sequence diversity was tested on a first batch of 96 accessions in rice. In one fragment, 3 SNPs were identified at in identical positions previously identified on 7 accessions. High throughput sequencing on the 280 rice accessions is now operational and the sequencing of the first PCR fragments of the ERECTA genes is underway.

DREB1A and DREB2A

The consensus primer definition using the CODEHOP methodology failed, but the strategy was then modified in order to retrieve corresponding sequences in the seven species. This consists in the reconciliation of species tree with gene tree, design of primers for each clade of the tree and primer optimisation. By using this methodology and species specific primers, putative DREB2A orthologs have been identified in rice, barley, sorghum, chickpea and common bean up to now. Sequencing of putative DREB2A orthologs in 8 genotypes of each of four species mentioned above were used for SNP analysis and constructing the phylogenetic tree. In case of rice and sorghum there was occurrence of one SNP at the position 977(C/A) and 126bp (T/C), respectively as per reference accession. One SNP (G/C) was observed in case of chickpea promoter region and there was no SNP found in case of chickpea CAP2 gene. Common bean DREB2A accounted for maximum number of SNPs, i.e., eight SNPs at positions 214 (T/A), 219 (T/A), 233 (C/A), 235 (C/G), 404 (T/C), 490 (G/A), 595 (T/C) and 598 (T/C); these 8 SNPs constituted two distinct haplotypes. Efforts to generate the DREBA2A amplicons in cassava and potato are in progress. DREB1A giving less clear results has been abandoned.

Tangible outputs delivered

- Lists of selected reference collections of 283 to 300 accessions for diversity analysis in seven crops (rice, barley, sorghum, bean, chickpea, potato and cassava), available for further genotyping and phenotyping proposals.
- Sequence collections of candidate genes in the seven species under consideration, representing six gene families
- Primer sequences (consensus and specific) allowing the amplification of candidate genes in the seven ADOC species and related species.
- First SNP discovery for candidate genes

Deviations from the workplan
The strategy based on consensus primers proposed in order to complete gene families in the seven ADOC species considered allowed identifying some new sequences, but at this stage we cannot affirm that the gene families are completed. We will stop this work very soon and consider different ways of defining orthology / paralogy relationships between the different members of gene families identified up to now. The reconciliation of species tree with gene tree is used to define DREB2A orthologs. For the other gene families, sequence alignments and orthology pipeline (using DORIO software) will be done by using only –fully sequenced genomes (rice, sorghum, Arabidopsis, poplar and grape), orthology relationship considered based on this information, and each of our target genes will then be assigned to an orthology group. Sequence allelic diversity will be studied for all members of gene families when possible, even if the whole sequence cannot be studied for very large genes. In this case, we will consider overlapping fragments of functional domains and/or specific 3' regions of the genes. Functional information will be added when available on phylogenetic trees, and comparative mapping data used when available.
As soon as SNP data will be delivered by OL4, they will be analysed by Ortho-labs in relation with crop-partners, on the basis of crop-clusters as initially planned (OL1 for cereals, OL2 for "root and tubers" crops and OL3 for legumes), and OL4 will complete a few missing data for all species to facilitate the global analysis in relation with gene specialists.

Data availability
The list of selected reference collections with passport information will be made available on the GCP central repository.
Sequence alignments and primer information need some more work before they will be integrated in the GCP database. It should be available in September for the ARM meeting.
SNP information will be integrated in a first step in a local database made available by OL4, and then integrated in the GCP repository at the end of the project, when analysis will be completed.

2006-03: SNP analysis of the genetic diversity along the rice genome (HAPLORYZA)

Principal Investigator:
K. McNally, IRRI
Collaborators:
C. Billot, Agropolis-CIRAD
B. Courtois, Agropolis-CIRAD
A. F. Abdelkhalik, Agropolis-CIRAD
G. Second, Agropolis-IRD
D. Brunel, INRA-CNG
M. Lathrop, CNG

Mid-year report

- Since October 2006, all DNA samples have been sent to CNG, France. They consist of 900 samples that include a core-collection (MiniGB), completed by samples taken from the GCP reference collection, three collections of accessions suspected to carry indica-japonica introgressions (from Madagascar, Guinea, Surinam) and wild samples. Caution has been taken to send enough DNA quantity as well as good quality DNA.
- The final set of 1536 was chosen from those with the highest Illumina quality scores such that 357 Mb of the genome is covered by 1 SNP per 320 kb and the remaining regions are covered by 1 SNP per 50 kb. The specific regions include 8 regions carrying candidate genes for drought of about 700 kb each ((totaling about 5.9 Mb), the short arm of Chr. 6 (15.4 Mb), and two regions of low SNP density on Chr. 7 and 2 regions of normal SNP density on Chr. 12 (~ 4 Mb).
- The Illumina chip including 1536 SNPs has been designed and produced by Illumina, and sent to CNG.

- First tests concerning DNA quality and concentrations have been performed.

Tangible outputs delivered

- A complete list of DNA is available.
- All DNA samples are available for analysis at CNG.
- All SNPs have been defined, were chosen regularly along the rice genome, including some more detailed areas, and are fully documented (exact location, inference in coding sequences, etc.).
- First tests of the SNP panel showed that
    a. DNA concentration of a sample should not be below 15 ng/ul in order to obtain sufficient PCR products for reliable detection.
    b. A large proportion of the SNPs will be usable (over 90%). These include good quality SNPs and polymorphic SNPs. The low level of monomorphism confirms that SNP were defined accurately, and that the indica sequence (93-11) shows a quality high enough to be used as a reference.

Deviations from the workplan

Determination of SNP primers was delayed due to Illumina scoring. However, since then, the experiments are going normally, and a first complete dataset should be available at the end of June 2007.

Data availability

Information regarding the 1536 SNPs chosen for the Illumina chip is stored in local databases at CIRAD and CNG. The complete dataset is not available yet. This dataset is expected by the end of June 2007.

2006-04: Phenotyping in the field: Global capacity accessible to the GCP – Inventory of phenotyping resources and capacity for the GCP

Principal Investigator:

Mahalakshmi Viswanthan, Consultant

Collaborators*:*

Abraham Blum, Consultant, Plantstress.com

Bioversity Contacts: Samy Gaiji, Project Coordinator, Bioversity Informatics

Toby Hodgkin, Director, Global Partnerships Programme

GCP consortium members**:** Challenge Programme on Water for Food

Mid-year report

Tangible outputs delivered

Based on the information on the drought traits available from the phenotyping workshop held in Montpellier, France in 2004 and from published literature a set of traits and methods were identified. As the drought phenotyping procedures span across multi-location field trials to glasshouse and growth chamber studies, two questionnaires were developed under Adobe Forms application. The questionnaires capture all the needed relevant information required to assess the phenotyping resources in terms of science, local capacity, methods and human resources. Brief description of the procedure adapted and more specifically information on published data is also incorporated into the questionnaire. Large scale phenotyping will occur essentially on the most prioriy crops (for potential impact of the GCP), which at this stage are: maize, rice, wheat, pearl millet, sorghum, groundnut, cowpea, chickpea, cassava and sweet potato (to be confirmed in the next strategy document of the GCP).

Both field and non-field questionnaire were sent to over 107 GCP and other potential researchers in the area of drought. Not all those addressed were positively known to carry out active phenotyping projects. So far we have a total of 23 field and 13 non-field returned questionnaires from a total of 18 locations.

CG centres which have responded include IRRI, CIMMYT, ICARDA, ICRISAT, CIAT and IITA. Others who have responded include China, South Africa, (consortium members), Australia, India, USA, Spain, Turkey, Thailand and Nigeria. Crops covered include wheat, bean, barley, rice, maize, groundnut, sorghum and cowpea. We will be studying the response more closely to check if any further information is required from the respondents. Database structure has been worked out. The information will soon be collated into the database. The data and information will be made available for users and also potential resource institution to update. We are currently seeking permission to make the information public from those who contributed.

Some key crops phenotyping is missing e.g., pearl millet, chickpea, cassava etc.

Deviations from the workplan
No deviation from the workplan but there is delay in starting of the project. Now that we are told about an impending funding constraint we will be seriously assessing the need to visit key locations and use more judiciously the funds to go those locations where the need is most.

Data availability
Questionnaire can be made available on the GCP website. Database is expected to be available only at the end of the project

2006-05: Development of a composite collection and the genotyping of faba bean

Principal Investigators:
Bonnie Furman, ICARDA
Michael Baum, ICARDA
Collaborators:
G. Duc, INRA-UR LEG,
M.J. Suso, Instituto de Agricultura Sostenible

Mid-year report
A global composite collection of 1000 accessions of faba bean was developed at ICARDA in collaboration with INRA-Dijon and IAS-Cordoba. The composite collection contains 505 accessions from the ICARDA global collection and 250 accessions from IAS and 245 accessions from INRA. The ICARDA and INRA materials have been grown and DNA extracted. IAS materials will follow in the growth chamber soon.

We have started to genotype accessions with 10 labeled primers and are scheduled to finalise the complete set of accessions for 20 functional primer pairs before the end of the year.

Tangible outputs delivered:
- A composite collection of 1000 accessions has been developed
- 33 functional SSR markers have been identified (Table 1)

Data availability
We have developed the core collection of 1000 accessions. Functional SSR markers for faba bean are also available now. We expect that we will finalize the genotyping of faba bean accessions before the end of 2007. Data are available under:
http://gcpcr.grinfo.net/index.php?app=datasets&inc=dataset_details&dataset_id=622

Deviations from the workplan
None

2006-06: Genotyping composite collection of finger millet [*Eleusine corcana* (L.) Gaertn]

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborators:
RK Varshney, ICRISAT
D Hoisington, ICRISAT
CLL Gowda, ICRISAT
CT Hash, ICRISAT
S Chandra, ICRISAT

Mid-year report

The finger millet composite collection (1000 accessions) was planted in field and DNA from single plants extracted at 22 days after sowing. A total of 31 SSR markers were obtained from the University of Georgia (Dr. K Devos) and initially screened on 8 diverse genotypes to identify 20 polymorphic markers. Five multiplexes for 20 markers have been optimised. PCR amplifications were performed for all genotypes with 19 primers. DNA fragments were denatured and genotyped using capillary electrophoresis on an ABI 3700 DNA sequencer. The electrophoretic data for all the 19 markers are being analysed. Genotyping of the remaining one SSR is in progress. All genotyping data will be analysed using appropriate computer programme and called allele data will be generated.

Tangible outputs delivered

- Genotyping completed on 1000 accessions with 19 SSR markers.
- Information on allele size available.

Future plan of work

- Genotyping the composite collection with remaining one marker.
- Data analysis for allele calling to procure information on population structure and genetic diversity.
- Identification of a reference collection consisting 300 diverse accessions.

Deviations from the workplan

None

Data availability

After completing the genotyping of the composite collection with the remaining one SSR, the complete data set of 20 SSR loci on 1000 accessions of finger millet composite collection will be made available to GCP database following the appropriate quality checks.

2006-29: Preparing IITA-cassava reference germplasm for distribution and association mapping

Principal Investigator:
Dominique Dumet, IITA
Collaborators:
Morag Ferguson, IITA
Martin Fregene, CIAT
Odu, Babajide, IITA

Mid-year report

Progress since last report

183 accessions out of the 198 selected have been introduced in vitro (+16% in comparison to last report - 92% achieved). Indexing has now started and is in progress for 119 accessions. It includes 2 virus screening levels: One on *in vitro* seedlings, the other one on acclimatised plants. So far, 109 accessions were found virus free after the first test. These accessions were sent to acclimatisation and, so far, 31 have been recorded virus free. They will be certified virus free by the Nigerian Plant Quarantine Service shortly. 10 accessions were rejected after the first screening. They will be re-introduced in vitro after thermo-treatment.

Tangible outputs delivered

Cassava In vitro collections certified virus free and ready for distribution/use

Deviations from the workplan

As mentioned before, the project only started in April/May 2006 – A no-cost-extension was agreed and the project should now end mid-October 2007.

Data availability

The collections should be ready for distribution by mid-October 2007

2006-30: Development and genotyping of composite collection of foxtail millet [*Setaria italica* (L.) Breauv].

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborators:
RK Varshney, ICRISAT
CT Hash, ICRISAT
D Hoisington, ICRISAT
CLL Gowda, ICRISAT
S Chandra, ICRISAT

Mid-year report

The foxtail millet composite collection consists of 500 accessions (451 landraces, 10 improved cultivars, 27 *Glauca sp.*, and 12 *Viridis subsp.*). The accessions were planted in field and DNA from single plants extracted at 22 days after sowing. Twenty SSR markers (15 pearl millet and 5 finger millet) were initially screened on 8 diverse accessions. Six multiplexes for all twenty markers have been optimised.

PCR amplifications were performed for 5 SSRs on 500 accessions. DNA fragments were denatured and genotyped using capillary electrophoresis on an ABI 3700 DNA sequencer. The electrophoretic data are being analysed. Genotyping is in progress for the remaining 15 markers. All genotyping data will be analysed using appropriate computer programme and called allele data will be generated.

Tangible outputs delivered

- DNA extracted from the entire composite collection.
- 20 polymorphic SSR markers selected from pearl millet (15) and finger millet (5).
- Genotyping completed on 500 accessions with 5 SSR markers.

Future plan of work

- The entire collection will be genotyped with remaining 15 markers.

- Data analysis to determine the population structure and genetic diversity.
- Identification of reference collection consisting 200 diverse accessions.

Deviations from the workplan:
None

Data availability
After completing the genotyping of the composite collection with the remaining 15 SSRs, the complete data set of 20 SSR loci on 500 accessions of foxtail millet composite collection will be made available to GCP database following the appropriate quality checks.

2006-31: Development and genotyping of composite collection of pearl millet (*Pennisetum glaucum* (L.) R. Br.)

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborators:
CT Hash, ICRISAT
S. Senthilvel, ICRISAT
RK Varshney, ICRISAT
D Hoisington, ICRISAT
KN Rai, ICRISAT
RP Thakur, ICRISAT
S Chandra, ICRISAT

Mid-rear report
The DNA extracted from bulked tissues (15 plants/accession) of 1023 accessions of pearl millet composite collection were quantified using picogreen® through Tecan® Spectro Fluor Plus plate reader and normalised to a uniform concentration of 5 ng/ul. A set of 20 pearl millet SSRs selected previously based on their ability to detect allele frequencies in a linear manner in bulk DNA samples were used to genotype the 1023 accessions through capillary electrophoresis performed with ABI 3700 Genetic analyser. The fragments were sized by Genescan® softwares and the allele calling was done by Genotyper®. The size of the fragments (in bp), peak height and peak area were scored for each accession. Up to 6 alleles were detected per locus. We are currently working to determine relative frequency of each allele in an accession based on their respective peak height. Allele calling for three markers was extremely difficult with bulked DNA samples due to poor amplification profiles. The possibility of repeating those marker analyses or using different primer pairs is being explored. The data is being curated for analysis.

Tangible outputs delivered
Raw genotypic data for 17 SSRs on 1023 accessions of pearl millet composite collection generated.

Deviations from the workplan:
None

Data availability
The data to GCP will be made available as and when complete data set is generated.

2006-32: Molecular characterisation of pigeonpea (Cajanus cajan L.) composite collection

Principal Investigator:
HD Upadhyaya, ICRISAT
Collaborators:
R Bhattacharjee, ICRISAT
D Hoisington, ICRISAT
S Chandra, ICRISAT
RK Varshney, ICRISAT
KB Saxena, ICRISAT

Mid-year report

The pigeonpea composite collection, consisting of 1000 accessions, was established at ICRISAT using the available phenotypic characterization and evaluation data (Table 1). The collection comprised accessions of pigeonpea mini core, comparator mini core, and representative accessions of landraces, breeding lines, genetic stocks, wild species and four control cultivars. At ICRISAT, this composite collection was genotyped using 17 SSR markers (Table 2). The data on remaining 3 markers is being generated. A fluorescent-based multiplex genotyping system was used to generate multiplexes that were used to fingerprint the composite collection. The amplified PCR products were separated by capillary electrophoresis in an ABI 3700 DNA Sequencer. SSR fragment sizes were called to two decimal places using the Genotyper v3.7 software. Final allele sizes were determined using the Allelobin software.

Using the existing data on 17 SSRs, a total of 184 alleles have been detected, ranging from 2 to 24 alleles with an average of 10.8 alleles per SSR locus (Table 2). Mean PIC value of 0.332 (ranging from 0.002 to 0.589) and gene diversity 0.360 (ranging from 0.002 – 0.649) (Table 3) was observed in the composite collection. Among the two biological types, wild types were more genetically diverse than cultivated types. Gene diversity ranged between 0.000 and 0.784 in wild types, and 0.002 and 0.646 in cultivated types. Accessions from Asia region 6 revealed high average gene diversity (0.542) while those from Europe revealed low gene diversity (0.206) (Table 3). Detection included 90 rare and 94 common alleles at 5%, and 37 rare and 147 common alleles at 1% level in the composite collection. Principal coordinate analysis delineated the accessions in two clusters. The cultivated and wild pigeonpea each formed two distinct clusters; however, a number of cultivated types also grouped with wild types indicating that cultivated types have originated from the wild pigeonpea.

Tangible outputs delivered

- Genotyping completed on 1000 accessions with 17 SSR markers.
- Information on allele size available.
- Preliminary data analysis completed.

Deviations from the workplan

None

Data availability

After completing the genotyping of the composite collection with the remaining three SSRs, the complete data set of 20 SSR loci on 1000 accessions of pigeonpea composite collection will be made available to GCP database following the appropriate quality checks.

2006-33: Development and genotyping of a composite germplasm sample of potato

Principal Investigator:
Marc Ghislain, CIP
Collaborators:
Jorge Núñez, CIP
Maria del Rosario Herrera, CIP
Guillermo Trujillo, CIP
Reinhard Simon, CIP
Edwin Rojas, CIP
One student, CIP

Final project report
Executive summary:
A large fingerprinting of potato landraces and breeders material was developed with 51 SSR markers which resulted in revisiting the proposed potato genetic identity (PGI) kit made of 18 SSR markers in 2004. The new SSR markers were mapped and characterised by their polymorphic information content (PIC). These results allowed us to select a new PGI kit based on the following criteria: 2 SSR markers per chromosome, high PIC value, and high quality of amplicons. The new PGI kit consists of 24 SSR markers. The comparison of similarity matrix of 531 landraces genotyped with the new PGI kit and with 51 SSR markers resulted with a high correlation (r=0.93) by the Mantel test. The new PGI kit is able to discriminate 96.99% of the 531 landraces, whereas the 51 SSR markers 99.06%. A SSR-specific ladder has been developed for all SSR markers of new PGI kit (Herrera et al., 2006). Selected advanced potato cultivars developed by the CIP breeding programme were genotyped using 51 SSR markers. The aim was to compare the gene pool used by potato breeders with the diversity of existing landraces. An analysis performed with 21 SSR showed that all the advanced cultivars (except two genotypes) are grouping together into a well-defined cluster with landraces of the Chilotanum Group. Such grouping is expected because the germplasm of the Chilotanum Group has been used extensively in potato breeding worldwide. Noteworthy, eight out of 185 SSR alleles (4.9%) were found only in the advanced cultivars. This extra allelic diversity could be due to the poor representation of the Chilotanum diversity in the CIP genebank or to the presence of SSR alleles from potato wild species. We favor the latter hypothesis because it is supported by pedigree information, and the Chilotanum material is a representative sample of an original group of 144 landraces, which we believe is fairly representative. The high PIC value (0.728) found in the advanced cultivars is comparable to those of groups of native potatoes. These results illustrate the high level of heterozygosity of the founder population of this breeding material as well as the introgression of DNA from potato wild species (Núñez et al., 2006a).

G4007.01: SP1 composite sets genotyping data: quality assessment and consolidation

Principal Investigator:
Jean-Francois Rami, CIRAD
Collaborators:
Dave Hoisington, ICRISAT
Marylin Warburton, CIMMYT

Mid-year report
The identification of validation labs is in progress.
Three validation labs have been identified among the GCP consortium members:

- CIMMYT
- ICRISAT

- AGROPOLIS-INRA-Genotyping platform of Clermont Ferrand.

The identification of an external validation lab is still pending.

The distribution of genotyping task between the four validation labs is in progress (see attached table)

Tangible outputs delivered
Not yet applicable

Deviations from the workplan
The identification of validation labs took more time than expected due to disengagement of SCRI and staff movements in CNG, both being external validation labs initially identified.

Data availability
No data produced for now

Subprogramme 2: Comparitive genomics for gene discovery

2005-09: Systematic evaluation of rice mutant collections for conditional phenotypes with emphasis on stress tolerance

Principal Investigators:
Andy Pereira, WUR
New AffiliationVBI, Virginia Tech
Collaborators:
PIs:
Hirohiko Hirochika, NIAS
Hei Leung, IRRI
Emmanuel Guiderdoni, Agropolis
Mathias Lorieux, IRD/CIAT
Manabu Ishitani, CIAT
Tiegang Lu, CAAS
Qifa Zhang/Lizhong Xiong, HAU
Co-PIs:
Gyn An, Pohang Univ. Science and Technology
Srinivasan Ramachandran Temasek Lifesciences Laboratory
Narayana Upadhyaya, CSIRO, Australia
Venkatesan Sundaresan, UC Davis

Final project report

Executive summary:
The rice genome sequence provides the basic framework for the functional analysis of monocot genomes. International efforts have produced rice mutant resources that are a powerful functional genomics tool to identify the function of the sequenced genes. This consortium proposes to create a platform to identify genes that can contribute to a phenotype of resistance/tolerance to abiotic/biotic stresses. Expression analysis of various plant genomes has revealed genes that respond to environmental stresses. In addition, the ongoing functional analysis of Arabidopsis and other model plants identifies genes and associated mechanisms involved in stress tolerance. This comparative genomics information will be integrated to identify a set of candidate rice genes predicted to be associated with drought and other stress responses.

In a reverse genetic strategy, we use sequence indexed knockout mutant resources developed by the rice international community to identify insertions in the target stress associated genes. These insertions include rice Tos17 transposons, heterologous Ac-Ds transposons and T-DNA inserts that are identified by available flanking sequences (FSTs), supplemented by PCR based screens for inserts. In addition TILLING is used to supplement the mutant coverage, and also provide non-transgenic stocks for field testing. To resolve gene redundancy, gain-of-function overexpression lines are generated by transformation of appropriate constructs, and available activation tag populations are also accessed for overexpression mutants.

To facilitate novel gene discovery a forward genetics mutant screen is carried out with random genotypes that have been sequence indexed, thus aiding their further analysis. The mutant genotypes are phenotyped for drought and disease stress parameters at appropriate growth stages in greenhouse and field based screens. Genotypes with altered stress tolerance phenotypes will are also tested for other abiotic/biotic stresses. Whole genome microarray analysis will be conducted to identify the downstream genes and characterise the stress response mechanism involved. The genes can be utilised directly by transformation to provide stress tolerance, or for cereal comparative genomics studies and allele mining for breeding.

2005-10: Wheat genetic stock assembly and utilisation

Principal Investigator:
Tom Payne

MID-YEAR REPORT NOT SUBMITTED

2005-11: Legume mutant resource development

Principal Investigator:
M. Blair, CIAT
Collaborators:
W. Broughton, University of Geneva
P. Lariguet, University of Geneva

Mid-year report
In the TILLING project for common beans we have made progress in two areas:

1) Mutagen Dose: We returned to the mutagen dosage issue as we had run into problems with low mutagenesis rates in recent lines from our collaboration with the Univ. of Geneva. As follow up at CIAT, we completed two ethyl methane-sulfonate (EMS) mutagenesis experiments with 5 treatment levels using BAT93 seed planted in the greenhouse and a protocol similar to one used in barley (Caldwell et al; 2004) that involves imbibing seeds with distilled water (8h) and then soaking them overnight (16 h) in EMS solutions.

Our first and second experiments included the doses in table 1 which are reported in terms of % and mM EMS given that various publications have used both methods of calculating concentration for the generation of common bean mutants through chemical mutagenesis. Each mutagenesis treatment involved treating 200 seeds, with a total of 2000 seeds mutagenized in both experiments. Germination rates were recorded for each treatment and the results are summarised in Table 2a and 2b. In general, the germination of treated seeds occurred later than seeds without EMS (Figure 1a and 1b) and germination was reduced with increases in EMS concentration. In addition, treated M1 seedlings showed morphological differences in comparison to control plants as has been seen for other mutagenesis programmes (Till et al; 2003)

Our experiments indicated that common bean is very susceptible to high concentration of EMS above 0.5 % and did best at concentrations between 0.1 and 0.3 % (Table1). These results agree with Yanhulov et al. (1980) who found that ideal concentrations were between 0.05 and 0.15 %; and with Barbosa et al. (1988) who found that concentrations of 0.25% yielded germination of 75 % and a high mutation frequency.

Table 1. Dosages used in the mutagenesis experiments with ethyl methane-sulfonate (EMS) in CIAT in 2006-2007.

