

Generation Challenge Programme

CULTIVATING PLANT DIVERSITY FOR THE RESOURCE POOR

Proceedings of GCP 2005 Annual
Research Meeting:
Mid-Year Project Reports



Generation Challenge Program, 2005. *2005 Generation Challenge Program Competitive and Commissioned Research Project Summaries*. Mexico D.F.: Generation Challenge Program.

© Generation Challenge Program 2005. All rights reserved.

Table of Contents

COMPETITIVE GRANTS	1
1. IDENTIFYING GENES RESPONSIBLE FOR FAILURE OF GRAIN FORMATION IN RICE AND WHEAT UNDER DROUGHT	1
2. REVITALISING MARGINAL LANDS: DISCOVERY OF GENES FOR TOLERANCE OF SALINE AND PHOSPHORUS DEFICIENT SOILS TO ENHANCE AND SUSTAIN PRODUCTIVITY.....	3
3. IDENTIFYING THE PHYSIOLOGICAL AND GENETIC TRAITS THAT MAKE CASSAVA ONE OF THE MOST DROUGHT TOLERANT CROPS	5
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 (TO BE FUNDED AT 50%)	5
5. UNLOCKING THE GENETIC DIVERSITY IN PEANUT'S WILD RELATIVES WITH GENOMIC AND GENETIC TOOLS TARGETED SUBPROGRAMME: SP3 - TRAIT CAPTURE FOR CROP IMPROVEMENT.....	6
6. MARKER DEVELOPMENT AND MARKER-ASSISTED SELECTION FOR STRIGA RESISTANCE IN COWPEA.....	9
7. MEASURING LINKAGE DISEQUILIBRIUM ACROSS THREE GENOMIC REGIONS IN RICE	10
8. TARGETED DISCOVERY OF SUPERIOR DISEASE QTL ALLELES IN THE MAIZE AND RICE GENOMES	12
9. DEVELOPMENT OF LOW-COST TECHNOLOGIES FOR PYRAMIDING USEFUL GENES FROM WILD RELATIVES OF CASSAVA INTO ELITE PROGENITORS	15
10. EXPLORING NATURAL GENETIC VARIATION: DEVELOPING GENOMIC RESOURCES AND INTROGRESSION LINES FOR FOUR AA GENOME RICE RELATIVES	19
11. FUNCTIONAL GENOMICS OF CROSS-SPECIES RESISTANCE TO FUNGAL DISEASES IN RICE AND WHEAT (CEREALIMMUNITY).....	23
12. DROUGHT TOLERANT RICE CULTIVARS FOR NORTH CHINA AND SOUTH/SOUTHEAST ASIA BY HIGHLY EFFICIENT PYRAMIDING OF QTLs FROM DIVERSE ORIGINS	26
13. DEVELOPMENT OF INFORMATIVE DNA MARKERS THROUGH ASSOCIATION MAPPING IN MAIZE TO IMPROVE DROUGHT TOLERANCE IN CEREALS	27
14. CHARACTERISATION OF GENETIC DIVERSITY OF MAIZE POPULATIONS: DOCUMENTING GLOBAL MAIZE MIGRATION FROM THE CENTRE OF ORIGIN	28
15. DETERMINATION OF A COMMON GENETIC BASIS FOR TISSUE GROWTH RATE UNDER WATER-LIMITED CONDITIONS ACROSS PLANT ORGANS AND GENOMES	30
16. ISOLATION AND CHARACTERISATION OF ALUMINUM TOLERANCE GENES IN THE CEREALS: AN INTEGRATED FUNCTIONAL GENOMIC, MOLECULAR GENETIC AND PHYSIOLOGICAL ANALYSIS.....	32
17. ALLELE MINING BASED ON NON-CODING REGULATORY SNPs IN BARLEY GERMPLASM	34
SP1 COMMISSIONED GRANTS	38
1A. COMPLETING GENOTYPING OF COMPOSITE GERMPLASM SET OF BARLEY	38
1B. COMPLETING GENOTYPING OF COMPOSITE GERMPLASM SET OF WHEAT.....	41
1C. COMPLETING GENOTYPING OF COMPOSITE GERMPLASM SET OF SORGHUM	41
1D. COMPLETING GENOTYPING OF COMPOSITE SET OF CHICKPEA	42
1E. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, MAIZE	43
1F. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, WHEAT	45
1G. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, RICE.....	46
1H. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, SORGHUM.....	49
1I. IN BARLEY GENOTYPING OF CGIAR CHALLENGE PROGRAM	50
1J. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, COMMON BEAN	51
1K. COWPEA GENOTYPING.....	53
1L. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, CASSAVA	54
1M. GENE POOL STRUCTURE OF CULTIVATED POTATOES ASSESSED BY SSR MARKER ANALYSES	55
1N. GENOTYPING OF COMPOSITE GERMPLASM SET, TIER 1, MUSA.....	56
2. SUPPORTING DISTRIBUTION OF REFERENCE GERMPLASM	58
3A. DEVELOPMENT OF COMPOSITE COLLECTION OF TIER 2 (ORPHAN) CROPS.....	58
3B. MOLECULAR CHARACTERISATION OF TIER 2 (ORPHAN) CROPS	59
3C. APPLICATION OF MOLECULAR MARKERS FOR GENE POOL DIVISION AND HETEROSIS ESTIMATION UNDER DROUGHT STRESS CONDITIONS IN SWEET POTATO	60

3D. ASSESSMENT OF GENETIC DIVERSITY OF WEST AFRICAN <i>DIOSCOREA SP</i> COLLECTION	63
3E. GENOTYPING A COMPOSITE GERMPLASM SET OF LENTIL	64
3F. MOLECULAR CHARACTERISATION OF TIER 2 (ORPHAN) CROPS.....	65
3G. MOLECULAR CHARACTERISATION OF TIER 2 CROPS: COCONUT	65
4. VALIDATION OF DIVERSITY ARRAYS TECHNOLOGY (DART) AS A PLATFORM FOR WHOLE GENOME PROFILING IN ORPHAN CROPS.....	66
5. ASSESSING ECOTILLING AS A METHODOLOGY FOR TARGETED GENOTYPING AND SNP DISCOVERY	68
6. SUPPORTING EMERGENCE OF REFERENCE DROUGHT TOLERANCE PHENOTYPING CENTRE	70
7. WHOLE PLANT PHYSIOLOGY MODELING OF DROUGHT TOLERANCE IN CEREALS	76
8. POPULATION STRUCTURE, PHENOTYPIC INFORMATION, AND ASSOCIATION STUDIES IN LONG- GENERATION CROPS.....	77
SP2 COMMISSIONED GRANTS	79
9. SYSTEMATIC EVALUATION OF RICE MUTANT COLLECTIONS FOR CONDITIONAL PHENOTYPES WITH EMPHASIS ON STRESS TOLERANCE.....	79
10. COLLECTION, DISTRIBUTION, PHENOTYPING, AND GENOTYPING DIRECTED TOWARDS UTILISATION OF EXISTING WHEAT GENETIC STOCKS TO ENHANCE TOLERANCE/RESISTANCE OF WHEAT CULTIVARS TO ABIOTIC AND BIOTIC STRESSES WITH EMPHASIS ON DROUGHT.....	81
11. LEGUME MUTANT RESOURCE DEVELOPMENT	86
12. A SATURATED POTATO MUTANT POPULATION FOR FUNCTIONAL GENOMICS AMONG SOLANACEAE AND TUBER CROPS	87
13. CROP GENE EXPRESSION PROFILES AND STRESS-GENE ARRAYS	88
14A. STRESS RESPONSE-ENRICHED EST RESOURCES FOR TARGETED SPECIES—PEARL MILLET....	90
14B. STRESS RESPONSE-ENRICHED EST RESOURCES FOR TARGETED SPECIES—COWPEA.....	92
15. MUSA GENOME FRAME-MAP CONSTRUCTION AND CONNECTION WITH THE RICE SEQUENCE ...	92
16. VALIDATION OF CONSERVED ORTHOLOGOUS MARKERS	97
17. COMPARATIVE QTL MAPPING FOR DROUGHT TOLERANCE.....	99
35. SEQUENCING MULTIPLE AND DIVERSE RICE VARIETIES: CONNECTING WHOLE-GENOME VARIATION WITH PHENOTYPE.....	100
SP3 COMMISSIONED GRANTS	101
18. DEVELOPMENT OF LOW COST GENE BASED TRAIT ASSAY TECHNOLOGIES IN CEREALS	101
19. EVALUATION AND DEPLOYMENT OF TRANSGENIC DROUGHT-TOLERANT VARIETIES	105
20. OPTIMISING MARKER-ASSISTED BREEDING SYSTEMS FOR DROUGHT TOLERANCE IN CEREALS THROUGH LINKAGE OF PHYSIOLOGICAL AND GENETIC MODELS	107
21. PLANNING FOR EFFECTIVE PRODUCT DEVELOPMENT, DELIVERY, AND USE.....	110
SP4 COMMISSIONED GRANTS	111
22. DEVELOPMENT OF GENERATIONCP DOMAIN (DATA) MODELS.....	111
23. IMPLEMENTATION OF WEB SERVICES TECHNOLOGY IN THE GCP CONSORTIUM	114
24. APPLICATION OF MOBY FOR GENERATIONCP CONSORTIUM.....	117
25. CREATION AND MAINTENANCE OF TEMPLATES FOR GCP DATA STORAGE IN REPOSITORIES...	118
26. CREATION AND MAINTENANCE OF GENERATION CP REPOSITORY	120
27. INTEGRATION OF THE HIGH PERFORMANCE COMPUTING (HPC)-FACILITIES IN THE GENERATION CP TOOLBOX.....	121
28. IMPROVEMENT OF QUALITY OF EXISTING GCP DATABASES	122
29A. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY (CIAT).....	124
29B. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY (CIMMYT).....	125
29C. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY (CIP).....	125
29D. CREATION OF BIO-INFORMATICS CAPACITY FOR CENTRAL AND WEST ASIA AND	127
29E. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY AT ICRISAT	128
29F. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY	129
29G. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY (IPGRI)	129
29H. CREATION OF INSTITUTIONAL BIOINFORMATICS CAPACITY	130
30. DEVELOPMENT OF DECISION SUPPORT SYSTEMS FOR SAMPLING GERMPLASM	130
31. DEVELOPMENT OF ORTHOLOG-FUNCTION DISPLAY TOOLS	131