Experiment 1 (Nov. 2006)		**Experiment 2 (Jan. 2007)**	
Treatment levels		Treatment Levels	
(% EMS basis)	(mM basis)	(% EMS basis)	(mM basis)
0.0	0.0	0.0	0.0
0.2	19.4	0.2	19.4
0.5	48.5	0.3	29.1
0.8	77.7	0.4	38.8
1.1	106.0	0.5	48.5

Table 2. Germination rate in ethyl methane-sulfonate (EMS) mutagenesis experiments.

a) **Experiment 1**				
Treatment	(% EMS basis)	(mM basis)	% germination (raw)	% germination (corrected)
1	0.0	0.0	94	100.0
2	0.2	19.4	65	69.2
3	0.5	48.5	1.5	1.60
4	0.8	77.7	0.4	0.43
5	1.1	106.0	2.5	2.66
b) **Experiment 2**				
Treatment	(% EMS basis)	(mM basis)	% germination (raw)	% germination (corrected)
1	0.0	0.0	57	100.0
2	0.2	19.4	34	60.0
3	0.3	29.1	19	33.3
4	0.4	38.8	4	7.00
5	0.5	48.5	1	1.70

2) CEL I testing: the CEL I nuclease we extracted last semester has been used to assay known SNPs in PCR products from allelic variants at the SR2 (SCAR), locus. We have used these initial experiments to establish detection platforms on both polyacrylamide and agarose using LICOR 4200 automated sequencers or EtBr staining, respectively. The activity of CEL I enzyme was compared to that of SurveyorTM enzyme, by means of digestion of the 632 bp G vs. C SNP from the Surveyor Kit.

Tangible outputs delivered

Our final goal for this part of the project was to augment the number of mutant families as we believed it is still necessary to develop a larger number of M1 plants even though we have reached the overall goal of over 3,000 M2 progeny. Therefore our tangible outpust have been:

- A mutagenesis protocol developed at CIAT that complements the one used by collaborators at Univ. of Geneva and another one being developed at USDA.
- M2:3 (1,152) and M3:4 (463) lines generated by the project.
- Dwarfing, leaf fasciation, leaf variegation, spindly growth and other visible mutants have been documented and photographed in both the greenhouse and the field growing cycles.
- DNA extracted of at least 2 individuals per family for genotypic screening.
- Publication submitted on mutagenesis goals of the project and one in preparation.

Deviations from the workplan

We have produced a large number of M2 progeny (3000 which was the goal of the project) and have increased or are increasing these by single seed descent to the M3 and M4 generations, however many of the lines are related within a more limited number of families. We have run into the expected problem of slow seed increase for common beans (unlike many cereals, common beans given their large seed size have a low multiplication ratio) and unexpected problems such as recent flooding in our production site and for this reason we are planting several M1:2 plants per family and single seed descents with remnant

seed to ensure we do not lose the line. A bottleneck is the production of M1 plants due to the high rates of deleterious mutants at EMS concentrations of 45 to 50 mM and the low rate of visible mutants when EMS concentrations are around 35 mM, therefore we have performed additional experiments to evaluate EMS concentration with previously imbibed seed soaked overnight in EMS solution. DNA extraction continues to be by a standard technique that gives good quality and we now have a over 1500 lines extracted but we are trying to work on a way to increase throughput. We have overcome the potential bottleneck of the extraction of CEL I endonuclease (McCallum et al., 2000) and have tested SNP detection on both acrylamide and agarose gels.

Data availability
A local database has been created for the mutant phenotypes identified so far and this is being put into a web-compatible format using Microsoft® Visual Basic for Excel.

2005-12: A saturated potato mutant population for functional genomics among Solanaceae and tuber crops

Principal Investigator:
Marc Ghislain, CIP
Collaborators:
Merideth Bonierbale, CIP
Alberto Salas, CIP
Maria del Rosario Herrera, CIP
Glenn Bryan, SCRI
Robbie Waugh, SCRI
Dani Zamir, Noa Issman - The Hebrew University of Jerusalem

Final project report
Executive summary:
A saturated mutant population of a wild potato species has been proposed as a genetic resource useful for forward and reverse genetics. For that purpose, the diploid tuber- bearing *Solanum verrucosum* Schlechtd. species was chosen because it is one of the rare self-compatible tuber-bearing species. Based on AFLP markers, two of eight *S. verrucosum* accessions of the CIP Gene Bank (SHGRF 4019 and TRHRG 161) appeared to be the most homozygous. Additional selfing has been performed at CIP on five plants from each accession to decrease residual heterozygosity. Additional selfing has been also performed on the CPC54 progeny at SCRI. Toxicity of ethyl methane sulfonate (EMS) for lethality of *S. verrucosum* seeds was tested both at CIP and SCRI at different doses. 0.5% EMS dose showed a LD of 20% and was selected as working dose. Phenotypic aberrations were observed on small samples of germinated mutagenised M1 seeds at both CIP and SCRI. Two thousands seeds from the progeny-3 of the TRHRG 161 accession (M0 seeds) have been mutagenized using the same EMS dose. The germination ratio was high and the seedlings have been transferred to a screenhouse at CIP Lima. In order to maximise flower production, the seedlings were established at higher altitude at CIP- Huancayo. However due to EMS treatment and infection with *Oidium*, the number of M1 plants decreased to1,082. Of these, only 382 plants presenting flowers were selfed. From these, approximately 65 000 M2 seeds have been obtained from 288 M1 plants. Additional M1 plants are currently being propagated to produce additional M2 seed stocks. A sample of this population will be screened for phenotypic mutations. Depending on the frequency of phenotypic mutations, the M1 plants will be increased or genotypic mutations will be assessed for specific traits. As a complementary activity of the potato mutant collection, an eggplant mutant collection is under development at the Hebrew University of Jerusalem. In addition a bioinformatics platform is now available to host future data on these future Solanaceae mutant collections.

2005-13: Crop gene expression profiles and stress-gene arrays

Principal Investigators:
Guozhen Liu, Beijing Genomics Institute
Shoshi Kikuchi, NIAS
Manabu Ishitani, CIAT

Final Project Report

Executive summary:

To understand the common stress tolerance mechanisms, we use wheat RNAs to hybridise BGI rice 60K rice Oligoarray. Plants from a drought tolerant wheat variety Hanxuan No.10 and a salinity tolerant variety tolerant variety 98160 were treated under different stresses. Cross species hybridisation showed that the hybridisation efficiency of wheat RNA to rice microarray is quite similar to that of rice RNA to rice microarray. About 50% of rice 60K sequences could be successfully hybridised with wheat mRNAs at the best. Wheat stress (drought and salinity) responsive genes have been identified by using wheat-rice cross species hybridisation.

To understand the diversity of physiological and molecular responses to drought in rice (Oryza sativa L), transcript profiling was done using 22K rice Oligoarray (Agilent technologies) in two contrasting genotypes namely, i) Apo, an up-land drought tolerant indica and ii) IR64, a popular high-yielding drought susceptible indica. At the whole-genome level, the tolerant Apo showed relatively more transcriptional changes than IR64 under drought stress. The two genotypes shared only a limited number of genes with common expression behaviour. Apo-specific upregulation include genes involved in general stress response, osmoprotection, cell wall growth, lipid metabolism and signaling. In addition Apo showed specific upregulation of several stress-related transcription factors carrying AP2, MYB, WRKY and NAC domains. Flavonoid biosynthesis and ROS scavenging pathways are specifically upregulated in the sensitive IR64. By viewing expression patterns through a sliding window of approximately 20 genes, we observed correlated expression of adjacent genes. We detected 14 regions of correlated expression in Apo and 5 in IR64. Some of the RCEs on chromosome 1, 7 and 8 were found to co-localize within the boundary of reported QTLs for drought tolerance, osmotic adjustment and cell membrane stability. Sequence analysis of upstream cis-element showed enrichment of ABRE and DREB elements in the genes specifically induced in Apo.

2005-14: Stress response-enriched EST resources for targeted species

Principal Investigator:
Sarah Hearne, IITA
Collaborators:
Morag Ferguson, IITA
Chris Town, TIGR
Richard Bishop, ILRI
Jean Hanson, ILRI
Jun Zhuang, TIGR (note TIGR has merged with other institutes to form the J. Craig Venter Institute)

Final project report

Executive summary:

Two normalised cowpea cDNA libraries comprising; 1, root and 2, leaf and stem tissue (including shoot meristem) were sequenced from the 5' end. The libraries comprised cDNAs from drought stressed and irrigated plants of the cowpea lines Dan Ila (type II drought tolerance), Tvu11986 (type I drought tolerance) Vu7778 (susceptible) and 12008D (a dual purpose cowpea line with good feed quality and reported drought tolerance) which differ in their drought tolerance as indicated. Some 41949 good quality sequences were generated with a mean read length of 581bp. Assembly and clustering into contiguous sequences yielded 16,954 unigenes of a mean size of 709bp comprising 7894 contigs (mean size 859bp)

and 9060 singletons (mean size 578bp). These data have been compared with the cowpea gene space sequences (GSS) of Michael Timko's research group in Virginia. In the final combined assembly a set of 49,831 unique sequences have been identified, these have a mean size of 505bp (max=4477bp and min=100bp) this set comprises 11241 contigs (mean size 838bp) and 38590 singletons (mean size 408bp). SSR and putative SNP's could be identified in both the EST and combined data set. Some 5172 potential markers have been identified from the EST data set alone; 3226 putative SNPs and 1805 SSRs (219 of which are polymorphic across the 5 genotypes in the combined EST/GSS data set). Primers have been designed for the SSRs. The output from this project increases the marker resource available for cowpea dramatically and will significantly improve our genotyping capabilities for diversity studies and MAS. In addition to the generation of markers, the combined data set is being utilised to generate a microarray chip to facilitate gene expression analysis, this will be available upon agreement from Michael Timko, Sarah Hearne and Richard Bishop.

2005-15: Targeted Musa genome sequencing and frame map construction

Principal Investigator:
Takuji Sasaki, NIAS
Collaborators:
Takashi Matsumoto, NIAS
Nicolas Roux, Bioversity International
Mathieu Rouard, Bioversity International
Isabelle Hippolyte, Agropolis-CIRAD
FrancChristophe Baurens, Agropolis-CIRAD
Frederic Bakry, Agropolis-CIRAD
Manoel Souza, EMBRAPA
Pat-Heslop–Harrison, University of Leicester
Jaroslav Dolezel, IEB

Mid-year report
Mapping SSR markers: (CIRAD)
Since December, we began the analysis of SSR genotyped on the BORLI population. We analysed the results using Join map software and we checked the results obtained using other softwares i.e. Mapmaker and Cartagène. These softwares analyse the data with different algorithms. They gave the same main grouping. The very first analysis showed the difficulty to construct a consensus map with both parents, because of translocation breakpoints probably different on Borneo and Pisang lilin linkage groups. Then, we did the analysis on each parent. At that stage, we pointed out the fact that the cross used for this map is very interesting on the point of view of translocation. As the translocation, are not bare by the same linkage groups, the comparison with the two maps will be very useful in testing and constructing the linkage groups.
Among the 178 heterozygous loci revealed on the progeny, 131 were heterozygous on Borneo, 158 on Pisang lilin. 108 heterozygous SSRs are common to both maps.

The expected linkage group number on *Musa acuminata* is 11.
Running Join map with a Lod score of 5 and recombination rate of 0.4 gave rise to 10 linkage groups (3 or more markers per group) on Pisang lilin and 13 groups (3 or more markers per group) on Borneo. On Pisang lilin 10 markers are grouping by pairs and 4 ungrouped, on Borneo, 4 markers are paired and 2 remained ungrouped. However the first group of Pisang lilin is probably an artificial grouping resulting from special pairing linked to translocation breakpoints. These preliminary results are shown on power point slides attached, and show that some points should be refined in relation to translocation breakpoints, markers order, and linkage group numbers.

For that purpose, we sent to Australia, half of the population to be analysed through DArTs technology in order to refine and, if possible, saturate the map produced. The results were sent last week, and we will then finish the analysis using these new data and those coming from RGAs analysis.

Defining BACs for sequencing through PCR and EST hybridisation approaches

Selection and analyses of BACs using biotic stress related genes: (CIRAD,UCB, Bioversity)
RGA08 probe have been hybridized onto banana BAC libraries MA4, MAC and MBP giving 11, 19 and 22 positive signals. One clone per library was selected based on fingerprint analysis and sent to NIAS for sequencing (MA4_52E23, MAC_91O16 and MBP_32N20). BAC sequences data have been treated through automatic, semi automatic and preliminary manual annotation and confirm the presence of multiple RGAs in all the 3 BACs. These analyses have been stored in a genome browser available on the GMGC website for the partners taking care of the manual annotation.

In order to determine if all the BACs are orthologous, Dot plot, Global blast analysis, BBMH on predicted proteins and phylogenetic analysis were performed.
Preliminary results are shown in figure attached.

Except the presence of RGAs in all the three BACs and putative orthologous relationships, the presence of a common gene in two of the three BACs (MA4_52E23 and MBP_32N20) and a truncated part of this gene in the third BAC (MAC_91O16) may indicate orthologous relationships between the BACs. Hybridisation with this gene and SSR analysis are under investigation for selecting homologous and orthologous BACs.

Selection of BACs using abiotic stress related genes: (University Leicester)
Nine clones were identified from A and B genome Musa BACs with heat, drought, salt and temperature-related genes homologous to those in rice, Arabidopsis and legumes. These were pipelined for sequencing, and additional candidates with identical homologues were identified for some of the genes, providing the opportunity in the future to generate contigs and longer sequences around the target genes.

Some 400 sequences representing nearly 200 retro-elements, nearly 200 repeats related to leucine-rich and NBS biotic stress genes, and about 20 SSRs and various duplications were analysed within the project.

In addition, work on genetic heterozygosity resulted in generation and analysis of about 70kb of DNA sequence from 6 Musa genotypes in genes, SSR and non-gene-non-SSR genomic regions. This was additional work to help determine Doubled haploid Pahang as the best candidate to further be sequenced.

BAC sequencing (NIAS)
New Musa BAC clones have been sequenced (the accumulated number was 15), but troubles were encountered in two of them: 99N22 and 125A12. In both cases, it was not possible to detect original sequences that were used for clone screening. We tried PCR amplification in our hands with the information of primer sequence, but no or very scarce amplification could be seen. Both clones could anyhow be sequenced, and 125A12 assembly could be submitted. But we could not assemble the 99N22 clone (ADH) sequence due to the heavy repeat sequences, spanning 60-70kb. Other clones are in the process to be sequenced. The list of BAC clones sequenced to date can be seen on the Global Musa Genomics Consortium Web site.

For the next batch of BACs for sequencing, we and Bioversity have arranged that we double-check clone identity via experiments prior to sequencing.

We have now received 10 new BACs, 4 by abiotic stress selection, 4 by microsynteny to rice genome in which both BAC ends hit the same region of rice genome, and two BACs selected by special interest (45SrDNA, and Cytogenetic marker). We have been informed the BAC end sequences for 4 microsyntenic clones, and compared them with the results of our own sequencing. All matched perfectly, so we have just started sequencing of 4 clones (3F3,3G18,4L11, and 9F20)in advance.

For the other clones we are waiting PCR primers information (we have to check one by one) for us to double-check.

Tangible outputs delivered

- 178 SSR markers on the BORLI map
- 15 BACs completely sequenced and 10 BACs in the process to be sequenced
- A poster on the biotic stress genes was presented, Diversity and evolution of NBS-type disease resistance gene homologues in Musa. Lead author Azhar Mohamad, at the European Plant Science meeting in Hungary, August 2006.
- A talk on the work was presented at the OECD meeting on Plant Domestication, Tsukuba, Japan October 2006.
- A talk was presented, as an observer-participant, at the IAEA/FAO programme on Musa and Cassava, Trivandrum, India, in February 2006.
- A public lecture on banana and genomic studies was presented at the University of Leicester in March 2007 (talk slides available at http://pathh.diinoweb.com/files/BananaBotanicGardenHeslopHarrison.pdf or www.biobanana.com)

PhD thesis of Azhar Mohamad, Genome organisation, evolution and biodiversity in Musa: Application to stress-related gene discovery and plant breeding http://pathh.diinoweb.com/files/Azhar_PhD_2006.pdf
A presentation “DNA and Genome Diversity in Bananas and Plantains (Musa)“ will be made at the European Cytogenetics Association meeting, Istanbul, Turkey (lead author Azhar Mohamad).

Deviations from the workplan

Grouping difficulties coming from translocation breakpoints and wait for new markers to help in the linkage group definition and saturation.

As an additional section, work was done to investigate polymorphism levels in genes, SSR, and anonymous genomic regions to measure heterozygosity levels in various Musa accessions. This work has concluded that a double-haploid is ideal for genomic sequencing, that SSR marker primer positions s can be selected to be cross-species transferable.

There is a delay in sequencing BAC clones within the deadline, because it took some time to choose the reasonable gene of interest, and select BACs from BAC library. We have been working on double-checking the selected clones, and continue shotgun sequencing up to a defined number.

Data availability

- Heterozygosity data being prepared for database submission.
- BAC sequencing data submitted to the GMGC database as available.

2005-17: Integrative genetic framework for comparative QTL mapping for drought tolerance in beans (CIAT and CIRAD/Agropolis)

Principal Investigators:
M. Blair, CIAT
N. Ahmadi, CIRAD / Agropolis
Collaborators:
M. Ishitani, CIAT
B. Courtois, CIRAD / Agropolis
J.F. Rami, CIRAD / Agropolis
A. Ghesquière, IRD, Agropolis
C. Tranchant, IRD, Agropolis

Mid-year report:
CIAT

We are wrapping up the development of markers for the DREB genes. Last year we reported on the evaluation of DREB1 and this year we have wrapped up the analysis of DREB1 and concentrated on DREB2 for which there are three copies identified so far (that we have called 2a, 2b and 2c). The following results are reported:

A) DREB1

Two PvDREB1 homologues were cloned from PCR reactions that used degenerate primers to isolate DREB candidates from common bean DNA. Degenerate primers were designed based on data-mining conducted with the CIAT and UNAM (Mexico) collections of ESTs (Ramirez et al., 2005) using the sequence information of DREB genes from *Brassica napus*, *Capsicum annum*, *Gosypium hirsitum*, *Helianthus annus*, *Oryza sativa*, *Glycine max*, *Arabidopsis thaliana* and *Lycopersicon esculentum* for the PvDREB2 homologues (see previous reports). The positive clones isolated with the degenerate primers were named PvDREB 1A and 1B. Genome walking experiments were conducted to obtain the full-length sequences for both genes. This allelic variants were sequenced for the parents of two mapping populations DOR 364 x G19833 and BAT 93 x Jalo EEP58. In addition the allele for BAT477, the parent of a drought tolerance mapping population, was also sequenced. Multiple sequence alignment were performed using Align X and Vector NTI (Invitrogen, Carlsbad CA) and were the foundation for SNP discovery and CAPS or dCAPS marker development.

Sequence comparisons for PvDREB 1B showed complete identity between allelic variants, in consequence no polymorphism for marker development was achieved and this gene will be difficult to map. In contrast, PvDREB 1A contained ten useful SNPs and a 9 bp insertion recognizable in the total sequence length which was present in Andean genotypes (G19833 and JaloEEP58) but absent in Mesoamerican genotypes (DOR 364 and BAT 93).

Based on the sequence information for PvDREB 1A, two different strategies were used for primer design. The first one was based on PCR primers flanking the 9 pb indel while the second was dCAPS design based on SNP2CAPS parameters (http://pgrc.ipkgatersleben.de/snp2caps/). Amplifications of the full length sequences were also used as a product for CAPS marker digestion and the conditions of digestions were set according to the enzyme.

Genetic mapping based on the indel allowed us to locate PvDREB1A in two mapping population (DOR 364 x G19833 and BAT 93 x Jalo EEP58) on linkage group b04 and the position was confirmed using the CAPS markers based on two restriction sites (*Alu*I and *Rsa*I). The PvDREB 1A indel marker and the PvDREB1A-*Rsa*I marker were located at almost the same position, but PvDREB1A-*Alu*I was loosely linked to these two markers. The mapping incongruity would most likely be explained because of partial digestion by *Alu*I.

The remaining ten SNPs for PvDREB1A were mainly distributed between the Andean and Mesoamerican genotypes and only one, located close to the 5' end of the sequence, was found distinguishing BAT 477 and DOR 364. A dCAPS marker was designed for this SNP but it was not successful and no further marker development was attempted given the difficulty of designing a dCAPS marker in this position. Therefore, the map position of PvDREB 1A in BAT 477 x DOR 364 population remains undiscovered.

B) DREB2

Three PvDREB 2 homologs were identified based on a database search of the common bean ESTs and were named PvDREB 2A, 2B and 2C. Since the sequences were partial in each case, genome walking (DNA walking speedup premix kit from Seegene Inc) and clone sequencing were conducted to obtain full length sequences. In the case of genome walking, sequencing was conducted with the Sp6 and T7 primers flanking the multiple cloning site of the pGEM®-T Easy Vector System (Promega Inc.). Allelic variants were sequenced using ABI PRIM 3700 (Macrogen Inc. Seoul, Korea) for the parents of two mapping populations DOR 364 x G19833 and BAT 93 x JaloEEP58. All sequences were based on purified PCR products resulting from primer amplification of genomic DNA with DREB based primers. Sequence quality was checked with Phred 4.25. sequence assembly was performed in sequencher 4.1.2 (Gene codes Inc), sequence alignment was performed Align X of the Vector NTI (Invitrogen Inc.) as well as BLAST searches. Protein domain searches were performed with Interproscan (www.ebi.ac.uk/cgibin/iprscan). Marker development was based on single strand confirmation polymorphism (SSCP) analysis on MDE non-denaturing gels (Cambrex Inc.).

i) *Pv DREB 2A:* PvDREB2A sequence was extended by genome walking in the 3' and 5' direction using the DOR364 genotype but the promoter and transcription start site have not been uncovered. The partial length sequence (1273 nt) obtained mostly for the 3' side of the fragment was used to design PCR primers which resulted in a single-copy amplification product which when tested for polymorphism on SSCP gels was found to vary for the Andean and Mesoamerican genotypes. This allowed the PvDREB 2A gene to be mapped to linkage group b02 in both the DOR364 x G19833 and BAT93 x JaloEEP558 populations and in a comparabale position. We are continuing to analyse the sequence variability for this gene but the actual sequence in DOR364 for PvDREB 2A appears to have 293 aa in ORF +2. When we obtain further sequence information at the beginning of the gene we will be able to make a fuller comparison between genotypes and across orthologs of DREB genes. For now PvDREB 2A shows a 25 aa difference at the beginning of the protein compared with GmDREBb, nevertheless the two amino acid sequences arehighly homologous (76%) as is the similarity at the nucleotide level (E^{-97}).

ii) Pv DREB 2B: Full-length sequence (1528 nt) were also obtained for PvDREB 2B through genome walking with partial sequences obtained through re-sequencing. In the full-length sequence we were able to find the 3' and 5' UTRs as well as the TATA box (positions 247 to 254) and the PvDREB 2B coding sequence corresponded to 294 aa. Amino acid and nucleotide sequence similarity analysis for PvDREB 2B showed it was homologous with GmDREB b (58% and E^{-26}, respectively). No polymorphisms have been found for PvDREB 2B in the sequence information but SSCP analysis shows a size difference for the amplification fragments in BAT93 and Jalo EEP558.

iii) Pv DREB 2C: Full-length sequence (1829 nt) was also obtained for PvDREB 2C based on sequencing of the original clone for the original G19833 EST using Sp6 and T7 primers flanking the multiple cloning site of the pCMVSport 6.0 vector and also with a new internal primer. BLAST searches showed homology between GmDREB3 (819 nt) and PvDREB 2C in two different positions (E value 10-47 and 10-11). The original EST did not have the AP2 domain that is characteristic of DREB genes but this was found with the additional sequencing of the cDNA. The open reading frame for the PvDREB 2C gene was found to encode 318 amino acids.

iv) Additional work on PvDREB2 clones: The conserved AP2 domain and the divergent 3' sequences of two DREB genes have been used to create overgo probes for hybridisation screening of the G19833 BAC library with the final goal of determining the genomic arrangement of the DREB genes and the number of independent chromosomal regions containing these homologues. So far one positive BAC clone has been obtained for PvDREB 2B. The sequence information is also being used to conduct gene expression studies with real time PCR primers for the DREB genes from common bean. So far PvDREB 2B and 2A are expressed in roots under drought stress, while PvDREB 2C appears not to be expressed but this could be due to it being a smaller sequence than was assayed. Now that we have full-length sequence for DREB2C the gene expression work will be repeated. Finally, the DREB2 clones are being used by Dr. Shinosaki for functional analysis in a yeast two-hybrid system with an *in vitro* ERBP promoter (RD29a).

Tangible outputs delivered

1. Cloned PCR fragments and sequences
2. Poster at ARM of the GCP

Deviations from the workplan

Low polymorphism has been a problem for DREB gene analysis although we have performed genome walking to obtain AP2 flanking sequences and complete coding sequences. This has allowed us to find a limited amount of polymorphism in the inter gene pool comparisons for a few of the genes analysed so far (PvDREB 1A and 2A). We have now obtained full-length ORF sequence for PvDREB 1A, 1B, 2B, and 2C but we are still working on sequencing PvDREB 2A which we will be wrapping up in the next few weeks.