32. DEVELOPMENT OF CROP GENE EXPRESSION DATABASE AND DATA MINING TOOLS	133
33. DEVELOPMENT OF AN INTEGRATED DECISION SUPPORT SYSTEM FOR MARKER-ASSISTED PLANT BREEDING	134
34. GENERATIONCP USE CASE AND SOFTWARE ENGINEERING COLLABORATION AND MANAGEMENT	136
SP5 COMMISSIONED GRANTS	137
CB1. TRAINING PROGRAMME ON GENETIC DIVERSITY ANALYSIS OF GERMPLASM	137
CB2. DEVELOPMENT OF TRAINING MATERIALS FOR A COURSE IN GENOMICS AND COMPARATIVE GENOMICS, AND DESIGN OF COURSE CURRICULUM	138
CB4. DEVELOPMENT OF TRAINING MATERIALS FOR A COURSE IN BIOINFORMATICS AND DESIGN OF COURSE CURRICULUM	139
CB5. DEVELOPMENT OF REFERENCE MOLECULAR MARKER KITS TO ANALYZE DIVERSITY OF GERMPLASM FOR THE YEAR 1 GCP CROPS - RICE	139
CB8. FUNCTIONAL GENOMICS TO IMPROVE AFRICAN CROPS	140
CB9. USE OF MOLECULAR MARKERS FOR MINING USEFUL ALLELIC DIVERSITY – A SUMMARY OF SP1 GENOTYPING FOR GERMPLASM SCIENTISTS	140
CB10. GR POLICIES: A WORKSHOP SESSION IN CHINA DEVOTED TO GENETIC RESOURCE POLICIES	141
CB11. CIMMYT PLANT GENETIC DIVERSITY AND MOLECULAR MARKER ASSISTED BREEDING: A TRAINING COURSE	141
CB11. CORNELL UNIVERSITY PLANT GENETIC DIVERSITY AND MOLECULAR MARKER ASSISTED BREEDING: A TRAINING COURSE	142
CB11. ICRISAT PLANT GENETIC DIVERSITY ANALYSIS AND MARKER-ASSISTED BREEDING	146
CB12. WRITING QUALITY PROJECT PROPOSALS THAT CONNECT AGRICULTURAL SCIENTISTS, STAKEHOLDERS, AND DONORS	149
CB13. THE INSTITUTE FOR GENOMIC DIVERSITY’S INTERACTIVE RESOURCE CENTRE	153
CB14. REGIONAL PGR COURSES	154
CB15. DISTANT POLICIES: A DISTANCE LEARNING MODULE FOR SCIENTISTS ON GENETIC RESOURCE POLICIES AND THEIR IMPLICATIONS FOR FREEDOM-TO-OPERATE	155
CB16. IP MATTERS: AN INTELLECTUAL PROPERTY AND ACCESS & BENEFIT SHARING-HELPDESK AND ON-LINE-RESOURCE FOR THE GCP COMMUNITY, PARTNERS AND STAKEHOLDERS	156
CB17. REPORTING FOR PRODUCT DISTRIBUTION: AN ASSET INVENTORY SYSTEM FOR THE GENERATION CHALLENGE PROGRAMME	156