Data availability

Sequence data available on request

CIRAD/Agropolis

Activities related to output 3.1

- The Nipponbare x Kasalath map data were positioned on physical map (TIGR V4). Their position on the integrative genetic map has to be interpolated.
- About 1400 QTLs have been projected on the on physical map (TIGR V4). Their position on the genetic integrative map has to be interpolated.
- Exploratory meta-analyses of QTLs involved in root development were undertaken, using the QTL's position on the physical map under Biomercator software. The results should be confirmed with analysis using the QTL's position on the genetic integrative map as it is required by the meta-QTL-analysis method developed by Goffinet and Gerber (2000).

Activities related to output 3.2

- Development of high throughput phenotyping method for root development using agar solid medium.
- Phenotyping of root development of CSSLs population of koshikarie x Kasalath and Nipponbare x Kasalath crosses using the high throughput phenotyping method (ongoing)
- Genotypic screening of BC4F2 population for 3 target QTLs and development of BC4-F3 lines.
- Genotypic analysis of the 6 insertion mutant co-localising with QTL 9.

Tangible outputs delivered

- Integrative map available on (http://rice-brcdb.cines.fr)
- Data base on rice QTLs involved in roots development and other drought-related traits (http://orygenesdb.cirad.fr/).

- The position of the sorghum stay-green QTLs on a single sorghum consensus map. The position of the sorghum stay-green QTLs on the rice pseudo-molecules.
- Phenotypic data (root development) produced for 3 CSSL populations from Caiapo x IRGC 103544, koshikarie x Kasalath and Nipponbare x Kasalath crosses.
- Effect of QTL 9 for deep root development confirmed in BC3F4 NILs
- BC4F3 lines with narrow chromosomal segments (1 cM for QTL9 and 3cM for QTL2) carrying target QTLs produced.

Deviations from the workplan
None

Data availability
- Root development data for 3 CSSL populations from Caiapo x IRGC 103544 (O. sativa x O. glaberrima), koshikarie x Kasalath and Nipponbare x Kasalath
- Integrative map based on 178 RI lines of the IR64 X Azucena cross genotyped with 226 SSR markers and on some 1700 additional markers from previous IR64 X Azucena map and from other maps have been *in silico* mapped or interpolated in their respective intervals spanned by the framework SSR markers. The completion of the Web-availability (http://orygenesdb.cirad.fr/) of these data is ongoing.
- Inventory of information available in the literature on QTLs involved in rice roots development and other drought-related traits: more that 1400 QTLs for some 90 traits involved in drought tolerance have been identified. The afferent information (QTL position, confidence interval, …) were synthesised in a database (http://orygenesdb.cirad.fr/)
- Establishment of the genomic position (rice physical map, TIGR V4), of 1400 drought tolerance related QTLs. The completion of the Web-availability (http://orygenesdb.cirad.fr/) of these data is ongoing.
- Position of sorghum stay-green QTLs on the rice Rice physical map.

2005-35: Sequencing multiple and diverse rice varieties: connecting whole-genome variation with phenotype

Principal Investigator:
Kenneth McNally, IRRI
Collaborators:
Dave Mackill, IRRI
Richard Bruskiewich, IRRI
Hei Leung, IRRI
Renee Stokowski, Perlegen Sciences, Inc
David Cox, Perlegen Sciences, Inc
Diana Star, Perlegen Sciences, Inc
Jan Leach, Colorado State University
C. Robin Buell, Kevin Childs, TIGR
Detlef Weigel - Max Planc, Tubingen

Mid-year report
- Final choice of the 100 Mb of Nippon-bare was resolved in September 2006. This fraction of the genome was chosen by comparing the Nippon-bare masked for repetitive regions to the gene models from the Rice Annotation project and TIGR databases with regions spanning more annotated bases preferentially included. The distribution of selected sequences indicated that the overwhelming majority of 100 kb windows across the genome would include one or more regions for SNP discovery.

- The varieties included in the project are Nipponbare, Tainung 67, LTH, M202, Azucena, Moroberekan, Cypress, Dom sufid, Aswina, Rayada, Dular, FR13A, N22, Swarna, Sadu-cho, Pokkali, Minghui 63, Zhenshan 97B, SHZ2 and IR64.
- Re-sequencing by high-density oligomer hybridization was split into two phases, the development phase to optimise conditions involving 379 kb of unique sequences from a 684 kb region Chromosome 3 and the discovery phase involving the remaining 99.7 Mb.
- **Development phase:** Perlegen determined that long-range PCR (LR-PCR) was the most efficient means for preparation of targets (the same method used for their human and mouse projects). They designed a chip for the 379 kb and 76 LR-PCR primers to amplify these target regions. They hybridised targets for the 20 varieties to these chips during August 2006. An initial summary table for the development phase results was sent by Perlegen in October 2006. In January 2007, we received the complete data set for the development phase that included coordinates of the tiled sequences, LR-PCR primers, model-based SNP calls for the 20 varieties, derived sequences for the tiled regions, and pseudo-trace files for each strand for these regions. Perlegen identified 2123 sites in the Nipponbare reference genome at which 1 or more of the other varieties were polymorphic by their model-based approach. A total of 12,501 SNPs were called at these sites, and 99% were biallelic. This equates to about 1 SNP/200 bp of the genome. Initial analysis indicates that LD extends to about 300 kb in this region.
- **Discovery phase:** Perlegen designed 5 wafers to tile the 99.7 Mb. They also designed 13,582 pairs of LR-PCR primers to cover these regions. From December 2006-February 2007, hybridizations of the LR-PCR amplicons for each of the 20 varieties. For each variety, 13,582 PCR reactions were performed and those products corresponding to tiled regions on a wafer were pooled for hybridization. On April 11, 2007, IRRI received the dataset from Perlegen including SNP calls, coordinates of the tiled sequences, LR-PCR primers, sequence calls for the 20 varieties for the tiled regions, and pseudo-trace files for the Watson and Crick strands. Perlegen's model-based approach to calling SNPs identified 257,598 sites in the Nipponbare reference sequence where one or more of the 20 varieties was polymorphic. Combining this data with the development phase data results in 259,721 SNP sites. Summary statistics for the combined data are shown in the following table.

Chr	IRGSP r4 (bp)	Tiled length	Non repetitive	SNP Sites	SNPs /kb	Total calls	Clear	Ambig-uous	bi-allelic
1	45064769	15411868	13377445	36993	2.7653	739860	205653	140140	36156
2	36823111	12695346	11028257	27628	2.5052	552560	146965	101119	27012
3	37257345	13687037	12353186	28228	2.2851	564560	156412	110220	27506
4	35863200	10674305	9111561	22226	2.4393	444520	102068	96606	21615
5	30039014	9054116	7799270	20433	2.6199	408660	114697	89165	19842
6	32124789	9491255	8000505	25549	3.1934	510980	133891	103692	24945
7	30357780	8989073	7646272	14202	1.8574	284040	67925	82363	13787
8	28530027	8179660	6923419	16117	2.3279	322340	79672	85411	15625
9	23843360	6900725	5840458	15349	2.6280	306980	72399	72929	14955

10	23661561	6405335	5355995	12248	2.2868	244960	63611	69859	11847
11	30828668	8063435	6768998	21606	3.1919	432120	97304	106348	21031
12	27757321	7333621	6243138	19142	3.0661	382840	86771	84495	18622
All	382150945	116885776	100448504	**259721**	**2.5856**	5194420	1327368	1142347	252943
							0.2555	0.2199	0.9739

- We are collaborating with Detlef Weigel's group at MPI-Tubingen for the data analysis and will apply the tools they developed for the analysis of SNP data from their Perlegen project on Arabidopsis.
- We undertook seed production for the 20 varieties over the dry season 2007 (December 2006 – March 2007). We have produced F1s from selected crosses in preparation for producing RILs for functional studies.

Tangible outputs delivered

We have obtained a dataset of SNP calls at 259,721 unique sites by Perlegen's model-based approach across the 20 varieties for the development and discovery phases. Perlegen also delivered associated files necessary for implementing further data analyses: coordinates of the tiled regions, long-range PCR primers, files of sequence calls for tiled regions, pseudo-trace files, etc.

Deviations from the workplan

Long-range PCR amplicons were used for preparing target DNAs for hybridisation to the wafers. This choice was based on the complexity of the rice genome and was the same approach used for Perlegen's SNP discovery for the human and mouse genomes.

Data Availability

Currently, the SNP data is being stored in local databases at IRRI, TIGR, and Max Planc – Tubingen. On finalizing the analysis, data will be made public via the OryzaSNP project site http://www.oryzasnp.org with links to the TIGR rice database. Projected date for public release is November 2007.

Subprogramme 3: Trait capture for crop improvement

2005-18: Development of low-cost gene-based trait assay technologies

Principal Investigator:
Casiana M. Vera Cruz, IRRI
Co-Principal Investigator:
Yunbi Xu, CIMMYT
Collaborators:
Jianli Wu, CNRRI
Joan Agarcio, PhilRice
Masdiar Bustamam, ICABIOGRAD
Usha Barwale Zehr, Mahyco Research Foundation
Valerie Verdier, IRD-Agropolis (in collaboration with NARS partners in Africa)
Martin Lagat, KARI
Chuan-Xiao Xie, CAAS
Firoz Hossain, IARI
Marcia Bunga Pabendon, ICERI

Mid-year report

I. Further technology development and optimisation of conditions for different technologies

A. Microplate Assay (PCR-ELISA^{TM})

The microplate assay mechanism involving oligonucleotide-specific hybridisation, as described by Osiowy, et. al (1998) and Knight, et. al (1999) was improved and validated to work with the *xa5* S and R probe. The 17mer oligonucleotides (5' biotin-labeled) containing the GTC=>GAG polymorphisms in the middle were designed to serve as capture probes. Gene-specific markers were used to amplify a region of the *xa5* gene containing the GTC=>GAG polymorphisms. The DIG-labeled PCR products (random-primed during PCR) were diluted 1:10 in sterile distilled water. The 5' Biotin-labeled probes specific to either the susceptible or resistant allele were allowed to hybridise with the PCR products at 37°C for 1.5hr. Probe-PCR product complex were immobilized on streptavidin-coated ELISA plates (Roche). Hybridization of PCR product and probe was detected using Roche's PCR-ELISA (DIG Detection Kit). Signals were only detected using the S probe but not the R probe. This work is being continued.
The probes for *Xa21* gene have already been designed and synthesized for evaluation and use in PCR-ELISA assay.

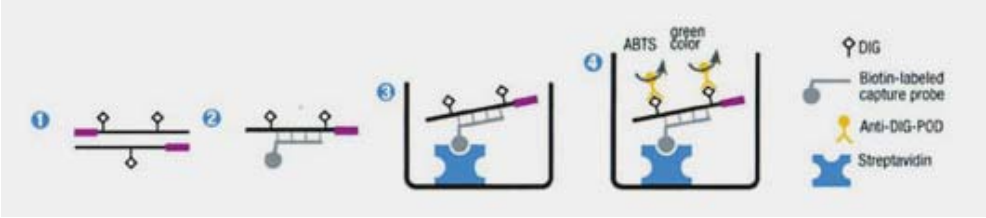


Fig. 1. Illustration of PCR-ELISA^{TM}: (a) PCR and DIG-labelling, (b) Hybridisation with biotin-labeled probe, (c) Immobilisation of probe-product complex on streptavidin-coated well, (d) Immunoassay using anti-DIG enzyme conjugate. PCR-ELISA is marketed by Roche. Source: http://www.roche-applied-science.com.

B. Dot-blot Assay

To further validate this assay, we used the previous optimised parameters and conditions to genotype developed population at the replicated yield trial (RYT), i.e. advanced elite lines. The RYT lines were grown and leaves were harvested for DNA extraction using the CTAB method resulting in good quality and high quantity of DNA. Although not all lines amplified using *Xa21* primers, dot blot has been successful. A phonotype of the lines against the races of the bacterial blight pathogen has also been done to complement the dot-blot results.

C. Microarray-based Genotyping Assay

New *xa5* probes for resistant and susceptible lines have been designed and synthesised for microarray-based genotyping of lines with and without *xa5*. To increase specificity of hybridisation, chaperone sequences have also been synthesized. We also designed a new DNA library for printing into the array, which included IRBB NILs and pyramids, IRBL NILs, popular megavarieties. Other popular varieties and germplasm used by other scientists at IRRI will be included. Most of these germplasm have already been requested.

Probes for *Xa21* gene have already been designed and synthesized for evaluation and use in MBG assay.

II. Evaluation, refinement, and deployment with NARS and SMEs

As one of the activities of the project is to transfer the technologies to our NARS and SME partners, a workshop has been planned, arranged and conducted with our project collaborator at Barwale Foundation through Dr. Dinesh Joshi and Dr. P. Kadirvel. Though the generous agreement of Dr. Usha Barwale Zehr, the Owner-Board Member of the Barwale Foundation, member of the IRRI and CIMMYT BOT, and member of the public-private committee of the CGIAR, the workshop was held at their own Barwale Knowledge and Study Center.

A. Barwale Foundation hosted the workshop

The excellent coordination and implementation of all arrangement for hosting the workshop by Dr. Dinesh Joshi, the Executive Director of the Barwale Foundation, the workshop was successfully conducted at the newly inaugurated Barwale Knowledge and Study Center located in Jalna, Maharashtra, India on 25-27 April 2007.

B. Laboratory Manual for the different technology platforms

A laboratory manual has been prepared for each technology, and stored in a zip file attached to this report. The manual consisted of the different technology platforms of microplate assay (PCR-ELISA^{TM}), dot-blot, fluorescence resonance energy transfer (FRET), microarray technology (microarray-based genotyping for MAS and Single Feature Polymorphism); bioinformatics section consisting of primer and probe design for SNP genotyping; 10 appendices consisting of list of reagents, solutions and other consumables for each technology platform, primers and probe sequences, calculation of optimal hybridisation and washing temperatures for oligonucleotide probes, methodology for competitive hybridisation adapted from Shirazawa et al. (2006), cost analysis; list and copies of recommended references; and list of suggested references.

Other information useful for the workshop were also included, such as workshop schedule, list of participants/resource persons/laboratory trainers/facilitators, and group assignments of participants during hands-on sessions.

C. Participants and resource speakers of the workshop

There were 18 participants who attended the workshop. They consisted of IRRI and CIMMYT NARES partners from Africa (2), China (2), India (2), Indonesia (2), and the Philippines (1). Seven participants from Barwale Foundation and Maharashtra Hybrid Seed Company and one participant each from India and Thailand, members of the Asia-Pacific Consortium on Agricultural Biotechnology coordinated by Dr. J.L. Karihaloo, also participated actively in the workshop.

Resource speakers from IRRI (3, self-funded) and IRD (1) were invited to present emerging technologies and platforms in the area of genomics, and their application in MAS for crop improvement. The workshop team members from IRRI (6) and CIMMYT (2) presented the concepts, principles and background of each technology, conducted the hands-on sessions, presented the lecture and computer hands-on for primer and probe design for SNP genotyping, and presented database management for high throughput technologies. A lecture on nonradioactive detection and fluorescence applications was also delivered by Mr. V. Voleti to supplement the lecture on FRET assay.

D. Validation of the technology by the participants

Using PCR-amplified DNA samples of breeding lines and control varieties brought from IRRI and CIMMYT, the 18 workshop participants validated the microplate assay, dot blot techniques and FRET assay technique.

1. Microplate Assay

For the microplate assay, samples used were developed lines (replicated yield trials) introgressed with the *xa5* bacterial blight resistance gene.

PCR products were hybridized with a 17mer 5' biotin-labeled probe complementary to the susceptible allele. Detection of the said allele was done using anti-DIG-POD Ab:ABTS interaction. Optical density was quantified using a BMG FLUOStar micropate (absorbance = A405nm). Susceptible lines can also be visually determined (Fig. 2). Further tests using these lines and control genotypes will be done to improve confidence on the data generated.

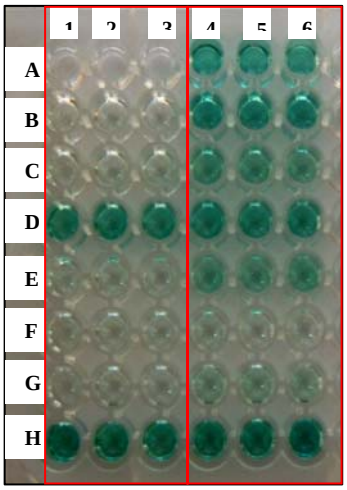


Fig. 2. ELISA plate showing detection of the susceptible *Xa5* allele. Samples were tested in triplicate. A1-A3: Blank; B1-B3: Negative control; C1-C3, Buffer blank; D1-D3: IR24; the rest, samples in triplicate.

2. Dot Blot Assay

For the dot blot assay, the technology was validated by the participants in groups of four using two sets of PCR products. The first two groups worked on the rice samples and used a set of recombinant inbred lines (RILs) containing *Xa21* bacterial blight resistance gene. The last two groups of participants worked on the maize samples, focusing on the *opaque2* allele.

Detection of the alleles was done using NBT/BCIP and blot analysis was visually observed (Fig. 3). Alternatively, to better handle and manage the data generated by this technique, a computer application called "ImageJ" was demonstrated to analyse a scanned image of the developed

membrane. This software not only improved data visualisation but also decreased possible errors during visual analysis.

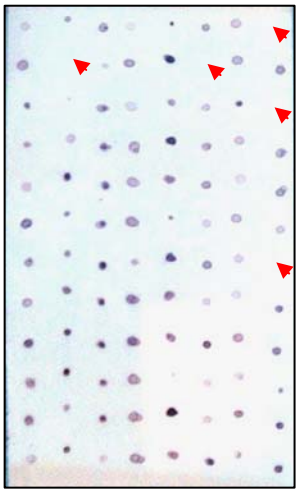


Fig.3. Membrane spotted with maize *O2* allele. Samples hybridised by the participants with wild type *O2* allele and detected using NBT/BCIP. Arrows indicate detection of mutant allele.

3. FRET Assay
No fluorescence reading was shown after the assay perhaps due to the instability of the reporter dye during transport of the labeled probe.

4. Microarray-based genotyping (MBG)
A demonstration of the manual hybridization of slide arrays with the probe was demonstrated to the participants. To supplement the demonstration session, a video presentation of the complete assay from printing of DNA to data scanning were presented and discussed in detail during the lecture session.

E. Workshop evaluation
A questionnaire was sent to the participants to assist us in evaluating the overall content of the workshop. So far, 50% of the participants responded. The summary of the workshop evaluation is included in the attached zip file (File 16).

III. Data management and LIMS (in collaboration with CRIL, IRRI)
A. Management for MAS data acquisition from high throughput technology platform
Data acquisition and management of genotype data are important components of gene-based technology development for correct identification and detection of target allele, especially when several hundreds or thousands of data are generated in a single assay such as in high throughput dot-blot assay or in microarray-based genotyping (MBG). Dot-blot and MBG are based on hybridization of SNP-based probes on glass or membrane filter. Being based on solid phase non-gel based genotyping system, unexpected false-positive or false-negative data may be generated on glass or membrane filters when applied on a breeding population. On the other hand, data generated from two liquid nongel-based genotyping techniques, i.e. microplate and FRET assays, could be generated by automated absorbance/fluorescence detection equipment, although it is also possible to obtain error during the experimental stage. In combination with the best condition for hybridization and detection for allele-specific probes, data management using adequate statistical analysis of quantified optical intensity in hybridized blot can be helpful. The threshold for positive and negative value can be obtained by comparison with positive and negative control and by calculating the average and standard deviation of selected spots.

In the case of dot-blot, genotyping and decision of false-positive spots could be done visually. But, if we use larger datasets, additional errors might be generated by improper position information, eye-balling error, or illusion of background.

Image analysis software called ImageJ (developed by Larry Reinking) was applied on dot-blot image to get numerical data for every spot, and make decision as to allele type of every genotype. Image processing by ImageJ allows relatively easy elimination of false positive spots by 'background substraction' and 'threshold' menu (Fig. 4). Further improvement on the use of the software is expected in consultation with R. Bruskiewich and J. Ulat of CRIL, IRRI.

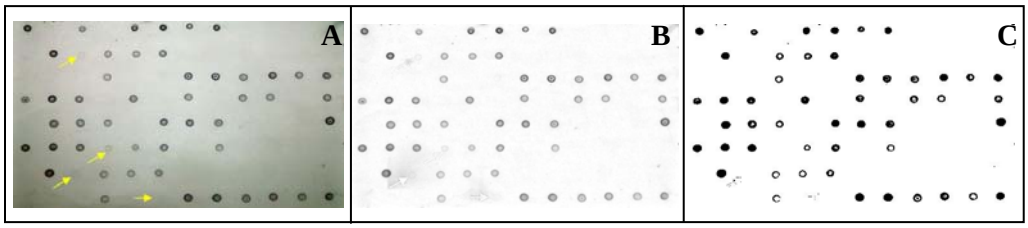


Fig. 4. Image processing for data acquisition by ImageJ of a previous dot blot data of *Xa21* R and S genotypes using Xa21R and Xa21S probes. A = raw data, B = after substracting background, and C = after using threshold.

A statistical analysis tool is available for deciding the allele-type of data generated by PCR-ELISA (microplate) and FRET technologies. Here, the data acquisition software has been developed and provided with the equipment. However, since the optical data show continuous data, it may be also possible to develop a module for data transformation.

B. LIMS (Laboratory Information Management System) development (CRIL, IRRI)

Regular meeting with Dr. Richard Bruskiewich and his team who are also involved in a GCP-SP4 project has been held since 19 December 2006. To assist us in this effort, we provided flowcharts of each technology that include the workflow of plant materials to be handled, e.g. DNA extraction of the parents, controls and breeding lines. Such information could be cross-linked to the information stored in IRIS/ICIS (Fig. 5). And following information will be provided 1) expected data for each step of the experiment; the output at each step will be used as a basis for mapping the information to the database tables in IRIS/ICIS or if necessary to add/create new tables/database; 2) provide output/data for each technology; 3) sample images of the results obtained from the different technologies.

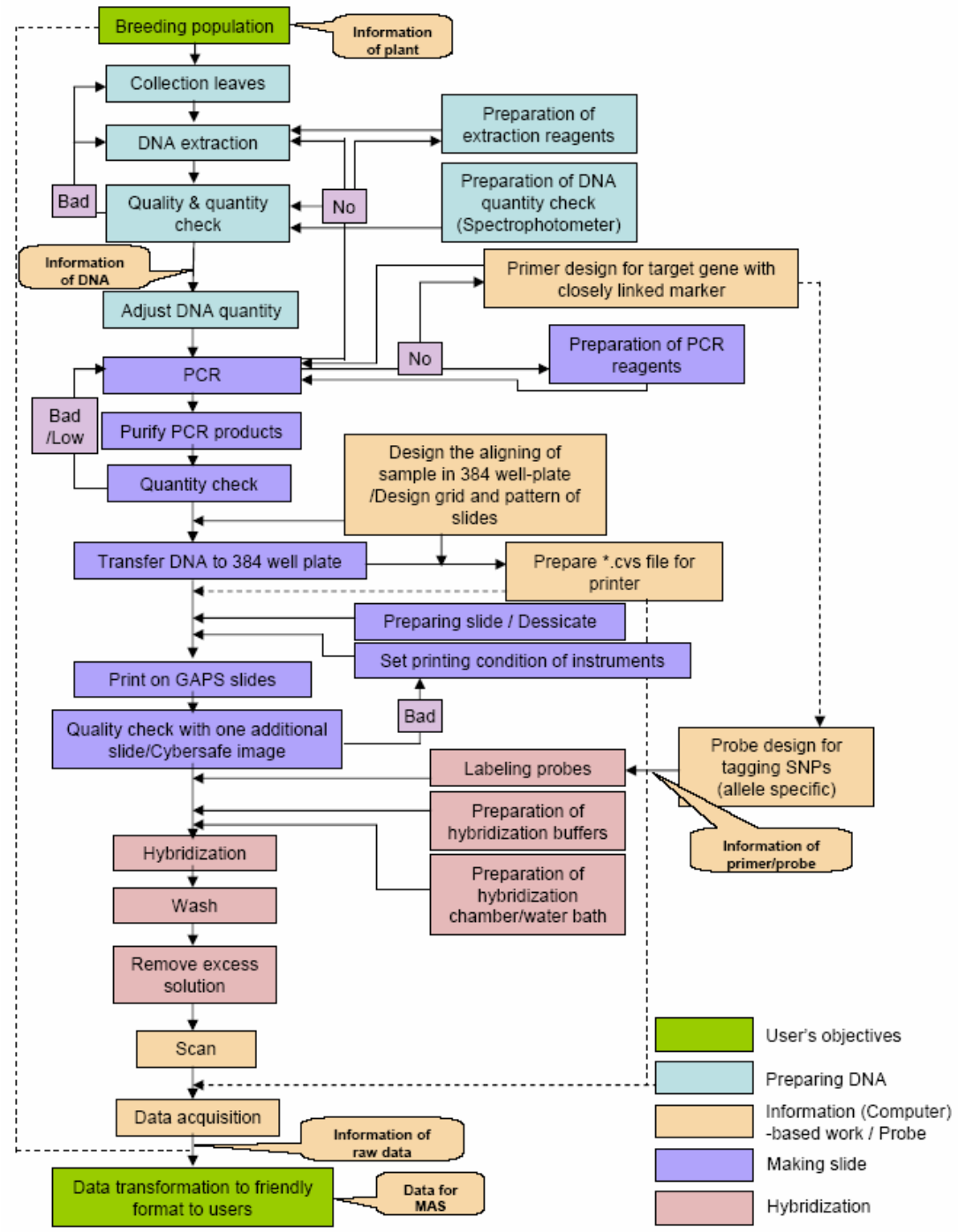


Fig. 5. Flowchart of MBG indicating workflow of the technology and data generation

We also need to identify final results that are important for data storage, and provide information on whether these genotype data are to be combined with phenotype at certain stage of the breeding process. This would determine compatibility with existing in-house databases.