COMPETITIVE GRANTS

1. Identifying Genes Responsible for Failure of Grain Formation in Rice and Wheat under Drought

Principal Investigator:

H. Renee Lafitte, IRRI

Co-Principal Investigators:

R. Richards, CSIRO, Australia

A.G. Condon, CSIRO, Australia

L. McIntyre, CSIRO, Australia

J. Bennett, IRRI

K. McNally, IRRI

G. Atlin, IRRI

S. Kikuchi, NIAS

R.C. Babu, Tamil Nadu Agricultural University, India

Zhengqiang Ma, Nanjing Agricultural University, China

Collaborating Scientist:

S. Robin, Tamil Nadu Agricultural University, India

MID-YEAR REPORT

The purpose of this project is to identify opportunities to enhance reproductive-stage drought tolerance in rice and wheat through physiological, genetic, and molecular analyses of two yield determinants that are highly sensitive to field-level stress—panicle exertion and floret fertility. The specific objectives of the project are:

1. To use rice and wheat genotypes with contrasting behavior under stress to identify the candidate genes that underlie differences in drought tolerance, using microarray analysis.
2. To short-list and validate candidate genes using tools such as QTL mapping, gene expression studies with recombinant inbred lines, reverse genetics of insertion and deletion mutants, and segregation analysis.
3. To identify novel alleles of the validated genes in genetic resources and assess their impact on relevant physiological traits under stress.

A key hypothesis under test is that yield-forming processes and stress-adaptive responses are in conflict (as illustrated by GA/ABA antagonism) under drought stress. Hormone treatments and analyses will also be used to specifically examine this question.

At heading, rice peduncles elongate at a rate of ~6 cm per day for about 4 days under well-watered conditions, a rate that can fall to zero under drought stress. Re-watering allows elongation to resume but a window of opportunity has been narrowed during stress, so that full elongation (~35 cm) is not achieved and a fraction of the panicle remains trapped within the flag leaf sheath. The trapped florets are usually sterile. Peduncle elongation in wheat is often insensitive to drought stress, and florets that are exerted during drought stress are quickly protected by such processes as deposition of waxy cuticles to reduce non-stomatal transpiration. We are interested in identifying the genes responsible for any sensitivity of wheat to reproductive-stage drought stress (during meiosis and grain filling) and those genes responsible for the much higher degree of sensitivity of rice to stress (during flowering). A key question is why peduncle elongation is drought-sensitive in most rice varieties but usually drought-insensitive in wheat. One possibility is that, for a given level of stress, the insensitive genotypes protect their peduncles from the direct effects of stress through

maintenance of peduncle water status, so we are learning how to equalize stress across genotypes and crops in spite of wide variation in root and shoot architecture, including major differences in the size, density and behavior of stomata.