Tangible outputs delivered

- Laboratory manual containing detailed protocol for each technology platform developed and distributed to all participants of the workshop (a zip file is attached to this report)
- Compilations of all presentations at the workshop (pdf files copied in a CD, to follow via courier or via FTP facility at IRRI, which can be downloaded directly by GCP)

- Submitted three proposals for consideration by GCP-SP5 call for CPSG on 01 May 2007

Deviations from the workplan
No deviation from the workplan.

Data availability
Since we have been developing different assay systems for the different technologies, no data on germplasm or breeding lines are available. All lines used have known phenotypes and genotypes. In collaboration with NARES partners in the future, we would be able to provide MAS data of breeding populations using these technologies.

2005-19: Evaluation and deployment of transgenic drought tolerant varieties

Principal Investigator:
J. Bennett, IRRI

MID-YEAR REPORT NOT SUBMITTED

2005-20: Optimising marker-assisted breeding systems for drought tolerance in cereals through linkage of physiological and genetic models

Principal Investigator:
Scott Chapman, CSIRO
Jiankang Wang (Co-PI), CIMMYT-China
Collaborators:
Simulations: Mark Dieters and Graeme Hammer, The University of Queensland
Wheat: Richard Trethowan, CIMMYT; David Bonnett, and Greg J. Rebetzke - CSIRO
Maize: Francois Tardieu, Claude Welcker - INRA
Private sector: Mark Cooper (advisory role) – Pioneer

Mid-year report
The main activities in this period have related to continued software development and licensing; two training courses; publication of two papers from the project; and a visit to Mexico and GCP. Details of these are given in the next section on outputs, and are summarised below.

1. A prototype of the file editing programme (QUGMAN) has been completed. This file allows genetic maps and QTL information (location and effect size) to be imported into formats that are suitable to run the QU-GENE simulation software. This software would benefit from further development with SP4 projects so that it can be integrated into more routine studies of introgression or selection of QTL.
2. The leaf elongation rate (LER) plug-in module has been tested with QuLine and breeding strategy simulations are being developed to apply these in the coming two months.
3. Development of a prototype hybrid breeding module (QuHybrid) has begun at CAAS with the assistance of a computer programmer funded by the project. This module will allow for simulations of testcross and recurrent reciprocal selection breeding methods, with marker-assisted selection.
4. Agreement with The University of Queensland has been reached on the new licencing arrangements for QU-GENE. From the end of June 2007, the executable software, including the major breeding module (QuLine) will be made freely available for download from the UQ website.

5. Two papers directly related to outputs of the project are in press: Wang et al 2007 Crop Science (published); Chapman 2007 Euphytica.
6. In March 2007, Dr. Wang presented seminars on QTL analysis, simulation and MAS at two training courses run at IRRI and at CAAS.
7. In May 2007, Drs. Chapman and Wang visited CIMMYT to work on the next coleoptile paper and to meet with GCP programme staff – J-M. Ribaut, P. Monneveux, J-C. Glaszmann. Also present were Jonathan Crouch (CIMMYT) and Graham McLaren (IRRI). The discussions included potential future needs of the GCP for capability in simulating MAS for drought-related traits, and a summary document was prepared. Dr. Monneveux emailed information about existing and new GCP projects that could benefit from interaction with the GCP-MAS project or any future initiative. A commitment was made by the parties to have a follow-up discussion in July/August 2007 about possible future directions.

Tangible outputs delivered

Many of the outputs for this project have been completed. However, there are still two papers in preparation that will be submitted before the end of the project. In May/June 2007, Dr. Chapman will travel to France to work on the paper that integrates the leaf elongation rate (LER) module with a QTL study.

These are listed below, together with an indication of their related outputs in the proposal and are followed by a list compared to the proposal outputs. (▼Fully supported by GCP SP3-20; ▽Partially supported by GCP SP3-20)

Case studies

1. Completion of Case Study I: Optimising the crossing and selection strategies to pyramid major genes in wheat when using marker assisted selection▼
(Outputs 1, 2). This preliminary study used existing CSIRO data to combine parental genes into the same genetic background, using markers based on known functional genes. A general paper with this example has been published (Chapman 2007 Euphytica).
2. Ongoing Case Study II: Building a genetic model for coleoptile length in wheat based on QTL studies and investigating MAS using both major genes and QTL▼
(Outputs 1, 2). This study extends the functional gene research to examine how to efficiently select for QTL for a drought related trait.
3. Ongoing Case Study III: Accounting for variability in the detection and use of markers for simple and complex traits by using the linkage of QU-GENE to a physiological model of LER (leaf elongation rate)▼
(Output 4). This study is being used to compare the genetic gain for different traits, in different environments, or in using QTL to select for improved leaf growth under drought.
4. Review of existing studies of simulated breeding of sorghum and identification of opportunities for utilisation of such data for the improvement of statistical methodologies. ▼
(Output 2). Compared with stochastic simulation, these physiological simulations have the advantage of representing the 'physiological' correlations among traits, and therefore provide a better test of potential statistical and QTL mapping software.
5. Completion of a design breeding approach in rice quality improvement▽
(Output 3). This research was largely completed prior to the start of the GCP project. However, the breeding strategies developed in this activity are being used elsewhere in this project.

Software tools

1. QuLine module▼
Maintenance and update of the breeding simulation tool (QuLine) (version 1.4 released in July 2006; a new version will be released in late 2007)

(Outputs 1, 2, 4). This tool can simulate all aspects of conventional and marker-assisted selection for inbreeding crops.

2. QugMan editor▼

Completion of a prototype editor (QugMan) for the QU-GENE input files

(Outputs 1, 2, 4). This tool allows marker and QTL maps to be used as direct inputs to the simulation platform. For example, consensus maps can be used for selection studies.

3. LER module▼

Development of a QU-Gene plug-in module (LER) for a physiological model of leaf elongation rate in response to drought

(Output 4). This plug-in enables direct input of gene information to a physiological model to allow the simulation of complex phenotypes in response to drought.

4. QTL Mapping software▽

Using the QU-GENE software library, developed a modified algorithm for improving the traditional composite interval mapping and associated software, i.e. ICIM (http://www.isbreeding.net)

(Output 2). This tool simulates populations of genotypes for testing of QTL mapping tools and will allow integration of QTL mapping as an important step in recurrent breeding strategies.

5. QuHybrid module▼

Preliminary design and development of the hybrid breeding application module for application in hybrid crops (i.e. QuHybrid)

(Outputs 1, 2). While QU-Line can be used to model the selfing components of a breeding programme this new module will allow simulation of breeding systems of hybrid crops including topcross testing and reciprocal selection schemes.

6. Availability of QU-Gene software▼

Through substantial negotiation with The University of Queensland, the QU-GENE/QuLine tools will be available under open-licence by June 2007

(Output 6). All restrictions to use of the software executables have been removed.

Training and exchanges

1. Workshop on training course on Quantitative Genetics and Statistical Methodology in Support of Germplasm Conservation and Crop Improvement at IRRI, The Philippines (March 14-16, 2007; 25 participants)▽(Output 3). J. Wang presented courses in breeding simulation and QTL mapping. From July 2007, IRRI intend to appoint a post-doc to work applications in breeding simulation, in consultation with J. Wang.

2. Training course on GE Interaction and Breeding Simulation in CAAS, Beijing, China (March 19-22, 2007; 48 participants)▽

(Outputs 3, 5). J. Wang and S. Chapman presented courses in breeding simulation and QTL mapping. The attendees included breeders and students from a range of crops including rice, wheat, maize, soybean and vegetables. Interactions with the breeders generated useful ideas for designing alternative breeding strategies in a range of crop species.

3. Project Travel: J. Wang visited Australia in 2005 for 3 months, in 2006 for 2 months; S. Chapman visited China in 2007 for 3 weeks and Mexico in May 2007 for 1 week.▼

Publications (published/accepted/in preparation) fully or partially sponsored by GCP SP3-20 (2006 to now)

Key:

▼ = Fully supported by GCP SP3-20

▽ = Partially supported by GCP SP3-20

* = Correspondence author)

▿Wang J, Wan X, Crossa J, Crouch J, Weng J, Zhai H, *Wan J (2006) QTL mapping of grain length in rice (*Oryza sativa* L.) using chromosome segment substitution lines, Genetical Research 88: 93-104.

▼Wang J, *Chapman S.C, Bonnett D.B, Rebetzke G. J, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection, Crop Science 47: 580-588.

▿Li H, Ye G, *Wang J. (2007) A modified algorithm for the improvement of composite interval mapping, Genetics 175: 361-374.

▿*Wang J, Pfeiffer W. H. (2007) Simulation approach and its applications in plant breeding, Scientia Agricultura Sinica 40: 1-12.

▼*Chapman S.C, Wang J, Rebetzke G.J, Bonnett D.G (2007) Accounting for variability in the detection and use of markers for Simple and complex traits. Proceedings of the Gene-Plant-Crop Conference, Wageningen April 23-26 2006. Kluwer Academic Publishers, The Netherlands

▿Wang J, Wan X, Li H, Pfeiffer W, Crouch J, *Wan J (2007) Application of identified QTL-marker associations in rice quality improvement through a design breeding approach. Theor. Appl. Genet. (in press).

▿Li H, Ye G, Li Z, *Wang J (2007) Inclusive composite interval mapping for digenic epistasis of quantitative traits. Genetics (resubmitted).

▼Wang J, *Chapman S. C, Bonnett D.B, Rebetzke G.J (2007) Integration of marker and phenotypic selection for coleoptile length in wheat. Crop Science (in preparation).

▼* Chapman S.C. (2007) Analyses of simulated plant breeding programmes for adaptation to drought environments. Euphytica (in press)

* Chapman S. C et al, Using physiological model of leaf elongation rate of maize in genetic simulation. J. Exp Bot. (in preparation).

▿* Jamieson P.D, Asseng S, Chapman S.C, Dreccer M.F, White J.W, McMaster G.S, Porter J.R, Semenov M.A. (2008) Modelling wheat production. In M. van Ginkel, World Wheat Book (in press).

Description against outputs of original proposal

1. Case studies for wheat and sorghum on explanatory power of physiological models. These will be undertaken to demonstrate the conditional effects of the interaction of genes and traits in conventional and marker-assisted selection for improved performance under different types of drought for drought-related traits.
   Software tools have been developed and case studies have been undertaken and published for wheat (Wang et al 2007), and have been revised for sorghum (Chapman 2007).
2. Breeding strategies for integration of drought traits with other traits. Using the case studies and for different levels of 'genetic understanding' alternative marker-assisted breeding strategies will be evaluated
   For wheat, a case study publishes the options to combine drought related genes and/or QTL with other function genes in scenarios for gene pyramiding (parental development) and for recurrent trait selection (population improvement). Development of a QTL-mapping activity in a related project (see papers) has resulted in improved tools for simulation of QTL-mapping that can be used in this project.
3. For rice, establish training relationships between IRRI and other partners with UQ. This will be built around a case study of multi-population data on QTL to report on design of breeding strategies to capture 'large QTL' for studies in submergence and salinity tolerance.
   In conjunction with other research, training courses in simulation and mapping were conducted at IRRI and CAAS (Beijing), and scientific support was provided for the establishment of an IRRI post-doc in breeding simulation. The work on submergence and salinity tolerance was put on hold due to the departure of key staff from IRRI.

4. For maize, demonstrate APSIM/QU-GENE link for existing QTL work on leaf growth. In conjunction with commissioned SP1 project (CIRAD/Luquet) a short report will be written on genetic analysis of the leaf elongation model.
   The LER module has been developed and is currently being used in simulations to compare the effects of trait, environment and QTL in selection methods. Publication underway.
5. Consultation with GCP consortium breeders (including legume and clonal crop breeders). This will be combined with data-gathering parts of 1 and 2 and to enable preparation of plans to apply the methodologies in 2 in other crops.
   Consultation to establish novel requirements of other breeders has occurred so far only at annual meetings and training courses. Direct engagement of breeders is considered to be necessary in any future work.
6. Develop process and template to make licence agreements between software owners (UQ/APSRU) and GCP participants (see below) to utilise existing and new tools.
   A completely free and open licence of executable code of QU-GENE has been established. APSIM is available for an annual maintenance fee from www.apsim.info
7. Using GCP web tools, establish a forum for iterative interaction between disciplines (molecular biologists, physiologists and breeders etc), crops, public and private sectors, and regions.
   A WIKI interaction page is being developed for the breeding simulation project.

Deviations from the workplan
As reported in Oct 2006, the project has been awarded a no-cost extension until 31 Dec 2007. Dr. Wang will continue to work on the project for this period, with salary paid by CIMMYT.

Data availability
Datasets are not a major output of this project. A GCP-Wiki page has been established for communication of results from the project, and will be developed further in the next 6 months. Published papers and relevant case studies will be published on that site.

2005-21: Planning for effective product development, delivery and use

Principal Investigator:
Victoria Henson Apollonio, CAS-IP, hosted by Bioversity International
Collaborators:
Karine Malgrand, Consultant
June Blalock, USDA
Zeze Sampaio, EMBRAPA
Andy Hall, UNU-MERIT
Michael Blakeney, Queen Mary Intellectual Property Institute
Helen Cordell, IPR Global
Elsie Quaite-Randall, McMaster University

Mid-year report (amendment to 15 October 2006 report)
The attached table details the proposed products of this project as included in the project proposal of November 2004, progress so far in producing these products and the reasons for deviation from the schedule of product delivery.

In this project, the investigators proposed to use data from four (originally three) case studies as examples for mapping and product development and delivery plans. Substantial progress has now been made in data collection and collation for the four case studies. This information is currently in the form of spreadsheets containing data identifying major products, product development stages, inputs (including third party inputs), input providers, initial (proximal) users of specific developmental stages and suggested end users of final products. However, the lack of a formal requirement for Principal Investigators (PIs) to provide

the necessary input into these case studies has been a major constraint to progress. This has delayed the investigators in delivering several of the proposed outputs.

With this information collected, it will be possible to quickly generate the product development and delivery maps that are essential prerequisites for generating many of the pending outputs; this task can easily be achieved during the proposed no-cost extension period (see accompanying request for no-cost extension). There have been several changes in the composition of the GCP Management Team since this proposal was finalized, which have led to many refinements of the GCP vision and strategy. However, based on the interest of several PIs at the 2006 Annual Research Meeting (such as R. Nelson and N. Ellis), the mapping techniques and the community of IP practitioners and technology transfer managers associated with this project continue to provide the GCP with an approach that is consistent with standards of technology transfer and intellectual asset management. In discussions with PIs of case-study projects, it was confirmed that this approach engages researchers in the challenge of getting up-stream research results into the hands of users as final products. At ARM06, scientists who had previously been reluctant to think about product delivery reported that this was an exciting approach.

It is notable that when this commissioned proposal was formulated with the previous SP3 leader during 2004 and 2005, it was envisaged that the project outputs would help define the GCP competitive call for grant proposals. In the mean time, SP5 has adopted many of the responsibilities in this area and has moved ahead in a somewhat different way than originally conceived in SP3. Nevertheless, it is hoped that the outputs from the project, as stated in the proposal, will still significantly contribute to "facilitate effective GCP planned product development and deployment with end-users through explicit holistic planning". The investigators remain open to suggestions from GCP management regarding how the outputs of this project can best serve their needs.

Proposed project product*	**Product status**	**Reason for deviation and plan to complete during no-cost extension**
1) Product development and delivery maps for three projects: genomics, transgenics, informatics (case studies): SP3-18, "Development of Low Cost Gene Based Trait Assay Technologies in Cereals" SP-19, "Evaluation and Deployment of Transgenic Drought Tolerant Varieties" SP3-20, "Optimising Marker-Assisted Breeding Systems for Drought Tolerance in Cereals Through Linkage of Physiological and SP3-Genetic Models"	Initial map for one project (SP3-18, case study); three maps under development (additional case was added at the request of the GCP: ("Unlocking the Genetic Diversity in Peanut's Wild Relatives With Genomic and Genetic Tools") This map was included in Poster and SP3/Plenary Session presentations at AGRM06	Delays in data collection** have made it difficult to complete the maps and move to more advanced steps of project. However, with most of the ground work for this now complete, we envisage it will be possible to quickly complete the development of these maps during the early part of the requested no-cost extension.
2) Product development and distribution plans for three projects that include:	Spreadsheet template was developed to collect project data for immediate construction of a project database with product development, constraints, delivery and user information.***	No deviation
a. identification of project products (both explicit research outputs and less obvious products	Products identified for each of the four case studies, both explicit and non-obvious. These lists of	No deviation

Proposed project product*	**Product status**	**Reason for deviation and plan to complete during no-cost extension**
such as networking techniques, etc.)	products are not exhaustive but intended to be illustrative.	
b. identification of potential users of at least one major product from each project	Potential user identified for at least one major product from each case study.	No deviation
c. development of strategies to involve end-users and stakeholders in the product development planning, testing and refinement process	Suggested strategies are being developed for the case study where product development and delivery maps are already available.	Delayed for some case studies pending completion of product development and delivery maps. These will be generated during the no-cost extension.
d. identification of possible product distribution routes for a least one major product from each project (transgenics, bioinformatics, genomics product types)	Preliminary work is ongoing but can not be finalized until product development and delivery maps are completed.	Delayed pending completion of product development and delivery maps. These will be generated during the no-cost extension.
e. evaluation of the plans	Not yet completed for any of the case studies.	Maps need to be completed (during the no-cost extension) before this can be accomplished in a meaningful way.
3) Definition and description of regulatory and legal constraints to product development and deployment, including intellectual property management, contractual/agreement provisions, biosafety and liability concerns	Information regarding IP and contractual issues affecting some products for each of the four cases has been collected. For the transgenic case study, issues regarding biosafety and liability may now be a moot point in view of the GCP's recent decisions regarding withdrawing from involvement in end-product development and delivery.	No deviation
4) Mitigation strategies to deal with legal and regulatory issues associated with product development and deployment		Mitigation strategies will be suggested for the genomics and bioinformatics case studies as soon as the product development and delivery maps and plans are drafted. This will be achieved during the no-cost extension.
5) A list of complementary skills, knowledge, access, and/or other assets requirement for efficient product development and for product deployment, including a list of potential partners to provide these assets.	Initial list completed for one case study.	Lists will be completed once product development and delivery maps are available. This will be achieved during the no-cost extension.
6) A training module that	This activity has been adopted on	Activity adopted by SP5. The

Proposed project product*	Product status	Reason for deviation and plan to complete during no-cost extension
will assist scientists applying for support from the GCP in the future in drafting product development and delivery plans. This module will be based on lessons learned in this project and on the experience and professional skills of the collaboration teams.	a broader scale by SP5. This project has been providing input to SP5 scientists, who have been developing a template to collect data from new grantees during proposal preparation. The collection of this data is aimed at providing grantees with a means to conceive of products and the delivery and use of these products in the development-to-delivery spectrum of up-stream research to farmers' fields'. In preparation for the proposed SP3 workshops, a consultant was contracted to assist in data gathering who also has the skills to assist the PI in preparing training materials and delivering training workshops.	project report here will continue to contribute to the SP5 process throughout the no-cost extension. Including, participation as a resource person in the proposed SP5 workshop on this topic.

G4007.03: "Community of Practices" concept applied to rice production in the Mekong region: Quick conversion of popular rice varieties with emphasis on drought, salinity and grain quality improvement

Principal Investigator:
Theerayut Toojinda, BIOTEC
Collaborators:
Jonaliza L. Siangliw, Rice Gene Discovery Unit (RDGU)
Sureeporn Kate Ngam, Ubon Ratchatani University (UBU)
Monthathip Chang, National Agricultural and Forestry Research Institute (NAFRI)
Men Sarom, Cambodian Agricultural Research and Development Institute (CARDI)
Toe Aung, Department of Agricultural Research (DAR)

Mid-year report
Collaborating institutes were informed that the proposal submitted to GCP was accepted and will start on 1 January 2007. Letters of support were received from four main collaborating institutes such as National Agricultural and Forestry Institute (NAFRI), Cambodian Agricultural Research and Development Institute (CARDI), Department of Agricultural Research (DAR) and Ubon Ratchatani University (UBN) committing their institutes to participate in the programmes outlined in the project proposal such as trainings to be conducted inside and outside of their country. This serves as sign of continues collaboration among countries in the Mekong Region.

Planning for the 2 year training programme has been initiated. The trainings to be held at Rice Gene Discovery Unit (RGDU), Thailand were scheduled on May 2007, October 2007, April 2008 and September 2008. For this year, the first training in May will begin on the 21st of May to the 30th (Table 1). The second training was set in October so that the participants will be able to attend BioAsia and the 2nd International Conference for Rice for the Future.

Regarding the workshop to be held this May, each institute had nominated two participants. Dr. Khun Leang Hak and Mr. Chou Vichet are joining from CARDI. NAFRI nominated Mr. Souvanh Thadavong and Mr. Khemkham Hongphakdy and for DAR, Daw Tin Tin Myint and Daw Thi Dar will join the training. All nominated participants are rice breeders. Dr. Sureeporn Katengam from UBN will be in charged with the IR57514 population which is also one of the populations developed during the first marker-assisted selection workshop sponsored by Rockefeller Foundation.

In the last concluded workshop on "Long-term practical training on Molecular Breeding on Rice for Mekong Region (LTMBR for MKR)" sponsored by Rockefeller Foundation and BIOTEC, the group of Cambodia identified drought as the main problem in rainfed areas. The Cambodian participants aim to improve cooking quality in drought tolerant variety named CAR3. They used rice variety 'Phkar Rum Duol' as donor. Phkar Rum Duol has good eating and cooking qualities but its yield is low under drought condition. Thus, cooking quality traits were transferred to CAR3 by MAS. At the end of the training, The Cambodian group identified 22 BC2F1 containing the cooking quality traits of Phkar Rum Duol based on marker profiles. After the end of training programme, the materials were allowed to self-pollinated and phenotypic selection was applied. Dr. Khun Leang Hak and Mr. Chou Vichet would like to continue two more round of MAS and backcrossing. They will bring the leaf tissue of 200 selected BC_2F_4 plants with CAR 3 background (which is resistant to drought) introgressed with cooking quality traits from Phkar Rum Duol as starting materials for this MAS training.

The Myanmar participants conducted two MAS activities. The first activity aims to improve salinity tolerance in new released variety 'IR53936-60-3-2-31' using rice variety 'Pokkali' as donor. Since yield in 4% of their planting areas in Myanmar is greatly affected by soil salinity and farmers living in this area are very poor. Therefore improving salinity tolerance is a need. At the end of the LTMBR for MKR programme they identified 22 BC_3F_1 carried QTL for salt tolerance on chromosome 1 from Pokkali. These materials were allowed to self-pollinated and phenotypic selection was applied in BC_3F_2 and BC_3F_3. Daw Tin Tin Myint and Daw Thi Dar will bring 200 BC_3F_4 lines for MAS selection and identify lines with salinity tolerance to be crossed to IR53936 to generate BC_4F_1. Then BC_4F_1 will be selected by MAS and the selected BC_4F_1 lines will be self-pollinated and phenotypic selection and MAS will be applied in BC_4F_2. The second activity aims to improve grain quality in high yielding variety 'Manawthukha' using Basmati as the donor. Manawthukha is popularly grown in most of rice growing areas in Myanmar, because it is well adapted, high yielding and with high milling recovery. However, its grain quality is inferior compared to Basmati, an aromatic fine grain rice. At the end of the LTMBR for MKR programme they identified 4 BC_3F_1. Myint Yi who got the MS scholarship from BIOTEC will continue this activity to be part of her thesis.

The Laos participants had chosen to improve the grain quality of the glutinous rice Thadokkham 1 or TDK1 by transferring genes from their local variety Homnangnouane, also a sticky rice. In the last programme, NAFRI ended up with 14 BC_2F_2 lines homozygous in 2 grain quality traits. Two hundred BC_2F_3 lines will be used as working materials by Mr. Souvanh Thadavong and Mr. Khemkham Hongphakdy to conduct MAS. The selected BC_2F_3 containing good cooking quality will be crossed to TDK1 to generate BC_3F_1. One more round of MAS backcrossing will be followed.

Another population to be continued in this training was developed by transferring cooking quality traits from a KDML105 introgression lines with good eating and seed quality traits, submergence tolerance and bacterial leaf blight resistance to the widely adapted rice variety IR57514. This population was handled by Thai participants from Ubon Ratchatani University and Rice Research Center at Ubon. At the end of the LTMBR for MKR programme, more than 500 BC3F1 seeds were generated and this will be used as the starting materials for MAS in this project.

Workshops to be held outside RGDU were scheduled in June 19-20, 2007 at CARDI, July 3-4, 2007 at NAFRI and August 20-21, 2007 at DAR. The workshop will emphasize on the application of

biotechnology in agriculture, plant genome and gene structure, DNA marker technology, genetic maps, QTL mapping and the use of marker-assisted selection in rice breeding. The video on plant molecular breeding and tangible example of MAS will be made also as part of the workshop in CARDI, NAFRI and DAR. RGDU would like to set the same workshop in all three institutes but in the case of DAR, RGDU may not be able to perform wet lab to demonstrate the use of markers in breeding programmes, thus video on how to use MAS will be shown. RGDU is now working with CARDI in the preparation of the workshop specifically on invitation letters from CARDI and equipment needed in the workshop.

Table 1. Schedule of the first training to be held at RGDU on May 21-30, 2007.

Outline for 1st MAS workshop: May 21-30, 2007

DATE	Day	Subject	Lecturer
21-May-07	Monday	Arrival/Check in	
		Welcoming participants to the training program	Dr. Apichart Vanavichit
		Planning for MAS training program	Dr. Theerayut Toojinda
		Welcome Party - Dinner	
22-May-07	Tuesday	Review basic molecular techniques:	
		Genomic DNA isolation and Agarose Gel Electrophoresis	Myint; Chanakarn
		DNA extraction (Lab)	RGDU staffs
23-May-07	Wednesday	DNA extraction (Lab)	RGDU staffs
24-May-07	Thursday	Review basic molecular techniques:	RGDU staffs
		Marker: Southern based	Meechai
		Marker: PCR based	Jutarat, Siriporn
		DNA quantification (Lab)	RGDU staffs
		Review basic molecular techniques: PCR	Vaiphot
		PCR (Lab)	RGDU staffs
25-May-07	Friday	Review basic molecular techniques:	Jirapong
		Polyacrylamide Gel Electrophoresis	
		PAGE-SSR (Lab)	RGDU staffs
		Band reading	Muny
		Band reading (Lab)	RGDU staffs
26-May-07	Saturday	PCR (Lab)	RGDU staffs
		PAGE-SSR (Lab)	RGDU staffs
		Band reading (Lab)	RGDU staffs
27-May-07	Sunday		
28-May-07	Monday	PCR (Lab)	RGDU staffs
		PAGE-SSR (Lab)	RGDU staffs
		Band reading (Lab)	RGDU staffs
		Data analysis (band reading)	Muny, Tanee
29-May-07	Tuesday	Data preparation	Participants
		Preparation of presentation	
30-May-07	Wednesday	Summarize Lab work	Dr.Theerayut Toojinda
		Group presentation	Dr. Apichart Vanavichit
		Discussion	
		Future planning	Dr.Theerayut Toojinda
		Departure	

Tangible outputs delivered

1. Support from collaborating institutes stating their interest and participation in the project either inside or outside of their country.
2. Plan for the 2 year training to be held at RGDU.
3. Planning and scheduling of the first training in May 2007.
4. Names of nominated participants from CARDI, NAFRI and DAR. Identifying the starting materials for this training from the results obtained in the last MAS training.
5. Plan for on site workshop to be held at CARDI, NAFRI and DAR.

Deviations from the workplan
The schedules of trainings at RGDU and workshops at CARDI, NAFRI and DAR were changed due to late processing of the money allocated for the project.

The participants were selected by their own institute as opposed to the plan that participants will be selected after the workshop has been done in each institute. The plan was changed because the workshop on site has not been done yet.

Data Availability
The letters of cooperation, names of nominated participants and plans for trainings in RGDU and outside RGDU are all kept in the RGDU server. The website for this MAS training will be posted as link to http://dna.kps.ku.ac.th by the end of September.

G4007.05: Bridging genomics, genetic resources and breeding to improve wheat and barley production in Morocco

Principal Investigator:
Abbad Andaloussi Fouad, INRA
Collaborators:
Nsarellah Nasserlehaq, INRA
Jlibene Mohammed, INRA
Lhaloui Saadia, INRA
Labhilili Mustapha, INRA
Saidi Seddik, INRA
Sripada M. Udupa, ICARDA
Roberto Tuberosa,\ University of Bologna
Mark E. Sorrells, Cornell University
Manilal William, CIMMYT
J. Perry Gustafson, University of Missouri

Mid-year report
Overall Note
The overall objective of this project is to bridge conventional wheat and barley improvement activities and biotechnology. The project starts with phenotyping of a set of local and GCP and CG shared germplasm. The molecular work will be done in the following seasons to relate to the on going field work. The work in this first season (2006-07) is hence mostly devoted to field experimentation and sample collection. Field and post harvest evaluation will be continue for most target traits (yield component and grain quality attributes).
The seeds received from the partners arrived very late in the season and special arrangements permitted to secure harvest for some. The season was particularly dry in most regions and sites. The early rains were unusually low and drought continued throughout the season. Planting in more than one station and planting in irrigated stations saved the outcome of the season. The remaining crop is in its late stages and is still under evaluation. In general, a good evaluation of drought and associated stresses is expected from this season data.

Bread wheat
Objectives: 1) Increase seed for further multidisciplinary evaluation , (2) evaluation of reaction to rusts, septoria and Hessian fly resistance, grain quality and drought tolerance (3) validation of molecular markers linked to septoria and rust resistance (4) new crosses to combine resistance to diseases and tolerance to drought (5) developing new populations for molecular marker studies.

Materials from the breeding programme as well as materials introduced from International centres, and the GCP INRA-France and a special CIMMYT collection of materials adapted to several mega-

environments were planted in different sites. These nurseries are meant for seed increase and field evaluation. Seed will be distributed for further precise field and lab evaluation in the future. Most of the sites were drought stricken. However, the some materials planted in sites (Tassaout, Sidi Allal Tazi, Haj Kaddour and Annoceur) where supplemental irrigation is possible, we are expecting seed increase and this seed can be used for conducting planned experiment during coming season. The irrigated or supplemented sites

Evaluations will be made for various agronomic traits including resistance to rusts, Septoria, and Hessian fly. Evaluation for grain quality traits will be done after harvest.
Two Recombinant inbred lines population are under increase and evaluation (phenotyping) and will be used next season for evaluation. The BERKUT/KRICHAUFF population meant for drought is under evaluation in Tassasout station. New crosses were made and the successful crosses will be harvested and selectively advanced. The new crosses were made between drought tolerant material and disease resistant materials (rusts, septoria, Hessian fly).

Durum wheat
The objective of the durum wheat work is to increase seed and phenotype several materials for disease resistance, insect resistance and drought tolerance a well as for grain quality. Another set of objectives is to develop mapping populations.
Materials from the three regional breeding programmes, materials introduced from International centres, the Root rot nursery, the GCP_Durum collection, and the GCP reference set received from INRA-France, were planted in different sites. Two sites are irrigated. These nurseries are meant for seed increase and field evaluation. Seed will be distributed for further precise field and lab evaluation in the future.
Evaluation of leaf stem and yellow rusts are on going. Evaluation of root rot and Hessian fly will be done in June. Evaluation for drought tolerance attribute is on going.
Laboratory evaluation for yield components and quality and drought related traits will be pursued during the summer. The Drought trial and grain quality trial consisting of 100 durum wheat varieties differing in drought tolerance and grain quality. They were planted in 4 sites: Two sites were drought stricken. The agromorphological studies, radiation studies and grain quality studies will be done according to plans.
'Stojocri' variety based Recombinant inbred lines populations are under development. These populations (stjcri / Telset2, Stjcri/ Cham1, Stjcri/ Karim, Stjcri/ Isly) are still in F4 generation and will be advanced in the greenhouse followed by the field in the coming season. They are meant for drought and heat tolerance marking. New crosses were made this season in view of developing new population with different parents than the one currently available. The successful crosses will be harvested in June.

Barley
The objective of the barley work was to characterise and identify germplasm tolerant to biotic and abiotic stresses. The reference set collection of barley was not received and hence locally developed material was chosen for this season.
A collection of 5400 lines selected in the dryland areas of Morocco was evaluated in small plots for adaptation to biotic and abiotic stresses. A second group of 110 lines were evaluated in larger 10 square meters plots. These lines are six rowed type barley, short to medium height and early to mid early in flowering and maturity. Protein content ranged from 11 to 16 %. These lines were planted and evaluated in Marchouch, Annoceur and Jemaa Shaim experiment stations. As in the other crops drought was predominant and caused some failures except for the irrigated stations. On going evaluation is concerned with agromorphology, phenology, and yield and its components. Quality attribute such as protein, β glucane, final yield and ash content will be evaluated.

Tangible outputs delivered
1. Reference sets of wheat was received and planted for seed increase
2. Local germplasm of wheat including improved materials were planted for seed increase
3. Local germplasm of barley including improved materials were planted for seed increase

4. Mapping population of bread wheat (BERKUT/KRICHAUFF) was received and planted for seed increase.

Deviations from the workplan
Due to unexpected drought during cropping season (2006-07) in Morocco and late arrival of seeds from the partners some activities may not able to perform or some observations could not be recorded. Such activities will be repeated during coming cropping seasons. Since, the crop is in the field and process of evaluation is in progress, exact details about such activities will be available at the end of the season (after 2-3 months). Therefore, details of activities to be repeated will be available in the next report.

In case of barley, reference set of the germplasm was not available during current cropping season. Once, the reference set received from ICARDA, the related activities will be planned.

Data availability
The project just started…

Subprogramme 4: Bioinformatics and crop information systems

2005-22: Development of GenerationCP domain models and ontology

Principal Investigator:
Richard Bruskiewich, IRRI
Collaborators:
Guy Davenport, CIMMYT-CRIL
Jonathan Crouch, CIMMYT-CRIL
Tom Hazekamp, Bioversity International
Elizabeth Arnaud, Bioversity International
Graham McLaren, IRRI
Thomas Metz, IRRI
Ken McNally, IRRI
Ruaraidh Sackville-Hamilton, IRRI
Martin Senger, EBI

Mid-year report
Project objectives

- *Domain Model Maintenance:* Fix and further refine the GCP domain models based on practical validation in GCP data curation and platform development**.**
- *Ontology Management embedded in the GCP Platform:* Collaborate with "Data Templates Development" and "Platform" tasks to incorporate easy end-user access to ontology terms from a GCP ontology dictionary, for data annotation.
- *GCP Ontology Curation:* Elaborate GCP domain models ontology through expert community scientific input from the GCP and external scientific communities.

Project outputs

- Production release of GCP domain models (with possible minor upgrade releases)
- Ontology management tools for ontology curator and end-user access embedded in GCP platform.
- Production release of a refined GCP domain model linked with priority sets of expert-reviewed dictionaries of GCP domain model ontology for 2007 release domain models.

Summary status report on project progress (see also tables below)

- GCP project meeting with some members of the team held at CIMMYT in January, took some design decisions with domain model development implications.
- Updated production release 1.1.* of the GCP domain model published in April 2007 with updated UML diagrams and supporting narratives, at http://pantheon.generationcp.org/demeter.
- GCP Java platform implementation of the Eclipse Modeling Framework (EMF); specification of the GCP domain model updated to latest GCP platform design decisions, such that current GCP platform middleware (Ceres) now uses the latest production model.
- GCP-funded ontology curator (Mr. Jeffrey Detras) hired late last year to replace previous staff member. The new staff member has been working steadily on converting/compiling priority ontology using the OBO Edit ontology curation standard. Protégé reviewed once again but put aside as an ontology tool (once again) as being more complex than the project needs at present.
- Ontology inventory on GCP ontology web site (http://ontology.generationcp.org) updated.
- GCP-funded ontology software engineer (Mr. Kevin Manansala) hired in March 2007 to assist: 1) with integration of ontology databases into GCP platform; 2) in data conversion requirements for importing external ontology/descriptors into the project; and 3) in the development of GCP platform interfaces (GCP DataConsumers) and web services using GCP ontology.

- Tom Hazekamp visited IRRI for one week in mid-March to discuss the development of specific priority crop ontology dictionaries for germplasm, passport and characterisation data. A first release of this "GCP Germplasm Ontology" made for further community in April.
- Job description for CIMMYT-hosted ontology postdoctoral fellow prepared and sent to CIMMYT for clearance. Formal approval by CIMMYT is still pending and recruitment has not started yet.
- The Plant Ontology (PO) Consortium and the Gramene database team (contact: Dr. Pankaj Jaiswal of Cornell University, a GCP core partner) have volunteered to commit NSF project money to hosting a small (~15 delegates) plant and trait ontology workshop at Cold Spring Harbor Laboratory (New York, USA) focused on GCP (CGIAR) mandate crops on June 10-11, 2007. Core members of task 22 project team will participate, in addition to other CGIAR/GCP representatives and some pertinent US collaborators.

Further information given in appendices

Key products developed by the project:

1. GCP domain models published at http://pantheon.generationcp.org/demeter as major release 1.0 and applied to GCP platform and network implementations to manage GCP data.
2. Data template and GCP platform tools cross-linked to ontology browser/selector tool connected to GCP ontology database.
3. GCP domain model major release ontology database published online with full documentation (at http://ontology.generationcp.org) and populated with priority sets of target ontology, for application to GCP data annotation.
4. Community strategy and tutorial materials about ontology usage, delivered to end-users in GCP training workshop at 2007 ARM.

Quantifiable outputs:

1. GCP domain models published at http://pantheon.generationcp.org/demeter as major release 1.0 and being applied to GCP platform and network implementations to manage GCP data. **Published.**
2. GCP domain models published at http://pantheon.generationcp.org/demeter as minor releases 1.#, where # is the minor release number. Minor releases will generally incrementally fix problem areas in the models or add methods to interfaces but not disrupt the overall topology of the models (hence, change impact will be localized) **In progress**.
3. GCP ontology database published online with full documentation (at http://ontology.generationcp.org). **In progress**.
4. GCP data template tool with embedded support for GCP ontology selection and community-curation. **Not yet started (task 25 programmer not yet recruited)**
5. GCP search engine and related applications containing embedded support for ontology-driven queries
6. Project PDF recruited; start-up project coordination meeting(s) convened for ontology tool tutorial training of task collaborators. **Posting to be cleared by CIMMYT for recruitment (as of April 2007)**
7. Assigned target ontology developed and loaded into ontology database for production use.**In progress**.
8. Strategy and policy document for long term community-curation of GCP ontology prepared for review at 2007 GCP annual review meeting, then posted as the GCP bioinformatics portal site, with linkages to relevant ontology resources (database, tools)
9. Ontology workshop training materials developed (in GCPWiki) then published online (in GCP Bioinformatics Portal).

10. Tutorial training to instruct GCP scientists about the community-curated development, application to data annotation, and data mining with, GCP ontology, delivered to interested GCP scientists in satellite workshop at 2007 GCP annual review meeting.
11. Scientific publication documenting the status of the GCP domain model and ontology project prepared and submitted.

Data production and availability section format

1**. The nature of the data to be produced:** GCP domain models, with associated ontology dictionaries.

2. **The format of the data:**

a. Domain models published as Unified Modeling Language (UML) diagrams with Java language, BioMOBY and XML schema implementations. **Published.**
b. Online (master copy) GCP ontology web site with:

i. Web browser interface- **in progress**
ii. Web services interface – **not yet provided**
iii. Downloadable text files in standard community formats (OWL, OBO)- **not yet provided**Possibly, downloadable (SQL?) copy of the database (for local mirroring by partners) **– not yet provided**

iv. 3. **Where and when the data will be posted:**

a. Domain models already available for download as documented UML diagrams at http://pantheon.generationcp.org with computable versions in CropForge (http://cropforge.org). **Published.**
b. GCP domain model ontology database and files accessible from http://ontology.generationcp.org. **In progress**

2005-23: Implementation of web service technology in the GCP Consortium

Principal Investigator:

Milko A. Škofič, Bioversity International/CGIAR System-wide Genetic Resources Programme (SGRP)

Collaborators:

Samy Gaiji, Bioversity/SGRP
Rajesh Sood, Bioversity/SGRP
Tom Hazekamp, Bioversity/SGRP
Mathieu Rouard, Bioversity
Martin Senger, IRRI
Markus Döring, Botanic Garden and Botanical Museum Berlin-Dahlem [BGBM]
Javier de la Torre, University of Madrid

Mid-year report

The latter part of 2006 was devoted to the development of TAPIR. This protocol and toolset was selected as the main web service deployment tool for databases. Particular emphasis has been placed on its ability to serve BioMOBY services natively. The software was released in November 2006 and the two following months were devoted to testing and fixing any remaining problems.

The beginning of 2007 has been devoted to testing the package and defining the strategy to bridge the schema-oriented approach of TAPIR with the object-oriented approach of BioMOBY. Consultation with the modeling group was undertaken to define a strategy and standards to map the TAPIR passport schema to the BioMOBY data types. Further consultation with Mathieu Rouard and Fernando Rojas was undertaken on how to handle EST databases.

A series of Concept Name Servers have been set up – one at the GCP and the other at the EURISCO site (http://eurisco.ecpgr.org). These Concept Name Servers provide schemas, alias files, TAPIR/BioMOBY configuration files and mapping configurations to all TAPIR users.

A portion of the GCP Central Registry (http://gcpcr.grinfo.net/include/webservices/index.php) has been set up as a web services resource centre, structured to include:

- ***A Technology section****: describing the three main web service technologies used in the GCP: BioCASE, BioMOBY and TAPIR – the section provides a general description of the protocols and provides links to further documentation;*
- ***An Installation section****: covering the installation procedures for BioCASE, TAPIR and BioMOBY – installation instructions are described, including links to further documentation and resource centres;*
- ***A Schemas section****: describing the XML schemas used by BioCASE and TAPIR, including links to download the schemas, CMF files needed by BioCASE and schema documentation; and*
- ***A Data sources section****: to monitor all registered data sources hourly; the system can currently track BioCASE and TAPIR servers. In the future, it may be necessary to add support to track BioMOBY services via the MOBY registry at IRRI. The data provider's connection status and speed are displayed in a weekly matrix. There is a plan to provide a daily view, including the metadata and capabilities of the individual providers.*

In February 2007, the Principal Investigator participated in a TAPIR development workshop at GBIF headquarters in Copenhagen; new developments were demonstrated and others were planned. Support for RDF and LSID resolution services will be of particular interest to the GCP.

In early April, a meeting with Mathieu Rouard, Principal Investigator of the GCP project 'Application and development of Web Services technology', took place at Bioversity headquarters in Rome. The purpose of this meeting was to analyse the relationship between the GCP domain model, the current BioMOBY data types and the XML schemas being developed in order to define standards, procedures and a strategy to map BioMOBY objects in TAPIR.

In the interim, a TAPIR server was installed at Bioversity. It will provide passport data from the EURISCO and SINGER catalogues, and serve test data sources. Another two TAPIR installations – one at Bioversity's office in Montpellier and one at CIAT – are now online; CIP and CGN will follow shortly.

Tangible outputs delivered

TAPIR has been finalised. This includes:

- Setting up a Concept Name Server;
- Creating a web services documentation and monitoring site;
- Installing TAPIR at Bioversity headquarters, Bioversity Montpellier and CIAT; and
- Participating in the TAPIR development workshop at GBIF headquarters.

Deviations from the workplan

Work has proceeded according to the workplan; no deviations are foreseen.

Data availability

This project acts as a service to other GCP projects in order to assist them in producing data and making them available.

2005-24: Application and development of web services technology

Principal Investigator:
M. Rouard, Bioversity International
Collaborators:
M. Conte, CIRAD
R. Bruskiewich, IRRI
M. Ruiz, CIRAD
G. Droc, CIRAD
M. Senger, EBI
M. Skofic, Bioversity
B. Lord Hendrix, IRRI

Mid-year report
A meeting held at CIMMYT in January 2007 provided an opportunity to refine the project workplan. This project is composed of the three main objectives listed below:

1) Improvement of the Moby core technology
A GCP web service provider package was developed and documented on the GCP wiki page:
http://cropwiki.irri.org/gcp/index.php/GCP_Web_Service_Provider_Package

This MOBY service provider package has been initiated with the goal of easing deployment of MOBY Web Services in International Rice Functional Genomics Consortium (IRFGC) member sites. With this package, web service developers do not have to start from scratch, but can utilise a pre-packed web application to deploy web services in their own sites/servers.

The project team is also reviewing the GCP Moby data types in order to fit with current needs in genetics and genomics.

2) Support SP1 use case
Implementation of passport data Moby services compatible with TAPIR data sources has been initiated. A meeting was held at Bioversity headquarters in Rome with Milko Skofic, Principal Investigator of the GCP project 'Implementation of web services technology in the GCP Consortium', in order to define and implement services to allow queries on passport data information. This work should be completed in the coming weeks.

A GCP data source for genotyping templates has also been developed. This data source allows access to data from a local Excel file. The project team will explore its potential application for remotely accessing templates stored in the GCP Central Registry.

3) Orthologs prediction tool
The pipeline for integrating a new sequence into an existing phylogeny has been developed and is currently in the testing phase; initial tests show consistent biological results. The next step is to develop the web services themselves, and then to integrate them into the GCP platform via KOIOS.

The Greenphyl database is located at the National Computer Center for Higher Education (CINES), which is a high-computing centre. There have been some problems with installation of the libraries onto that server. The project team must either wait for the new server to be set up in few weeks or migrate Greenphyl onto a CIRAD server, which could be tedious. This issue should be resolved soon but is slowing down the process of web services development.

Tangible outputs delivered (see above)
- GCP web service provider package
- Orthologous prediction pipeline

Deviations from the workplan
To date, the project is proceeding according to the workplan. Some problems have been experienced recently with installation of the Greenphyl database onto the server at CINES, which has slowed down the process of web services development. However, these problems should be resolved soon and are not likely to result in any significant deviation from the workplan.

Data availability
This project is not supposed to produce data. It provides access to datasets produced by some others GCP projects and also an also to relevant tools to analyse them.

2005-25: Creation and maintenance of templates for GenerationCP data storage in repositories

Principal Investigator:
Guy Davenport, CIMMYT
Collaborators:
Richard Bruskiewich, IRRI
Tom Hazekamp, Bioversity International
Jane Morris, ACGT (University of Pretoria)

Mid-year report
Work on the new versions of the GCP templates has started using input from the GCP Genotyping Data Quality Workshop held at IRRI, 19-23 February (http://cropwiki.irri.org/gcp/index.php/Genotyping_Data_Quality_Workshop) and from the data already available in the Central Registry.

In collaboration with the GCP Central Registry we did a review of the data in the Central Registry to see if they conformed to the GCP templates. The results of this review are reported in the Central Registry report. The providers of data not conforming to the GCP templates will be approached and given help to reformat or complete their datasets.

As part of a no-cost extension to the 2006 project the web interface for GCP is expect to be finished by August 2007, with a prototype available for testing at the end of June 2007.

Due to staff recruitment problems at IRRI we have not started the envisioned improvements to the template editor and validator

Tangible outputs delivered
No so far, first deliverable is planned for August 2006

Deviations from the workplan
None

Data availability
Template definitions and documentation are available from http://www.generationcp/templates/definitions and the templates themselves are available under a creative commons license from CropForge (http://cropforge.org/frs/?group_id=18)

2005-26: Management of the Generation Challenge Programme Central Registry

Principal Investigator:
Tom Hazekamp, Bioversity International/CGIAR System-wide Genetic Resources Programme [SGRP]
Collaborators:
Guy Davenport, CIMMYT
Samy Gaiji, Bioversity/SGRP
Milko Skofic, Bioversity/SGRP
Rajesh Sood, Bioversity/SGRP

Mid-year report

Technical Management of Central Registry

During the reporting period, day-to-day maintenance of the GCP Central Registry was undertaken and some changes were made to the Central Registry web pages. More detailed statistics on the type of resources held by the Central Registry are now shown. In addition, some core functions such as registering and searching for datasets have been positioned further toward the front end to increase user-friendliness.

Strengthening the Central Registry's resource collection

In an effort to strengthen the Central Registry's resource collection, several Principal Investigators of GCP projects were approached to register and or upload their datasets. These included the PIs of projects that were initiated in 2006 and those who reported that their project had produced data sets in an online user survey (see below).

The list of projects for which resources can be registered and uploaded at the Central Registry now includes SP5 projects; until recently, it contained only SP1, SP2 and SP3 projects. The first SP5 contribution, a reference microsatellite kit for barley, has been registered and uploaded.
In April 2007, the GCP Project Officer provided access to 2006 GCP project reports. These reports are being used to check for availability of project datasets and guide further data-collection efforts.

In November 2006 and April 2007, Central Registry status reports were generated. These reports were used to inform the leaders of SPs1, 2 and 3 of the submissions made by PIs working in their subprogrammes and to solicit feedback in order to increase the amount of resources held by the Central Registry. During the reporting period, the number of registered datasets increased from 78 to 101 (29%). The number of uploaded files increased from 21 to 33 (57%).

Development of the Central Registry

Colleagues from CIP used some of the metadata held by the Central Registry, along with Mondrian software, to determine how (graphical) visualisation tools could be used to provide an insight into the resources held by the Registry. A basic prototype was produced, but has not yet been incorporated into the Central Registry site.