Another possibility is that stored fructans in wheat tissues provide a superior source of carbon during drought stress and allow peduncle elongation to continue. The fructan fructosyltransferases (FFTs) of wheat evolved from vacuolar invertases but the complete sequencing of the rice genome revealed that rice contains two vacuolar invertase genes (*OsVIN1* and *OsVIN2*) and no FFT genes. Furthermore, expression of recombinant *OsVIN1* and *OsVIN2* cDNAs in the yeast *Pichia pastoris* enabled us to show that these genes encode authentic invertases that lack significant FFT activity, so we do not consider it worthwhile to screen rice germplasm for undiscovered fructan accumulators or search for growth or stress conditions that might induce fructan accumulation in rice. Efforts to engineer fructan accumulation in rice through the introduction of FFT genes from wheat is a more promising approach, but it will, however, have to take note of the high fructan hydrolase activity of *OsVIN1*.

A third possibility is that sensitive rice genotypes do not use available carbohydrate stores (sucrose and starch) efficiently for peduncle elongation during drought stress and that more efficient usage would improve elongation. The small size of the peduncle before elongation gives it only a very limited capacity to store carbohydrates for its own elongation and these stores are in any event used rapidly at the start of elongation. More important are the starch and sucrose stores of the leaf sheaths. To benefit from these stores, rice must convert the starch to sucrose and the released sucrose must be taken up from the phloem of the peduncle, especially at the base where cell division and elongation occur. We have shown that the starch reserves of the leaf sheaths are indeed degraded during drought stress at heading (whereas in well-watered plants they would be mobilized only during grain filling), through activation of specific members of the α - and β -amylase gene families. By contrast, the cell-wall invertases of the peduncle are all down-regulated by stress, so that the diversion of sucrose from the phloem to the peduncle (via the combined action of cell-wall invertases and hexose transporters) does not occur. The most active cell-wall invertase gene in peduncles is *OsCIN2*, and its expression occurs around the phloem at the base of the peduncle. Its expression is driven by GA and is blocked by ABA, so we conclude that ABA-GA antagonism is likely to play a large role in the inefficient use of carbohydrate by peduncles during drought stress.

We are using proteomics and microarray analysis to characterize further the signal transduction pathways by which GA and ABA exert their effects on peduncle elongation. Chips from Agilent (with ~22K genes) and Beijing Genomics Institute (with ~60K genes) have been hybridized with rice and wheat RNA. Hybridizations of the 22K array with three biological replications of peduncle RNA from well-watered, drought-stressed and re-watered IR64 plants and well-watered plants of IR64 mutant (*eui10*) showing rapid peduncle elongation have been completed. Analysis of the data is underway. Detailed RT-PCR examination of the ABA-GA signal transduction pathways in peduncles has also been initiated, through examination of gene families of the ABRE-binding transcription factors and their post-synthetic modulators (the PKABA1-like protein kinases and protein phosphatases 2C). Of particular interest is the behavior of these genes at the base of the peduncle where cell division and elongation occur.

Wheat florets are most drought-sensitive at the meiosis stage and the grain filling stage. By contrast, rice florets are often most sensitive at the flowering stage. We are using microarray analysis and proteomics to compare gene expression in florets inside and outside the flag leaf

sheath. Of particular interest is the possibility that ethylene accumulates more in unexserted florets and contributes to sterility through alteration in the ABA-GA balance. Another focal point is the impact of exsertion on the deposition of cuticular wax on the surface of the exserted florets as a mechanism to reduce non-stomatal transpiration and thus delay reduction in water status.

In both wheat and rice, we have a strong interest in carbohydrate allocation to developing floral organs and cell- and tissue-types. As in the peduncle, drought stress down-regulates all cell-wall invertases of the anthers and therefore disrupts the flow of carbon to anthers and pollen grains. We are using RNA *in situ* hybridization to identify which cell-wall invertases and hexose transporters operate in each of the tissues of rice and wheat showing drought-sensitive development.