As an added-value activity, the Central Registry project created a dedicated section on the GCP wiki site for assessment of analysis tools (http://cropwiki.irri.org/gcp/index.php/Assessments_of_Analysis_Tools_for_SP1,_SP2_and_SP3). Thirty reviews of various analysis tools provided by Brigitte Courtois of AGROPOLIS were re-formatted and uploaded into this section.

Content Management

New or updated entries into the Registry have been checked and obvious mistakes corrected. PIs have been informed if the corrections to their entries are more significant than simply re-formatting.

The GCP project 'Creation and maintenance of templates for Generation CP data storage in repositories', led by Guy Davenport of CIMMYT, has commenced the validation of uploaded data files against GCP data templates. The first validation run revealed that the majority of uploaded data files require some form of modification to pass the validation test. CIMMYT will work with data-file providers to fix formatting and validation problems. A stand-alone version of the templates validation tool created by CIMMYT is available and a web-based version is expected to be released by the end of August. These tools will enable data providers to validate their own data files before or during data-file uploads.

Help desk
Help has mainly been provided to users by email. A Frequently Asked Questions (FAQ) section on the Central Registry's website is kept up to date with information on some of the most common issues encountered by users. References to the Help Desk were made more visible on the Central Registry 'Welcome' page (http://data.generationcp.org).

User survey
An online user survey regarding the GCP Central Registry and data templates was developed in collaboration with the GCP project 'Creation and maintenance of templates for Generation CP data storage in repositories'; 72 PIs were invited to participate in the survey and 40 (55%) responded. The survey showed a high awareness of the Central Registry by the respondents (96%) but a moderate use (40%) to date.

A small follow-up survey of 16 respondents was also undertaken; the majority of these respondents had used the Central Registry. Respondents were asked to score the functionality and ease of using various components of the Registry, with most components receiving a 'good' to 'excellent' score. The ease of using the query system, the FAQ and Help Desk functions received scores of 'good' to 'average'. Some of the core functions such as 'register', 'search' and 'download', and linkages to 'help' functions have already been made more prominent to increase user-friendliness. Each request for help is now monitored separately.

Tangible outputs delivered

- **Central Registry application**
  A fully functional Central Registry application (http://data.generationcp.org) has been developed to allow registration of datasets, uploading of data files and registration of web service entry points. It features flexible query facilities and allows users to download centrally stored data files. As of 1 May 2007, 101 datasets have been registered at the Central Registry and 33 files related to these datasets are available for direct download.

- **Section on GCP wiki site on assessments of analysis tools for SP1, SP2 and SP3**
  Thirty assessments of analysis tools, mainly provided by AGROPOLIS, were re- formatted and uploaded to a dedicated section on the GCP wiki site (http://cropwiki.irri.org/gcp/index.php/Assessments_of_Analysis_Tools_for_SP1%2C_SP2_and_SP3 – requires login).

Deviations from the workplan
The project has proceeded according to the workplan; no deviations are foreseen.

Data availability
Metadata on datasets produced by GCP funded research are available online at http://data.generationcp.org.

2005-27: Integration of the High Performance Computing (HPC)-facilities in the GenerationCP toolbox

Principal Investigator:
Anthony Collins, CIP
Collaborators:
Reinhard Simon, CIP
Roland Schlafeitner, CIP
Subhash Chandra, ICRISAT
Jayashree Balaji, ICRISAT
David Hoisington, ICRISAT
Rajeev Varshney, ICRISAT
Richard Bruskiewich, IRRI
Thomas Metz, IRRI
Manuel Ruiz, CIRAD
Guy Davenport, CIMMYT
Marcos Costa, EMBRAPA
Jorge Franco, CIMMYT
Marilyn Warburton, CIMMYT
Jiankang Wang, CIMMYT
Etienne de Villiers, ILRI

Mid-year report

Introduction

Due to the no-cost extension of the 2006 project through May 2007, this update report presents some global highlights as a hybrid of results overlapping the 2006 and 2007 project plans by lead centres CIP, ICRISAT and IRRI. Subsequent reporting for the 2007 ARM in September will bring into clear focus the Activities, Quantifiable Outputs, and Key Products proposed as 2007 goals.

CIP:

Data analysis support to gene expression analysis. Key staff left and staff with appropriate skills needed to be provided with time. The task has now partially been re-assigned and should be delivered by end of June 2007 (see deviations). However, the overall CIP HPC workload has steadily increased, including by providing supplementary processing for IRRI's heavy HPC workload.

CIP also continues to coordinate and support the installation of common HPC programmes including the Rmpi statistics package, as used by CIMMYT, CIP, ILRI, and IRRI, thus nurturing a growing community of practice. An Eclipse based user interface to R and HPC has also been developed (prototype – Ref. Slide 2)

ICRISAT:

During the year 2006, the collection of comparative genomics and population genetics software tools have been expanded with the implementation of several more freely available tools either parallelized or optimised to work on the Paracel HPC at ICRISAT. Tools include parallelized "structure" software with web UI, visualisation tool for Structure output, ClustalW-MPI and Parallel Tree Puzzle, format conversion tool for popular population genetics software including TFPGA, and improvements in the SNP detection pipeline implemented during 2005-2006.

The web interfaces to the parallel structure programme allow for job submission, output access and retrieval (http://hpc.icrisat.cgiar.org/webstructure/login.php). Visualisation of structure output for even large datasets has been made feasible through the implementation of a visualisation tool, which we have called "VisualStruct". This tool is an improvement over the freely available "distruct" programme (Rosenberg, 2004). Besides software for format conversion to programmes like TFPGA are also available

through the user interfaces. The software ClustalW-MPI (***http://web.bii.a-star.edu.sg/~kuobin/clustalg/***) for multiple sequence alignment and parallel TreePuzzle (***http://www.tree-puzzle.de/***) are also available through the web pages on the HPC.

Use of the tools available on the HPC: The comparative genomics software is being used by students and staff from the genomics laboratory at ICRISAT. The population genetics software are also proving to be useful for national partners NRCS (National Research Centre for Sorghum), NBPGR (National Bureau of Plant Genetic Resources) and ANGRAU agricultural University. ICRISAT is also attempting to publicize the availability of software tools on the HPC at workshops held in ICRISAT where participants come from national and international institutes in this region.

Reference:

Rosenberg NA, (2004) Distruct: a programme for the graphical display of population structure. Mol.Ecol.Notes. 4, 137-138.

IRRI:

The IRRI HPC has strongly consolidated usage for the year 2006-2007, with an increasing variety of user applications being used for bioinformatics computation both within and beyond the GCP community:

- R Statistical software using both Rmpi and Rpvm for parallel/cluster execution by RMauloen, EDeomano, VBartolome, CJunxing
- Paracel Blast, NCBI Blast, BLAT by VUlat
- CIS-Element Searching Tool by Koji Doi, a Perl application.

As confirmed by the Ganglia Cluster Toolkit, cluster's main CPU has been extensively used by many user and/or applications: refer slide 1 attached.

Tangible outputs delivered – highlights:

ICRISAT:

The software "Structure" with web UI
Visualisation tool for population structure
ClustalW-MPI implemented on the HPC
Parallel TreePuzzle implemented on the HPC
Format conversion possible through the web interface; converts to formats required for popular population genetics analysis software.
Further improvements to the SNP detection pipeline implemented on the HPC.

Deviations from the workplan:

CIP:

The activity is delayed due to staff leaving the project. Presently, part of this activity (transfer of Westfall & Young MinP and MaxP Resampling-based multiple testing algorithms) is being re-assigned.

The installation of Expresso is being reconsidered due to lack of adequate human resources. Simon proposes to rather consider installation of tools more fitting available skills in R, PHP and Java. One such tool appropriate for analysis of microarray experiments on the HPC would be CARMA (http://www.u.arizona.edu/~jhoying/carma.html, http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16542461). CARMA would be a kind of replacement of Expresso in the sense of a complete analysis pipeline tool. It misses some of Expressos novel features like generation of first order logical statements. However, it would provide an easy-to-use solution through a web interface. Other tools considered for installation on the HPC include now the AROMA package (http://www.maths.lth.se/help/R/aroma/) based on R.

ICRISAT:
The parallelisation of MUSCLE was found unfeasible after a detailed study of the algorithm and its components.
Reference:
Robert C Edgar. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics** 2004, 5:113.

Data availability:
NYA

2005-31: Development of ortholog-function display tools

Principal Investigator:
Richard Bruskiewich, IRRI
Collaborators:
Kimmen Sjölander, University of California at Berkeley
Brigitte Courtois, CIRAD
Manuel Ruiz, CIRAD
Christophe Perin, CIRAD
Mathieu Conte, CIRAD
Mathieu Rouard, Bioversity International
Marcos Costa, EMBRAPA
Georgios Pappas, EMBRAPA
Natalia Martins, EMBRAPA

Mid-year report

- Dr. Samart Wanchana, GCP-funded postdoctoral fellow hired on this project, continues curation of the comparative stress gene catalog. Dr. Wanchana sent data sets to Dr. Kimmen Sjölander for bulk analysis and continues using local high performance computing resources for local analysis.
- Dr. Wanchana is also collaborating closely with Dr. Ramil Mauleon, GCP-funded postdoctoral fellow on the 2006-08 SP4 analysis task for SP2 data. In particular, significantly expressed probe sets from stress treatment microarray experiments from Dr. Mauleon are being aligned to gene families by Dr. Wanchana.
- Complementary research on stress responsive genes in the rice genome undertaken by Dr. Wanchana as a part of the Rice Annotation Project (http://rapdb.lab.nig.ac.jp/).
- Further development of the GCP platform interfaces to the GMOD Chado-based project database continued during the reporting period albeit at a slower pace than before due to deficiencies in the software engineering team on the project (partly resolved as of March).
- EMBRAPA team recommitted during the January GCP platform meeting at CIMMYT to work on Apollo and Genoma linkages to GMOD Chado, which could help data curation in this project (if delivered in a timely fashion).
- Progress on CIRAD-hosted "GreenPhyl" comparative gene resource reviewed at rice functional genomics meeting last fall. Mathieu Rouard of Bioversity, Task 24 team leader in 2007, has committed to wrap GreenPhyl with web services, for query from Dayhoff web site (see below).

Tangible outputs delivered

- http://dayhoff.generationcp.org web site represents visible product of the project. The site and its data access modalities are to be elaborated further by the next GCP ARM.

Deviations from the workplan

- Due to the delay in the 2 -year post-doc recruitment in 2005, gene catalog curation was delayed by one year. Project completion is now projected to December 31, 2007.

2005-32: Development of crop gene expression database and data mining tools

Principal Investigator:
Shoshi Kikuchi, NIAS
Collaborators:
Richard Bruskiewich, IRRI
Hei Leung, IRRI
Masaru Takeya, NIAS

Final project report
Executive summary:
In this project, we have established a database storing the data from several rice microarray systems, and a pipeline system for the cis-element search in the promoter region of the genes which turned out to be the same cluster after the microarray experiments. A project database has been basically specified with the customisation of MAXD system. MAXD is a data warehouse and visualisation environment for genomic expression data. It is being developed in the University of Manchester by the Microarray Bioinformatics Group. We have used a second version of Maxdload2 accurately. Currently about 300 gene expression data from the cDNA-based microarray system originally stored in RED (Rice Expression Database) and about 200 gene expression data from the 22K oligomicroarray system (The Agilent Technologies G4138A) are in storage.

To speculate the common transcription factor(s) that may control the expression of the genes which show the similar expression profile, one of the ways is to search the common sequence motifs (cis-elements) locating the promoter region of the genes. It is very annoying work for the researchers to find the common cis-elements in the promoter region of the clustered gene set after microarray analysis. For such needs, we have established a WEB-based pipeline system to solve the problem. 5' upstream sequences of the genes based on the mapping data of the full-length cDNA sequences to rice genome sequence, from those 60mer-oligo were synthesized and printed on the 22K oligoarray system, and the MEME programme http://meme.sdsc.edu/meme/intro.html for the key engine for the motif sequence search in the promoter region are the components of this pipeline system. The pipeline system has been announced to the public through several mailing lists on March 14, 2007. And the address is http://hpc.irri.cgiar.org:80/tool/nias/ces .

2005-34: GenerationCP software engineering and collaboration platform

Principal Investigator:
Thomas Metz, IRRI
Collaborators:
None

Mid-year report

- Both systems, CropForge (http://cropforge.org/) and GCPWiki (http://cropwiki.irri.org/gcp/), have been serviced and maintained during the reporting period, with small software updates applied at the application and operating system levels.
- The CropForge system now hosts 70 projects, 22 of which are related to the GenerationCP, and has 157 registered users. It is in continuous use by GCP software developers.
- The CropWiki system has now 188 registered users and has about 458 content pages and 403 uploaded files. There have been a total of 50,702 page views, and 8,682 page edits since the wiki was setup. A major use in the recent period was for workshop documentation and reports (e.g. http://cropwiki.irri.org/gcp/index.php/Genotyping_Data_Quality_Workshop) and for the development of a SP5 bioinformatics course. (see: http://cropwiki.irri.org/gcp/index.php/SP5/Bioinformatics101/index.html).

- A system for monitoring the external connectivity to the CropForge and CropWiki servers has been installed in November 2006. The system produces connectivity charts for the last 3 hours, 30 hours, 10 days, and 400 days for both servers.

  See: CropForge: http://smokeping.cheng-metz.de/index.cgi?target=Network.IRRI.CropForge

  See: CropWiki: http://smokeping.cheng-metz.de/index.cgi?target=Network.IRRI.CropWiki

  The systems also monitor connectivity to other servers at IRRI and CIMMYT and has been used as a diagnostic tool in case of networking problems.
- A new and openly accessible wiki site for the GCP has been established at: http://cropwiki.irri.org/generationcp. For anonymous users, this site looks like a normal website while for logged-in users, it appears like a wiki. This new wiki is meant for the collaborative dissemination and support of GCP products.
- The IRRI helpdesk (Jonathan Michal Mendoza) was able to resolve all problems that appeared during the reporting period, usually in less than two working days.

Tangible outputs delivered

- Both systems have now been running for more than 2 years and are in constant use. Regular software and platform upgrades have been applied.
- A helpdesk quickly resolved all user problems.
- A system for the continuous monitoring and charting of connectivity to both servers is in place and publicly accessible.
- A new generationcp wiki with public access and released under Creative Commons open access license has been established.

Deviations from the workplan

Project is on track with only minor deviation from the project time table.

Data availability

The Software Engineering and Collaboration Platform is designed to make all the uploaded information (source code, text, messages, images, documents) transparently accessible to the users. Technically, the GCPWiki system is access- restricted largely to the GCP participant, but practically all non-GCP scientists working in the field who have requested access have obtained it. The CropForge system allows open access to the entire content. A new generationcp wiki is globally accessible and the content is released under a Creative Commons open access license.

2006-08: Data analysis support for existing projects in SP2 with emphasis on integrating results from microarray and mapping experiments

Principal Investigator:

Guy Davenport, CIMMYT

Collaborators:

Richard Bruskiewich, IRRI
Hei Leung, IRRI
Shoshi Kikuchi, NIAS
K. Satoh, NIAS
Masaru Takeya, NIAS
Andreas Magusin, JIC
Jose Crossa, CIMMYT
Yunbi Xu , CIMMYT

Mid-year report

Two general activities were carried out in microarray data analysis at IRRI.

The first major activity was customising the microarray expression database (maxd-GCP) to load GCP-specific data. Data for two array platforms (Agilent 22k rice oligoarray platform – catalog no. G4138A, and University of Arizona maize oligoarray chip) were coded as XML scripts loadable into maxd-GCP. These are available at the **cropforge** site http://cropforge.org/projects/gcpmicroarray/ .

The second major activity was the development of data analysis software for microarray analysis. Two sub-activities were done: (1) Implementation of published analysis algorithms as R and perl scripts, for pipeline analysis, and (2) Reconfiguration of pre-existing open-source or freeware programmes for the rice array system (initially).

For (1), the following methods that can be incorporated into an analysis pipeline are:

a) Significance analysis – Microarray ANOVA (R/MAANOVA, Wu et al., 2002). This is used for statistical determination of differentially expressed genes with multiple factors such as treatments, genotypes, and their interactions, using a mixed linear model. The **R/MAANOVA** package is available at http://www.jax.org/staff/churchill/labsite/software/Rmaanova/index.html, while implementing R scripts and sample data are available at the **cropforge** site.

b) Gene pair correlation analysis –computes the correlation (selectable from Pearson, Spearman-Rank, etc) of all pairs of genes across several experiments. Implementing R script and sample data available at the **cropforge** site.

c) Genome region correlation (RCE) analysis - determines if contiguous blocks of genes show correlated expression across different but relevant microarray experiments. The method of Spellman and Rubin (2002) is implemented in R and perl and is available at the **cropforge** site.

d) Differentially expressed gene (DEG) aggregation analysis - determines if DEGs from relevant microarray experiments aggregate in specific regions of the genome (Choi, Satoh, et al, in prep. 2007). This analysis is implemented as a frequency statistics test, wherein the ratio of count of DEGs:nonDEGs in a block of the genome is tested for significant difference with the genome-wide DEG:nonDEG ratio (fixed ratio test). Implementing perl and R scripts and sample data available at the **cropforge** site.

For (2), open-source or freeware applications for interactive microarray analysis were adapted to enable their use in the rice microarray system. These are the following:

1. Gene expression profile clustering using TIGR Multiple Experiment Viewer (**TMEV**, Saeed et al., 2003, available at http://www.tm4.org/mev.html). Several gene expression profile clustering solutions are implemented in TMEV, and routinely explored in order to come up with the most biologically relevant expression clusters. Rice array-specific gene annotation files for use in TMEV is available at the **cropforge** site.

2. Enrichment / over-representation analysis using **EASE** - determines if a characteristic or attribute is enriched in a subset of genes of interest (for example, is a particular pathway, gene function, or QTL-bin over-represented in a list of genes?). This analysis is implemented as a frequency statistics test by the programme EASE (Enrichment Analysis Systematic Explorer, Hosack et al. 2004) which compares the subset gene with the genome wide observation of a particular characteristic. Standalone EASE tool available at http://david.abcc.ncifcrf.gov/content.jsp?file=/ease/ease1.htm&type=1. The mapping of the genes/features of each microarray platform for EASE is available at the **cropforge** site.

3. Pathway binning and visualisation using **MAPMAN** (http://gabi.rzpd.de/projects/MapMan/) – determine and visualize which metabolic pathway / reaction / process a subset gene group belongs to. The mapping of TIGR5 rice gene models to curated rice pathways (using the Affymetrix-rice curated pathways done by the ANU Research School of Biological Sciences,

http://bioinfoserver.rsbs.anu.edu.au/utils/GeneBins/download.php), for association) is available at the **cropforge** site.

For further development are the following:

1. Integration of the pipeline analysis systems in the GCP Bioinformatics portal, and use of CGIAR HPC resources as analysis engine.

2. Mapping of the features of diverse GCP microarray systems to enable cross-platform (cross-crop) integrated analysis and data mining.

3. Continuous software development for new/interesting analysis algorithms encountered.

4. Adaption/reconfiguration of open-source or freeware applications for all GCP microarray systems.

At CIMMYT we expect to appoint the post-doctoral fellow starting 1 August 2007.

At the Plant and Animal Genome in January 2007 we held a project management meeting. In February Andreas Magusin (JIC) and Dr K. Satoh (NIAS) visited IRRI to meet with staff there. Access grid meetings with staff at IRRI and NIAS were held daily during the meeting.

References:

Hosack D.A, Dennis G. JR., Sherman B.T., Lane H.C, and Lempicki R.A. (2003) Identifying biological themes within lists of genes with EASE. Genome Biol. 4: R70-R70.8.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabatin, Braisted J, Klapa M, Currier T, Thia Garanjan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Rylstov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34(2):374-8.

Spellman PT, Rubin GM. (2002) Evidence for large domains of similarly expressed genes in the Drosophila genome. J. Biol. 1:5, 5.1 – 5.8.

Wu H, Kerr MK, Cui X., Churchill G. (2002) R/Maanova: An extensive R environment for the analysis of spotted cDNA microarray experiments. From the website http://www.jax.org/staff/churchill/labsite/software/anova/

Tangible outputs delivered

- XML scripts to load data for two array platforms (Agilent 22k rice oligoarray platform – catalog no. G4138A, and University of Arizona maize oligoarray chip) into maxd-GCP.
- Perl and R scripts and sample data for genome region correlation (RCE), differentially expressed gene (DEG) aggregation analysis
- Rice array-specific gene annotation files for use in TMEV
- Mapping of the genes/features of each microarray platform for EASE
- Mapping of TIGR5 rice gene models to curated rice pathways

Deviations from the workplan

No deviations from the workplan. A no-cost extension until July 2007 was envisioned in the original workplan. Milestones and devliables are delayed due to the late appointment of CIMMYT post-doc.

Data availability
All outputs are available on CropForge (http://cropforge.org/frs/?group_id=73)

2006-16: Development of an integrated GCP information platform

Principal Investigator:
Graham McLaren, IRRI
Collaborators:
Sub-Task I – SP1 Use-Case (lead from CIRAD):
Manuel Ruiz, CIRAD
Brigitte Courtois, CIRAD
Sub-Task II – SP2 Use-Case (lead from NIAS):
Shoshi Kikuchi, NIAS
Maseru Takeya, NIAS
Richard Bruskiewich, IRRI
Ramil Mauleon, IRRI
Marcos Costa, EMBRAPA
Georgios Pappas, EMBRAPA
Sub-Task III SP3 Use-case (lead from CIMMYT):
Guy Davenport, CIMMYT
Kyle Braak, CIMMYT
Manilal William, CIMMYT
Key Collaborators from related projects:
Martin Senger, Consultant with EBI
Mathieu Rouard, Bioversity

Mid-year report
After Use-Case discussions with biologists at a meeting at PAG in January, a planning workshop was held at CIMMYT from 22-26 January 2007. Use-cases were defined around identified biological problems associated with each GCP Sub Programme. Workplans were made to develop platform components to solve these problems. The task was subdivided into three sub-tasks each supporting problems from one of the biological GCP SPs.
(See http://cropwiki.irri.org/gcp/index.php/Platform_Development_Activities_for_2007)

Sub-Task I will support querying of genotype data, combining genotyping datasets, converting genotyping data between different formats, providing a pipeline to different analysis applications, checking genotyping data quality and connecting to mapping data.
Progress from January to May 2007 is as follows:

- Connection to genotyping and passport data sources
    - Integration and update of GCP genotyping data into TropGene : sorghum, coconut, musa, rice (in progress)
    - Existing TropGene GCP datasources need still to be refined to be compliant with the current GCP model
    - Bioversity team is beginning to develop Tapir/BioMoby web services for passport databases
    - CIMMYT is integrating genotyping data into IMIS and IWIS (in progress)
- CIRAD is developing a prototypic Web application for the query and the combination of different genotyping studies. This prototype will be shown to targeted users at the end of may in order to collect feedbacks.
- Connection with genetic and physical maps (GBrowse, CMap, OrygenesDB) : the version 1 of the GenMapping Web Query is still functional but not yet connected to the GCP Web platform (delays in having student assigned to this work)

Sub-Task II proposes to make progress on the fundamental problem of Functional Genomics: Identification of the list of candidate genes whose expression, inferred directly by transcript expression (microarray) and/or position (QTL mapping), exhibit statistically significant association with target trait values (phenotypes), for a given germplasm, under specified conditions (treatments, environment).
Progress to date is as follows:

- A review of available Portal Technology to construct a container for GCP Web applications has failed to produce a clear solution so GCP Web applications will continue to be built independently for now. The Koios search engine prototype developed in 2006 is being re-factored to use the GCP DataConsumer API and a new prototype will be available for testing by end of June.
- GCP project microarray data has been deposited at NCBI-GEO and MaxDLoad software integrated with GCP data sources, but MAXD database not yet wrapped as a GCP datasource. 91 microarray data sets produced in Kikuchi's lab to have been deposited to NCBI-GEO (GSE7530,31,32). Sample information of these data was described according to the MIAMI-Plant standard. These data and information were also sent for customization of the MAXD-base DB to IRRI.
- NIAS cis-element data mining tool was opened to the public on14th March. The next iteration in tool functionality is being implemented – motif searching in downstream regions as well as upstream and more freedom in selecting reference gene lists. Motif searching for coding regions and downstream regions have been extended. Specific motifs in such regions will be useful for comprehensive understanding of gene regulatory mechanism. We prepared corresponding sequence data of those regions as well as upstream regions for each TU. These data can be obtained from KOME. Currently, little is known about the role and function of motifs in these regions. The tool tries to pick up 'specific' motifs for user-defined gene lists, by comparison with the large reference list including all transcription units derived from KOME and Pseudomolecule ver 4.0. When the specificity is evaluated for the cis-element candidates of the limited number of genes, any considerable candidates may seem 'highly specific', no matter how the threshold is set or what statistic test is used to filter out false-positive candidates. This is not desirable to ensure sufficient sensitivity to detect really meaningful cis-elements. To solve this issue, we planed to prepare multiple reference data sets of more limited size, which are defined by some appropriate factors, such as 'drought stress responsive'. Users will be able to assign a data set freely along with their interest.
- Martin Senger has completed the coding for Soaplab version 2. A student is working on a prototype for a GCP DataConsumer using the Soaplab protocol to invoke specific software tools on the IRRI HPC.
- Basic MOBY Client DataSource design strategy and code framework established and the framework applied to some simple MOBY service use cases. A prototype GCP Web Service Provider implementing mutant phenotype use case has been developed into a package for deployment by collaborators (shared with a non-GCP mutant database team from Taiwan).

Sub-task III proposes to develop the GENOMEDIUM application interface to facilitate comparative mapping and QTL queries for breeders to select parents for crossing. Data will come from GCP and public data sources and analysis will be done using existing applications such as CMTV and the CIRAD GenMapping servlet.
Progress to date is as follows:

- After the January developer's meeting a number of changes were proposed in order to allow the use cases to be implemented (http://pantheon.generationcp.org/Releases.html). GenoMedium has been updated to use this new version of Pantheon.
- Support for changing the data source through the GUI has now been added.
- The genotyping use case has been updated to reflect these changes http://pantheon.generationcp.org/usecases/genotyping
- The tutorial on DataSources (http://pantheon.generationcp.org/TutorialDataSources.html) and DataConsumers (http://pantheon.generationcp.org/TutorialDataConsumers.html have been updated.
- Instructions on how to define DataSources and DataConsumers) in GenoMedium are also available (http://pantheon.generationcp.org/TutorialPantheonEclipse.html).

Tangible outputs delivered
None

Deviations from the workplan
Software development at IRRI is considerably slowed by the resignation of two GCP developers and the completion of studies by several students. Active recruiting is on- going well as recruitment of student interns from Canada for four-month summer internships.

2006-17: GenerationCP data quality improvement and assurance

Principal Investigator:
Thomas Metz, IRRI
Collaborators:
T.v. Hintum, WUR
B. Jayashree, ICRISAT
F. Atien, Bioversity International
D. Marshall, SCRI

Mid-year report
After an extensive review and discussion of the project during the ARM 2006, the project plan was changed for 2007. The emphasis of the project is now on funding 3 workshops in strategic areas of data quality. These workshops, to be conducted during 2007, are now in the planning phase as follows:

Workshop on passport data quality assessment and improvement:
This workshop is planned to be held at Bioversity HQ in Rome, either 18-22 June or 25-29 June. Participants will be from GCP data provider institutions. A final list of participant participants will be available shortly.

Workshop on ISO quality certification for genebanks and service labs:
In September 2007, a three to four day workshop will be jointly organised with GPG2 in Wageningen to discuss the quality management of genebanks and data production labs. The invitees will mainly consist of genebank managers (GPG2 budget plus self- funded participants) and production lab managers (GCP budget).

Workshop on ICRISAT LIMS adoption and collaborative development:
The first circular for the workshop was sent out late March wherein IRRI responded offering to send two of their bioinformatics staff to this workshop. CIMMYT, ICARDA and EMBRAPA also responded to the announcement and are interested in keeping the software going. Besides the GCP participants, there are participants from the private sector (a seed company- MAHYCO) who have installed the LIMS after agreeing to the GPL. They will participate in the workshop at their own expense. The team at ICRISAT has been working on the LIMS code, its structure and documentation in anticipation of the workshop. A collaborating open source software developer has been contacted who can serve as facilitator during the workshop.

Upon request, the project has been granted a no-cost extension for the 2006 workplan until 31 December 2007. In line with the 2006 workplan, the following workshop was conducted:

Genotyping data quality workshop:
This workshop took place at IRRI HQ, Los Banos, Philippines last 19-23 February 2007. The workshop had 9 external participants (CIP, CIMMYT, ICARDA, CIRAD, INRA, ICRISAT, SCRI, Diversity Arrays), 6 participants from IRRI, and 5 contributors via video/audio conference (CIP, CIMMYT, GCP).

The workshop is documented on the GCPWiki (http://cropwiki.irri.org/gcp/index.php/Genotyping_Data_Quality_Workshop).
As part of the 2006 workplan also, a LoA with CIMMYT on the documentation of the use of linked PDAs-GPS handheld devices for field data collection was finalised.

Tangible outputs delivered

- A LoA with CIMMYT on the documentation of the use of linked PDAs-GPS handheld devices for field data collection was signed.
- A workshop on Genotyping Data Quality was held at IRRI HQ. The workshop documentation is available on the GCPWiki.

Deviations from the workplan
The workplan for the 2007 data quality project is on track with minor adjustments. The 2006 workplan is still being executed under a no-cost extension to 31 December 2007. Several contributions from collaborators are still in the pipeline.

Data availability
All information outputs of this project are posted to the collaborative workspaces on CropForge and GCPWiki.

2006-18: Creation of institutional bioinformatics capacity (CIAT)

Principal Investigator:
Joe Tohme, CIAT
Collaborators:
Mathias Lorieux, CIAT
Alex Garcia, CIAT
Fernando Rojas, CIAT

Final project report
Executive summary:
We have designed and implemented software pipelines to support the analysis of EST and genomic sequences; these include EST assembly and SNP discovery and multiplexing. We have also developed a flexible Laboratory Information Management System that is currently being used not only as a central repository but also as the source of raw data for analysis pipelines; intermediate results are also being stored in this repository. Laboratory process workflows are being managed via our LIMS allowing in this way to accurately capture data. Ontologies, where possible, have been used to increase the accuracy of the descriptions and the replicability of the experiments.

Domain experts have been involved in the development of those tools supporting their research, not as informants but as active designers of their own technology. This dynamic relationship between engineers and domain experts has been possible in part because of our educational efforts; throughout 2006 four major training activities were carried out. These activities also allowed us to actively engage local universities in useful collaborations related to data integration in biological sciences.

2006-19: Creation of institutional bioinformatics capacity (CIMMYT)

Principal Investigator:
Guy Davenport, CIMMYT
Collaborators:
None

Final project report
Executive summary:

- The Research Informatics Laboratory is now established at CIMMYT
- Systems for the management of CIMMYT genomic have been established.
- An Access Grid node has been installed at CIMMYT
- One international and one national student have visited CIMMYT for training in bioinformatics

2006-20: Creation of institutional bioinformatics capacity (CIP)

Principal Investigator:
Reinhard Simon
Collaborators:
Merideth Bonierbale, CIP
Roland Schafleitner, CIP
(Ruth Grene, Virginia Tech - Cancelled)
(Lenwood Heath, Virginia Tech – Cancelled)
Luis Avila, CIP
(Enver Tarazona, CIP- Cancelled)

Mid-year report
Regarding Expresso wrapping this needs now to to be reconsidered. See below.
Regarding integration of systems-biology tools (COPASI, GEPASI) is being addressed.

Tangible outputs delivered
Contributions to micro-array related data processing and analysis work at CIP as part of the following publication:
Schafleitner R, Gutierrez Rosales RO, Gaudin A, Alvarado Aliaga CA, Nomberto Martinez G, Tincopa Marca RL, Avila Bolivar L, Mendiburu Delgado F, Simon R, Bonierbale M. "Capturing candidate drought tolerance traits in two native Andean potato clones by transcription profiling of field grown plants under water stress". Submitted to Plant Physiology and Biochemistry
http://research.cip.cgiar.org/confluence/display/FGenomics/Home

Deviations from the workplan
The task regarding the integration of EXPRESSO is still on hold and likely to be cancelled. Other main contributions include support of microarray data management on an ad-hoc basis (by L. Avila). One more staff (E. Tarazona) has also left the team for personal reasons and is now mainly back-stopping under a part-time consultancy contract regarding installation issues of R on the HPC (in collaboration with the HPC team at CIP). Thus, overall the main intention now is to replace the implementation of Expresso with a related one using R-based tools.

Data availability
http://research.cip.cgiar.org/confluence/display/FGenomics/Home

2006-21: Creation of institutional bioinformatics capacity (ICARDA)

Principal Investigator:
M. Singh, ICARDA
Collaborators:
M. Baum, ICARDA
K. Chabane, ICARDA
A. Akintunde, ICARDA
M. von Korff, ICARDA
K. El-Shamaa, ICARDA
H. Simo, ICARDA
H. Abed, ICARDA

Final project report
Executive summary:
A number of software on population structure, phylogenetic analysis, marker trait association and sequence assembly were learnt and support to the scientists and students working on SP1 – SP3 was provided. – STRUCTURE, Tassel, and DARwin – were examined for their functionalities, self training was carried out on these and support was provided to the scientists. Two staff members were trained in the areas of bioinformatics and statistical genomics at conferences and workshops during 2006. We need to continue upgrading the bioinformatics resources.

2006-22: Creation of institutional bioinformatics capacity (ICRISAT)

Principal Investigator:
Chandra S, ICRISAT
Collaborators:
Jayashree B, ICRISAT
Hoisington D, ICRISAT

Final project report
Executive summary:
Scientists in ICRISAT's Global Theme on Biotechnology employ a range of modern genomic technologies in their efforts to enhance the efficiency and effectiveness of crop improvement, including high-throughput genotyping. The increasing volumes resulting from the use of these technologies demand efficient information management platforms that allow the data to be explored, analysed and interpreted effectively and efficiently. The role of this project on the Creation of Institutional Bioinformatics Capacity at ICRISAT under the Generation Challenge Programme has been a starting point to the better management of biological information at the Institute. The projects has fuelled several bioinformatics activities; especially those on human capacity building, and updating and maintenance of existing infrastructure. Over the last three years, the Bioinformatics team have provided support to the user community at ICRISAT in terms of (a) Building awareness of new data sources and newer versions of commonly used bioinformatics tools, (b) Providing support with using available information management systems/tools and (c) Supporting existing databases published online, regularly updating and maintaining them with user inputs to increase their usefulness.

2006-23: Creation of institutional bioinformatics capacity (IITA)

Principal Investigator:
Dong-Jin Kim, IITA
Collaborators:
None

Mid-year report
The IITA/GCP Bioinformatics workshop was held at ILRI-Nairobi on 12th and 13th of October 2006. The goal of this workshop was to build research capacity in the area of genomics and bioinformatics in IITA-Nairobi as well as the NARS in this region. The following key issues were raised during the open discussion forum at the workshop; 1. Building a Bioinformatics capacity on the web: Since majority of the Scientists/ Researchers have access to the World Wide Web, it can be used as a medium to communicate research ideas and also to host research tools which can be accessible to all. 2. Integrating with the other research areas: Different scientists have different research interests but they may all share common computational tools to realise their objectives. There is need to offer different research groups information that is specific to their research interests. 3. Need for a more intensive Bioinformatics Workshop: Intensive training in Bioinformatics is needed so as to enable researchers get a thorough understanding of computational tools needed to generate and analyse biological data in a faster manner.

Therefore, we are working on developing a web interface that will provide a communication platform for participating scientists. A draft web structure have been developed, and to be developed further for a functional bioinformatics web site at the end of the project.

Tangible outputs delivered
Training of 30 participants in the bioinformatics area

Deviations from the workplan
None

2006-24: Creation of institutional bioinformatics capacity (Bioversity International)

Principal Investigator:
Samy Gaiji, Bioversity International/CGIAR System-wide,
Genetic Resources Programme (SGRP)
Collaborators:
None

Final project report
Executive summary:
This project was intended to further support the target activities involved in implementing the Subprogramme 4 workplan – including other GCP-commissioned projects being implemented at Bioversity International. The main activities included recruiting additional bioinformatics staff members at Bioversity and supporting existing bioinformatics experts. Bioversity proposed to allocate the project funds to recruitment a Bioinformatics Senior Scientist based at Bioversity headquarters in Rome, Italy. This person was to be dedicated to implementing GCP bioinformatics activities led by Bioversity as well as providing leadership and support of the overall GCP Subprogramme 4 platform and implementation design.

Expected outputs of this activity included management of GCP-commissioned projects, development and deployment of Web Services technology and the GCP Central Registry, content management, the

establishment of a help desk for GCP information activities and enhanced public awareness of Subprogramme 4 initiatives.

A Senior Bioinformatics Scientist was selected after a competitive recruitment process and a candidate was offered the post, however this candidate did not accept the position. As a contingency measure, arrangements were made to recruit consultants for specific bioinformatics tasks (including further refinement of the Central Registry, upgrade of Web Services technology and deployment of Web Services at specific sites). All these activities were carried out as planned and are continuing with additional GCP-commissioned activities in 2007.

2006-25: Creation of institutional bioinformatics capacity (IRRI)

Principal Investigator:
Graham McLaren, IRRI
Collaborators:
R. Bruskiewich, IRRI
T.Metz, IRRI

Final project report
Executive summary:
In 2005, Task 29 has supported the positions of two NRS bioinformatics staff as well as bioinformatics training for IRRI staff and resource material. Bioinformatics training focused on proactive acquisition of new skills in bioinformatics which will be required as we progress in molecular evaluation of germplasm with microarrays and GxE analysis of phenotypic evaluation.

- Thomas Metz attended a workshop on Practical DNA Microarray Analysis in Munich, Germany from May 9th to 12th , 2005.
- Arllet Portugal attended courses on Programming with C# and Programming the MS .NET framework using C# in Manila, Philippines from July 11th to 15th and 18th to 22nd , 2005.
- Graham McLaren attended the course on Computational and Statistical Aspects of Microarray Analysis in Bressanone, Italy from June 19th□24th , 2005.
- Emily Deomano attended the advanced course on the design and analysis of multi-environment trials: conventional and QTL based methods in Zaragoza, Spain from September 12 to 23rd, 2005.
- Subscriptions to three journals were also taken out under this project: 1) Journal of Bioinformatics; 2) Briefings in Bioinformatics; and 3) Applied Bioinformatics.
- In 2006, one NRS staff position was funded from this project to improve critical mass in research informatics at IRRI.
- Arllet Portugal attended a training course in Singapore on management of software development projects to better equip her to manage GCP software development teams.
- One NRS staff member travelled from IRRI to CIMMYT in May 2006 to attend the ICIS workshop and work with colleagues at CIMMYT on common solutions to the management of genotyping data in cereals.

2006-34: Installation and implementation of the ICRISAT LIMS at the Biosciences Eastern and Central Africa (BecA) facility and IITA-Ibadan

Principal Investigators:
Dave Hoisington, ICRISAT
Jayashree Balaji, ICRISAT
Collaborators:
Etienne deVilliers, ILRI-Nairobi

Morag Ferguson, IITA-Nairobi
Sarah Hearne, IITA-Nairobi
Santie deVilliers, ICRISAT-Nairobi
Rosemary Mutegi, ICRISAT-Nairobi

Final project report

Executive summary:

The LIMS developed by ICRISAT meets the needs of a moderately high-throughput molecular genotyping facility. The ICRISAT LIMS was developed with partial support from the GCP and is being used at the genomics laboratory facilities in ICRISAT-Patancheru. During 2005, the existing ICRISAT MS-Windows based LIMS was recoded as a platform independent, multi-user LIMS that manages workflow and information in the Applied Genomics Laboratory (AGL) at ICRISAT-Patancheru. The LIMS was developed as modules, thus any number of modules can be added depending upon the laboratory using the application. The goal of the LIMS transfer project was to transfer this application to the Biosciences Eastern and Central Africa (BecA) laboratories in Nairobi, Kenya and IITA in Ibadan, Nigeria for use by the scientists and technical staff at these locations. Both ICRISAT and IITA already have scientists based in BecA and have on-going research projects in genomics of sorghum and cassava, as well as other crops. Since the LIMS could be adapted through extensions and modifications, it was our interest to adapt the software to the genomics laboratory at BecA and IITA. This was seen as a better alternative to investing in proprietary software and then customizing it. It also provided an opportunity to document actual adoption of GCP products by other CGIAR, GCP and NARS scientists. Through teleconferences and demonstrations, BecA and IITA user requirements were gathered and additions needed in the LIMS application identified. These modifications were carried out at ICRISAT-Patancheru and the application transferred to BecA in August 2006 and IITA-Ibadan in November 2006. A workshop was held at BecA-Nairobi in August 2006, to provide hands-on training in its use and support to ILRI, IITA and ICRISAT staff. Several African NARS partners also participated in the workshop.

2006-35: Support for existing projects in SP1 on Germplasm data analysis (GDA)

Principal Investigator:
Marco Bink, WUR
Collaborators:
Hans Jansen, WUR
Fred van Eeuwijk, WUR
Marcos Malosetti , WUR
Xavier Perrier, CIRAD

Jean-Louis Noyer, CIRAD (to be confirmed)
Jean-Francois Rami, CIRAD (to be confirmed)
Kodjo Tomekpe, CARBAP (to be confirmed)
Jorge Rojas, PROINPA (to be confirmed)
Elisa Mihovilovich, CIP (to be confirmed)
Paula Hurtado, CIAT (to be confirmed)
Boris Sagredo, INIA (to be confirmed)
Lalith Perera, CRI (to be confirmed)
Emmanuel Okogbenin, NRCRI/CIAT (to be confirmed)
Ange-Marie Risterucci, Agropolis-CIRAD (to be confirmed)
Humberto Gomez, GCP / GSS
Jean-Cristophe Glaszmann, CIRAD
Carmen de Vicente, GCP

Mid-year report

[A] Portal / Website

The training materials have been entered into a first version of a website entitled 'Germplasm Data Analysis'. The training materials have been divided into three main sections; the three sections have been divided again into subsections:

- Molecular data : Quality checks
  : Structural analysis
  : Relatedness
  : Haplotyping
- Phenotypic data : Description of genetic material
  : Quality checks
  : Analysis of phenotypic data
- Sampling and analysis: Germplasm sampling
  : Disequilibrium analysis (including linkage analysis)
  : Association analysis (including QTL analysis)

[B] Bilateral Support / Helpdesk / Workshop

Two SP1 projects were actively supported by scientists of CIRAD, in the course of data analyses support new challenges/problems were discovered and are currently under investigation (see Appendix 2 for details).

As a follow-up on the successful workshop on statistical analysis of datasets from SP1 projects in October 2006 in Zaragoza, an additional workshop was planned for April 2007 in Zaragoza (again). The workshop intends to concentrate on the data generated by the Genotyping Support Service (GSS, H. Gomez). Unfortunately, this workshop has been postponed due to major delay in delivery of the marker data from GSS.

[C] Improved tools for Germplasm Data Analysis

CIRAD started to develop useful clearly explained step-wise procedures for reticulations in a diversity tree. A first algorithm was written and we are trying to reduce its complexity. The tree viewer in Darwin has been modified to display reticulations in a tree (see also appendix 1).

Tangible outputs delivered

The first version of web site has been sent on CD to Dr. Carmen de Vicente. It will be made available to partners and other interested scientist via the GCP website through a link.

Deviations from the workplan

At present the website mainly consists of links to presentations. The coming three months the website will be upgraded in order to become a medium for self reference and learning. It is aimed that each topic consists of a layered structure:

- an introduction
- an advanced section
- references to literature and links to available software
- examples
- exercises

Confirmation of partners is still pending as we wish to have the website entitled 'Germplasm Data Analysis' publicly available through the GCP website. The availability of this website will encourage these partners to collaborate in this project.

Data availability

Not applicable to this project

G4007.11: Refinement and distribution of iMAS for use by NARS and other user communities

Principal Investigators:
Subhash Chandra, ICRISAT
DA Hoisington, ICRISAT
Collaborators:
Jayashree Balaji, ICRISAT
Graham McLaren, IRRI
Guy Davenport, CIMMYT
Jose Crossa, CIMMYT

Mid-year report:

The goal of this three-year project (2005-07) is to develop an integrated decision support system, called **iMAS**, to seamlessly facilitate marker-assisted plant breeding by

- Integrating freely available quality software involved in the process of identification and application of trait-linked markers, and
- Providing simple-to-understand-and-use online decision guidelines to correctly use these software, interpret and use their outputs.

To achieve this goal, the project has been structured into the following logical sequence of nine activities.

A1: Analyse potentially useful free software
A2: Select software for inclusion in Imas
A3: Develop iMAS system
A4: Develop & incorporate online decision guidelines
A5: Test iMAS system
A6: Refine iMAS system
A7: Develop iMAS user manual
A8: Release of and Training in iMAS
A9: Consultation and support

Work completed during Oct 06 – May 07

A3: Revision of the mapping population module in light of the feedback received from end users in the Sept 06 workshop was undertaken and completed.

A4: Online decision guidelines for IRRISTAT, GMendel, PlabQTL, PopMin and GGT have been revised in the light of the feedback received from end users in the Sept 06 workshop.

A5-A6: The system, as it was revised, was simultaneously tested, further refined and problems resolved. The system was extensively tested on diverse real datasets during a one-week international workshop held at ICRISAT in May 07 by 20 prospective users who came from African, South American and Asian countries. The system was in general found to work well but requires some further refinement.

A7: A draft of the iMAS user manual has been developed and is being revised based on feedback received from the May 07 workshop participants.

Tangible outputs delivered

The Phenotyping and the Mapping Population modules were revised and extensively tested on 20 diverse real data sets in the May 07 workshop. Many participants were able to complete their analysis in a few hours for which they earlier needed months to complete.

Deviations from the workplan

It has been decided during the May 07 workshop to drop the Germplasm module for Association Analysis from the system.

Data availability
Not applicable tothis project

G4007.12: Development of tools and technology to increase the functionality of the GCP information platform

Principal Investigator:
Martin Senger, IRRI Crop Research Informatics Laboratory (IRRI-CRIL)
Visitor, European Bioinformatics Institute
Collaborators:
Graham McLaren, IRRI-CRIL
Richard Bruskiewich, IRRI-CRIL
Guy Davenport, CIMMYT-CRIL
Manuel Ruiz, CIRAD/Agropolis
Masaru Takeya, NIAS
Peter Rice, EBI

Mid-year report
Both technologies and tools have been developed, documented, and presented. See the list in the next paragraph. All code has been published and made available to the GCP developers.

Tangible outputs delivered
GCP Platform Component Validator
GCP Data Transformer specification
Soaplab2 development – phase 2 finished
Three new tutorial on GCP Platform
Established technology on sharing components (Maven)

Deviations from the workplan
Taverna web portal is delayed. Planned now for the second half of the year.

Data availability
All code and documentation have been published at cropforge repository. There are no biological data.

Subprogramme 5: Capacity-building and enabling delivery

2005-CB02: Development of training materials for a course in genomics and comparative genomics, and design of course curriculum

Principal Investigator:
Theresa Fulton, Institute for Genomic Diversity, Cornell University
Collaborators:
None directly

Final project report
Executive summary:
The goal of this project was to develop training materials and a curriculum for a course in genomics and comparative genomics, to be used either as a self-tutorial or as the basis for a course of approximately 2 weeks duration; including definitions of terms, illustrations of concepts, photographs, real-life examples, appropriate applications, lists of key references, and other items as appropriate.

This module is now complete and is available for download at
http://www.igd.cornell.edu/Comparative%20Genomics/Comparative%20Genomics%20Proj.html

2005-CB04: Development of training materials for a course in bioinformatics and design of course curriculum

Principal Investigator:
Richard Bruskiewich, IRRI
Collaborators:
Teresa Fulton, Cornell University

Mid-year report

- Locally recruited Filipino national staff biologist was hired as of January 2, 2007 for 12 months on this project to compile and organise course materials.

Tangible outputs delivered

- GCPWiki is hosting the evolving course materials at: http://cropwiki.irri.org/gcp/index.php/GCP_Online_Bioinformatics_Course.
- The course will be converted into a fully interactive web site with some expert advice by IRRI's Training Center and Communications and Publications Services.

Deviations from the workplan

- Delayed completion to December 31, 2007 (no cost extension requested and approved). ~70% completion of the course materials by the September GCP ARM.

Data availability

- N/A. Online course. Preliminary version could be presented at next ARM.

2005-CB05e: Development of reference molecular marker kits to analyse diversity of germplasm for the year 1 GCP crops (maize*)*

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborators:
Chaba Jampatong, National Corn and Sorghum Research Center
Susanne Dreisigacker, CIMMYT
Ana Lidia Gomez, CIMMYT

Final project report
Executive summary:
Dr. Jampatong Chaba arrived from the National Corn and Sorghum Research Center, Thailand, where she works with Dr. Pichet Grudloyma. Dr. Chaba has screened 1189 maize inbred lines from CIMMYT, IITA, and CAAS for 35 SSR markers. She has found 139 unique alleles using these 35 SSR markers, which can be all represented with a total of 65 genotypes. She re-amplified these to verify the alleles and make sure they do indeed amplify only one clear allele of the previously reported size, and also to check its banding pattern on manual acrylimide gels. Dr. Chaba has also searched through 468 maize populations for 26 SSR markers looking for unique alleles.

2005-CB05f: Development of reference molecular marker kits to analyse diversity of germplasm for the year 1 GCP crops (wheat*)*

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborators:
Li Genying, Crop Institute, Shandong Academy of Agricultral Sciences.
Susanne Dreisigacker, CIMMYT
Leticia Diaz, CIMMYT

Final project report
Executive summary:
Dr. Genyinig Li arrived on August 10, 2005, from the Crop Institute, Shandong Academy of Agricultural Sciences, where she works with Dr. Xia Xianchun. Dr. Li has screened 2506 wheat accessions and 38 markers to select a final 138 genotypes to re-fingerprint and develop the SSR allele kit. The kit consisted of 270 alleles amplified from these 38 SSR markers, all of which could be seen in the list of 138 genotypes, taken as a group.

2005-CB05i: Development of reference molecular marker kits to analyse diversity of germplasm for the year 1 GCP crops (barley*)*

Principal Investigator:
Michael Baum, ICARDA
Collaborators:
Dr. Wafa Choumane, ICARDA

Final project report
Executive summary:
In order to establish a microsatellite reference kit for barley, 13 different genotypes of barley were selected to be used as reference DNA samples and were analysed with 20 SSR and EST-SSR markers.

The accessions were selected on the basis of the geographical origin, the taxa, the domestication, the number of rows/spike, and the polymorphism at the microsatellite loci. The microsatellite primers were selected on the basis of clear amplification, the level of allele number, and the distribution on the

chromosomes. The total number of alleles detected in the DNA reference kit at 20 loci was 99 with an average of 4.95 alleles per locus. The genic
diversity ranged from 0.260 at the locus SCSS 14079 to 0.899 at the locus SCSSR 3907. The value of genetic diversity in the DNA reference kit is 0.6625.

2005-05j: Development of reference molecular marker kits to analyse diversity of germplasm for the year 1 GCP crops (chickpea*)*

Principal Investigator:
DA Hoisington, ICRISAT
Collaborators:
RK Varshney, ICRISAT
HD Upadhyaya, ICRISAT

Final project report
Executive summary:
A chickpea SSR marker reference kit has been developed based on the genotyping of the global chickpea composite collection (3000 accessions) with 35 simple sequence repeat (SSR) markers. The kit consists of three pools of chickpea accessions along with supporting documentation on the SSR markers, PCR and detection conditions, and the allele sizes for each of the 35 SSR loci. All details on the marker such as the primer sequence, annealing temperature, source of the markers, diagnostic alleles in the pools of the reference genotypes and allelic patterns of the reference genotypes in pool have been posted on ICRISAT's web page at http://www.icrisat.org/gt-bt/Marker_Kits.htm. The developed SSR kit will have a wide range of applications, especially for genetic diversity studies in chickpea. By using the markers and reference accessions in the kit, scientists in other laboratories will be able to compare the genotypic data they obtain on their germplasm with that obtain using the global composite collection.

2005-CB11a: Training course on plant genetic diversity and molecular marker-assisted breeding

Principal Investigator:
Marilyn Warburton, CIMMYT
Collaborators:
Patricio Hinrichson, INIA
Mario Paredes, INIA
Viviana Becerra, INIA
Carlos Magni, University of Chile
Jorge Franco, Universidad de la Republica
Boris Sagredo, INIA
Gisela Orjeda, Universidad Peruana Cayetano Heredia
Isabel Vales, Oregon State University
Eduardo Oyanadel, Pontificia Universidad Católica de Valparaíso
Mariano Bulos, Nidera Seeds
Phillip White, Horticulture Research International
Carmen de Vicente, Bioversity
Edith Talesnik - IFFIVE-INTA, Argentina and International Fund for Science

Final project report
Executive summary:
The GCP Training workshop entitled "Plant Genetic Diversity and Molecular Marker Assisted Breeding" was geared towards National Programme scientists with the desire and possibilities to utilise markers (via diversity analyses and Marker Assisted Selection) in their breeding programmes. This course was offered in conjunction INIA La Platina in Santiago, Chile. Attending the workshop were 17 women and 8 men

from 8 countries. Most held permanent positions as young scientists in universities or National Agriculture programmes, and a few were finishing up their dissertations. The students were without exception excellent and hard working. Many report that they have begun to form networks to keep working together after the workshop, including a rice and a potato network to date. The resource people for the course included breeders, molecular geneticists, and statisticians from both the public and private sector. There were 11 speakers from 6 countries, and all were rated well by the students. The topics covered during the course can be seen in the workshop schedule. The course budget came from the GCP, the Syngenta Trust and the Kirkhouse Trust. The total budget spent on the workshop was under budget, and the surplus were spent on minigrants for students and resource people. $2,500 each were awarded to one resource person and the INIA laboratory staff for their outstanding contributions to the course. Five grants were approved and funded for the students.

2005-CB13: The Interactive Resource Center & Helpdesk

Principal Investigator:
Theresa Fulton, Institute for Genomic Diversity at Cornell University
Collaborators:
Various members of IGD, including but not limited to:
Sharon Mitchell
Martha Hamblin
Charlotte Acharya
Others

Mid-year report
A small survey was sent out to assess next priorities of the Resource Center (only to sorghum users, to alert them of the new sorghum resources posted). Responses included the need for more protocols, and more detail in the protocols, and more tutorials, particularly on software and plant breeding topics, including phenotyping. A more comprehensive survey is planned in the coming months to assess: 1) if these are also the needs of the broader user community, and 2) more specifically which software programmes users need help with, and what topics more information is needed on.

To further awareness of the helpdesk, a number of new contacts were formed and reciprocal links posted, including:

- The African Molecular Marker Applications Network (AMMANET http://africancrops.net). I also sent an item to their listserve advising members of the IRC, which generated positive feedback.
- The Agricultural Biotechnology Network in Africa (ABNETA http://www.abneta.org/site/). A member of the ABNETA site (David Priest) contacted the IRC to send a note of thanks for the resources (in particular the "how to" section), we have now each posted reciprocal links, and plans have been made to exchange additional laboratory protocols.
- To better link to other GCP resources, links have been added to the GCP Bioinformatics Portal and the IP helpdesk.
- Updates and links to the ongoing Sorghum Sequencing project by the DOE have been posted periodically as well as a letter of information from the PI Andy Paterson. News about the cassava genome sequencing and new pearl millet markers being developed have also been posted.
- Questions recently received and answered by the Helpdesk team included queries about Mapmaker software, SNP data analysis, DNA extractions, and others.

Since the last report, approximately 2000 'unique' (as called by the Statistics counter) users have visited the site.

Tangible outputs delivered
New information posted on the site, as noted above.

Deviations from the workplan
None

Data availability
Links to new data have been posted in the news section (including new sorghum sequence data and news about the sequencing of the cassava genome.

2005-CB17: Reporting for product distribution: An asset inventory system for the Generation Challenge Programme

Principal Investigator:
Victoria Henson-Apollonio (CAS-IP, hosted by Bioversity International)
Collaborators:
Hanumanth Rao, ICRISAT
Pro bono legal support provided by Helen Cordell and Shawn Sullivan
Additional technical support provided by Thordis Moeller

Final project report
Executive summary:
To ensure the use of Generation Challenge Programme (GCP) assets, these research results need to be identified and to be free of constraints affecting their exchange and distribution. Because of the importance of reporting GCP on intellectual property (IP) and information associated with third party inputs, the GCP Consortium Agreement includes *Article 25 – IP Management* regarding GCP IP and their inputs. This article requires that each Consortium Member submit: (i) an annual inventory of background and pre-existing IP that has been used as inputs into GCP assets; and (ii) a list of IP that has been created by GCP-funded research at the institution. *Schedule 4 of* the consortium agreement *(IP Management),* contains a form detailing the way in which these assets are to be reported to GCP management. To make this reporting process more efficient and to allow GCP to compile a dynamic database of this information, this project has created an online version of Schedule 4 that assists individual grantees and their institutions in meeting GCP reporting requirements. This online inventory system contains a series of five user-friendly web pages with input fields and pull-down menus, which link to a database of existing project information in order to simplify the reporting process and to collate all information in one location automatically. In addition, there are links embedded in the reporting pages that allow users to access explanatory information. While this online system should facilitate an increase in reporting activity regarding GCP intellectual assets created by grantees, it should be monitored and revised over time to ensure full reporting and ease of use by both scientists and GCP management.

2005-CB21: Fellowships and travel grants (2005)

Principal Investigator:
M Carmen de Vicente, GCP
Collaborators:
N/A

Final project report
Executive summary:
A total of 23 applicants submitted a project proposal to the Fellowship Programme in two different calls. Eight fellowships were awarded. The origin of the winners was as follows: three from Africa (Senegal, Kenya and Nigeria), four from Asia (India--2, Indonesia and Pakistan) and one from America (Mexico). The Travel Grant calls for applications opened late 2004, and continued until all grants were awarded. A total of 42 applications were received through direct submission to SP5. Fourteen applicants from 10 countries received a travel grant for hands-on training at advanced research institutions, to attend international conferences, or to participate in training courses.

In addition, the GCP awarded 14 grants to attend the INTERDROUGHT II conference (of which one person could not attend due to visa issues), held in Rome, Italy, September 24-28, 2005. Applications (80) were received by the organisers of the conference. Ten awards were given for travel and registration and four for registration only.

2005-CB23: Genotyping support service

Principal Investigator:
M Carmen de Vicente, GCP
Collaborator:
Jean-Christophe Glaszmann, CIRAD

Final project report
Executive summary:
The Generation Challenge Programme (GCP) mandate includes unlocking the genetic diversity of crop germplasm by using genomics to discover the genes and alleles responsible for the expression of complex agronomic traits. The results of this research are useful for the biological sciences in general and especially for research applied to crop breeding, which is the continuous process of producing new genetic combinations to create varieties. With a better understanding of the traits that condition plant and crop performance, plant breeders can create better varieties faster, and provide farming communities with more suitable products. The GCP is committed to ensuring that this work reaches crop users to provide consumers with better and more abundant harvests as a result of the enhanced availability of better cultivars.

The Genotyping Support Service (GSS) was developed to facilitate the access of National Agricultural Research Systems (NARS) to genotyping technologies, and in turn to bridge the gaps between research undertaken in laboratories and that conducted in the field. With this service, the GCP offers cost-efficient genotyping services worldwide, access to data and support (which builds local capacity) and training in statistics for proper interpretation of genotype and phenotype data to raise the productivity of local scientists.

In the pilot phase, the GSS contacted 22 National Programmes working in one of the following crops: cassava, coconut, groundnut, *Musa* and potato. Out of the NARS that expressed interest, eight applications were selected to benefit from the service in this term. This phase entailed the identification of steps needing legal documents, consistent with the Consortium Agreement, and the preparation and use of those documents. It also tested different options of service providers, i.e. Consortium members versus outsourcing. Details on the description of services and number of samples genotyped are in the report. The training activity of the project was postponed to take place when all data become available.
A no-cost extension to this project was awarded until 15 May 2007.

2006-13: Targeting and impact analysis of Generation Challenge Programme (GCP) technologies

Principal Investigator:
Glenn Hyman, CIAT
Collaborators:
Peter Jones, CIAT
Sam Fujisaka, CIAT
Stan Wood, IFPRI
John Dixon, CIMMYT

Mid-year report
1) Geographic distribution and the magnitude of poverty in the 14 GCP priority farming systems.
An inventory of poverty maps was carried out for the 14 priority systems. The inventory

includes information on the dates of the poverty assessments, the types of variables used, the level of geographic detail and other information from each assessment. Income and consumption variables are within most of the maps. In a few cases, such as Pakistan, the variables are unique and will be difficult to compare across countries. We think we will have about 90% of the study covered with good secondary information on human welfare levels.

2) Crop-specific modeling of drought severity and type
A complimentary copy of the most recent DSSAT package has been obtained for the study. This was due to the CIAT-MarkSim/ICASA linkage. The environments of the farming systems have been examined and it was determined that for each crop studied a set of four representative environments should be chosen. It is probably that all the GCP crops might not be included in the analysis because there is no model available. For all the main crops, there is a good model, for some, such as cassava the model is rudimentary, but works. There is as yet no DSSAT model for sweet potato. Preliminary work has been done on conceptualising the FORTRAN routine that will calculate drought susceptibility indices.

3) Analysis of GCP priority farming systems and their patterns of technology adoption at disaggregated levels within each region
We have acquired some preliminary data on technology adoption from the CGIAR system-wide panel on impact assessment (SPIA). They reported adoption rates at national levels. Our aim is to find their original sources in an effort to know about the subnational distributions of adoption patterns. We will then link this information to the GCP priority farming systems. Short assessment of 4 farming systems were carried out and written into a draft report. The systems are Mesoamerican maize- beans, Southeast Asia lowland rice, South Asia rainfed mixed and South Asia rice-wheat.

4) Economic assessment of the potential market-scale impacts of a variety of household- level adoption scenarios
The economic assessment work is being carried out under the HarvestChoice project funded to IFPRI by the Gates Foundation. That project is developing a global household-level data set. Some activities include translation of data sets to English and review of variable that we plan to use.

Tangible outputs delivered
None delivered so far

Deviations from the workplan
There have been no deviations from the workplan**.**

Data availability
Poverty data sets have been collected. These will be made available to exclusively to GCP researchers. A catalogue that explains each poverty variable has been developed.

2006-14: *Ex ante* impact analysis of marker-assisted selection technologies supported by the Generation Challenge Programme (GCP)

Principal Investigator:
George W. Norton, Virginia Tech
Collaborators:
Jeffrey Alwang, Virginia Tech

Mid-year report

This impact assessment project began in 2007, with the first steps being to: (1) construct technology-impact pathways for rice and cassava molecular-assisted breeding projects, (2) gather secondary data on prices, production, area, and trade for the two crops in the target countries, and (3) gather crop loss data. One graduate student was hired on the project for the rice component beginning in January and a second graduate student for the cassava part beginning in May. Both students are preparing their thesis proposals.

Construction of technology-impact pathways: Reports for the rice and cassava projects were reviewed and a set of questions developed for the scientists involved in the two projects. A trip to IRRI was made in March by George Norton to interview the PI and collaborators of the Marginal Lands (rice) project to understand the steps in their research and the intended products, their timing, and likely geographic areas of impact. Notes from those reports, interviews, and other literature review are being summarised into a brief report to accompany the technology impact pathway map. That report is still being finalized. A brief trip was made to CIAT in April by George Norton to interview the PI and collaborators on the project for Pyramiding Useful Genes from Wild Relatives of Cassava into Elite Progenitors. The purpose for that trip as well was to understand the steps in the research and the intended products, their timing, and likely geographic areas of impact. Notes from those reports, interviews, and other literature review are also being summarized in a brief report and a technology impact pathway map.

Gather secondary data on prices, production, area, and trade: Disaggregated secondary data on production, area, prices, and trade for Bangladesh, India, Indonesia, and the Philippines for rice, and Uganda, Nigeria, Ghana, and Brazil for cassava are being assembled, drawing in part on the IFPRI/Minnesota database.

Gather crop loss data: Some existing data on crop losses in rice due to salinity and phosphorous retention in soils have been collected, but the effort is not yet completed for the target countries. Data on crop losses for CMD, green mites, white flies, horn-flies, and post-harvest physiological deterioration in the four target countries are partially collected.

Tangible outputs delivered

None

Deviations from the workplan

By the time the agreement was completed in February, one of the graduate students potentially slated for the project (the cassava part) had begun work on another project. However, a different student was hired in May and the unexpended wages for the first 4 months will be spent on a second graduate student for the cassava part (three students totally on the project, with the third one to be hired in August) so we should be back on schedule soon. We lost two weeks of work with the shootings at Virginia Tech in April, but that time will be made up and we do not anticipate a delay in completing the first year projected outputs.

Data availability

No data sets completed yet

2006-15: Fellowships and travel grants (2006)

Principal Investigator:
M Carmen de Vicente, GCP
Collaborators:
N/A

Final project report
Executive summary:
The first call for applications for 2006 fellowships was released on 1st December 2005, and the second one on 15th June 2006. A total of 38 applicantions were received and 6 awarded. The origin of the fellow winners was as follows: Ghana, Kenya, Bangladesh, Thailand, Egypt and Indonesia.

The Travel Grant calls for applications opened on 30th November 2006, and continued until all grants were awarded. A total of 101 applications were received through direct submission to SubProgramme 5 (SP5). Nineteen applicants from 13 countries received a travel grant for hands-on training at advanced research institutions, to attend international conferences, or to participate in training courses.

2006-28: Regional PGR courses

Principal Investigator:
Marja Thijssen, Wageningen UR
Collaborators:
Niels Louwaars, Wageningen UR
Walter de Boef, Consultant to Wageningen UR
Victoria Henson-Apollonio, CAS – Bioversity International
Zeze Sampaio, EMBRAPA

Final project report
Executive summary:
Much of the success of the GCP programme depends upon having access to materials and technologies under clear agreements. The purpose of this project was to create a curriculum and training materials for regional courses on genetic resource policies. The objectives of these courses are to create awareness, extend relevant knowledge and share experiences among scientists and science managers, enhancing their capacity to develop or strengthen institutional policies for handling Intellectual Property Rights and Access and Benefit Sharing. The primary beneficiaries are scientists and science managers from GCP partner organisations. The design of the curriculum on genetic resources issues with special focus on institutional policies and tools for Freedom to Operate, IPR and ABS, has been based on experiences of courses in Sweden, Germany/Bioversity International, but differs from these in specifically addressing the institutional context. The training materials have been developed by Wageningen UR with inputs from Bioversity International and EMBRAPA. Training materials were first used in two-week international course with similar objectives in Wageningen in June 2005 and June 2006 (outside the responsibilities of GCP), and finally tested in a regional workshop in Karaj, Iran in 2006, with a group of 22 professionals from Central West Asia and Africa. Participants worked in biotechnology, crop improvement, seed supply and genetic resources conservation and were coming from both NARS and civil society. In general the course was evaluated very well, considered useful and innovative, both in terms of content and teaching methods. With a pre- and post course test, a clear increase in knowledge on and insight in genetic resources policies and their institutional implications could be measured.

2006-36: Capacity-building & research activities in Sub-Saharan Africa

Principal Investigator:
ACCI/University of KwaZulu-Natal
Collaborators:
Prof MD Laing, Director ACCI

Mid-year report:
The project involves hiring a GCP Molecular Plant Breeder, to join the staff of the ACCI. The scientist will then teach the ACCI students MAS techniques, supervise PhD projects, develop projects with GCP collaborators in the 13 African countries in which we operate, and assist in developing a regional genotyping and protein sequencing centre.

Progress to date:
The key activity to date has been to appoint the Molecular Breeder and to get him cleared to work in South Africa.

Timetable of the UKZN appointment process:

- End of September 2006: GCP commits to the project. Contract is signed, which allows the appointment process to be initiated, but only then. UKZN does not allow the process to be initiated until the contract is concluded.
- October 2006: selection committee is chosen (DVC of College, Dean, and 7 others), and a draft advertisement drawn up.
- November 2006 Advertisement: the entire selection committee reviews, edits and passes the advertisement. Advertisement is placed in local newspapers and international listservers
- December 20, 2006: Selection committee meets to review the applications and shortlist candidates.
- December 20 2006-January 20th 2007 UKZN Xmas break, during which no HR activities occur.
- March 2007 Telephonic interviews. Selection Committee chose Dr Schafleitner, a potato and sweet potato molecular breeder at CIP in Peru. Negotiations re conditions of the appointment take place between between Dr Schafleitner and UKZN.
- April 2007 Agreement on contract. Human Resources Division at UKZN ask Dr Schafleitner for documentation for application for a work permit.
- April 2007 Submission of documentation to Dr Schafleitner to SAQA, the SA Quality Assurance unity of Dept of Education, who then check that his academic qualifications are legitimate.
- June 2007. SAQA certification in provided, some 3 months later.
- July 2007 Full documentation submitted to Dept of Home Affairs, requesting a Work Permit for a Non-South African. This will take a further 3-4 months. HR do not foresee unusual complications as Dr Schafleitner will be appointed under the Scarce Skills provision which has been used for many external appointments to UKZN.
- September 2007. Work permit should be granted. Then Dr Schafleitner may purchase air tickets and move to South Africa.
- October/November 2007 Dr Schafleitner and family will be in South Africa and he will be at work with the ACCI.

Inherent delays in the hiring process:
No action could take place until the GCP / UKZN contract was signed. The UKZN process was followed as soon as we got the go ahead and signed the contract with the GCP. Given the size of the committee, and that it had both the DVC and Dean of the Faculty (with very full diaries), these processes happened as fast as was possible.

SAQA certification is a national requirement. We have no control over their efficiency. They took 3 months to issue the certificate.

Home Affairs grants work permits. Again, we have no control over their efficiency. Typically they take 3-5 months to grant a foreigner a work permit. However, the appointment of a Professor of Microbiology took place in October 2006, and he is still waiting for a work permit.

South Africa has some of the most complex labour relations and employment requirements of any country in the world. UKZN is committed to implementing this in full. We also have to deal with a bureaucracy in South Africa which has a limited capacity. Hence, the process of appointing a foreign academic to UKZN is now taking 9-12 month to complete, if one can find a suitable candidate.

Impact on the overall objectives of the GCP / ACCI Project
The contract was dated at the end of September 2006, and the narrative of the contract proposed that the selection process would be carried out for an appointment early 2007.
The appointment was made in early 2007, as expected. However, the Molecular Breeder will only be in place at UKZN by October / November 2007, due to external forces.

It is therefore proposed that we amend the timetable of the project,
The reality is that the soonest we can hope to have Dr Schafleitner in South Africa, working at the ACCI, is October-November 2007. Hence, we need to extend the timeline of the project, ideally to December 2012 (to coincide with the graduation of a cohort of students), to ensure an effective 5 years of delivering real work and actual outcomes from the project.
Note that it has no financial implications for the project, other than the GCP giving the ACCI permission to roll over funds for 2007 into 2008.

Parallel activities
In the meantime we are working on three other directions:

1. UKZN has appointed a senior scientist to manage the Molecular Biology Unit, which houses the DNA and Protein sequencers and the Real-time PCR. And we have found funds to upgrade the equipment and software, so the unit should be substantially enhanced by the time Dr Schafleitner arrives here. We are also seeking to enhance the management of the MBU so it runs more smoothly, and I hope to appoint Dr Schafleitner onto the management committee, so he can assist in developing the MBU into a regional genotyping unit. This is where most of his research work will take place.
2. We have also been successful in securing access to a molecular biology laboratory for Dr Schafleitner and his students to work in. Given the severe shortage of space in this building, it was a major coup to succeed in this venture.
3. I have been working with Dr Phillippe Monveaux of GCP, to develop a regional project on MAS for maize streak virus resistance with Mozambique, using the MBU and Dr Schafleitner's expertise to do the genotyping.

Tangible outputs delivered to date
Dr Schafleitner has been appointed.
A molecular biology lab has been secured for him.
The MBU resources, which he will need, have been rejuvenated and upgraded.

Deviations from the workplan
Essentially the project is running about 9 months behind schedule due to delays with the initial project evaluation by GCP, which then delayed the hiring process as the go-ahead came over December, when UKZN HR department stops hiring people for 6 weeks. And now we are delayed by bureaucratic

processes of the SA Dept of Home Affairs. The reality is that hiring a foreign scientist into South Africa typically takes 6-12 months.

The net result is that we need to request a change to the timetable of the project, requesting a 10-12 month delay in the implementation of the active part of the project, extending the life of the project to December 2012.

Data availability
Not applicable to this project

G4007.14: Fellowship and travel grants (2007):

Principal Investigator:
Carmen de Vicente, GCP
Collaborators:
None

Mid-year report:
Fellowships

- Date of announcement: December 8th, 2006.
- Number of applications received: 12
- Winners: Table 1 summarises the winners of the single annual call for Fellowships in 2007.

Travel Grants
Travel grants for hands-on training:

- Date of announcement: January 30th, 2007
- Number of applications received: 25
- Winners: Table 2 summarises the winners of the single annual call for Travel Grants in 2007

Travel grants for the GCP 2007 Annual Research Meeting, Benoni, South Africa,12–16 September, 2007:

- A selection was made between NARS partner candidates from African collaborators.
- Winners: Table 3 summarises the winners of travel grants to attend the forthcoming ARM.

Travel grants awarded for the GCP Workshop on product management and delivery in GCP rice research in Asia, Bangkok, Thailand, November 6th and 8th, 2007:

- Selection of winners was made amongst the collaborators of the GCP rice research projects in Asia
- Winners: Table 4 summarises the winners of travel grants to attend the GCP workshop on product management and delivery in GCP rice research in Asia.

Tangible outputs delivered
Five (5) fellowships awarded.
Eight (8) travel grants for hands on training awarded.
Twenty-four (24) travel grants for participation in special conferences/meetings awarded.

Deviations from the workplan
None

Data availability
No data available

For more information on the Generation Challenge Programme, please visit our Web site at www.generationcp.org

or contact us:
info@generationcp.org
Apdo. Postal 6-641 06600 Mexico D.F, Mexico
Telephone +52 55 5804 2004, Fax: +52 55 5804 7558

Consortium members

African Centre for Gene Technologies (ACGT) • Agropolis • Bioversity International • Brazilian Agricultural Research Corporation (EMBRAPA) • Chinese Academy of Agricultural Sciences (CAAS) • Cornell University • Indian Council for Agricultural Research (ICAR) • International Center for Tropical Agriculture (CIAT) • International Maize and Wheat Improvement Center (CIMMYT) • International Potato Center (CIP) • International Center for Agricultural Research in the Dry Areas (ICARDA) • International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) • International Institute for Tropical Agriculture(IITA) • International Rice Research Institute (IRRI) • John Innes Centre • National Institute of Agrobiological Sciences (NIAS-Japan) • Wageningen University • Africa Rice Center (WARDA)

Provisional Members:

Centro de Investigación y de Estudios Avanzados (CINVESTAV) • Institut National de la Recherche Agronomique (INRA) • Instituto Agronomico per l'Oltremare (IAO) • National Center for Genetic Engineering and Biotechnology (BIOTEC)