A segregating population of ~900 lines has been developed to the F4 stage for the cross IR64 x Moroberekan. These two parents differ significantly in their tolerance to drought at heading. In particular, under drought stress in the IRRI Phytotron, Moroberekan shows greater floret fertility in the top four rachis branches than IR64 and this is correlated with clumping of the pollen in IR64. Pollen clumping can reduce the number of pollen grains released onto stigmatic surfaces for fertilization or the ability of pollen to germinate on the stigmas. We are currently using the segregating population to test the hypothesis that the accumulation and breakdown of a particular anther glycoprotein governs the self-adhesion and stress tolerance of the floret.

Tangible outputs delivered:

Publication,

Ji X.M., Raveendran M., Oane R, Ismail A, Lafitte R, Bruskiwich R, Cheng S.H., Bennett J. Tissue-specific expression and drought responsiveness of cell-wall invertase genes of rice at flowering. *Plant Molecular Biology* (in press).

Deviations from the work plan:

None.

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

Co-Principal Investigators:

Matthias Wissuwa, IRRI

Glenn B. Gregorio, IRRI

David J. Mackill, IRRI

Eduardo Blumwald, University of California, Davis,

Emmanuel Delhaize, CSIRO, Australia

Zeba Seraj, Dhaka University, Bangladesh

Masdiar Bustamam, Indonesian Centre for Agricultural Biotechnology and Genetic Resources and Research Development, ICABGRRD

Collaborating Scientists:

Massahiro Yano, NIAS

Timothy J. Close, University of California, Riverside

MID-YEAR REPORT

- Planning workshop was conducted during 17-18 May, 2005. All PIs on the project participated beside collaborators from ABRII, Iran. Collaborators from UC Riverside participated through a telecom conference during the meeting. The meeting reviewed the progress made by each group and refined the workplan and responsibilities of each team.
- Personnel hiring in progress (Consultant on board and a postdoctoral research fellow will join in August 15)
- Annotated all the genes in *Pup1* region and studied their expression using RT-PCR. Two genes were identified as putative candidates for further analysis
- Selection of parental genotypes for backcrossing *Saltol* and *Pup1* after consultation with breeders and collaborators. Recurrent parents for *Saltol*: IR64, Swarna+sub1, Samba Masuri and BR28; *Pup1*: IR64, IR36, IR71525-19-1-1, IR74371-3-1-1 (aerobic varieties)
- Backcrossing initiated (i.e. planting of parental lines).
- Started screening DNA markers for polymorphism between parents
- Allelic screening using five putative DNA markers for salinity tolerance (*Saltol*) on 90 elite salt tolerant genotypes and parental lines completed, data is being analyzed.
- Identified all the genes located in the *Saltol* region and the borders regions limiting *Saltol* in chromosome I. Using information available in the databases, the function of each gene was analyzed and the genes were classified according to their putative functions (see figures).
- Initiated analysis of the expression of these genes using Northern blots and RT-PCR (in progress).
- Initiated greenhouse experiments using Pokkali, FL478 (tolerant) and IR29 (sensitive). The plants were exposed at 100 mM NaCl (50 mM NaCl stepwise) and RNA was isolated at different stages of growth. The RNA is being used for Northern blots, RT-PCR and DNA microarray comparison (using rice DNA arrays of Professor Ju-Kon Kim of Myongji University, Korea who is collaborating with UCD).
- Kasalath BAC clones in *Pup1* region identified
- Comparison of genome sequences indicated genomic rearrangements between *indica* and *japonica* genotypes in *Pup1* region
- Established experimental protocol and conducted one experiment under variable levels of P using NILs that vary in *Pup1* introgression obtained from IRRI. Assessed few traits including root length, dry matter, and P content.
- Evaluated predicted genes at *Pup1* region and investigated their ESTs in databases. Based on these result Real Time Beacon primers designed for 11 genes. Primers for other genes in the region will be designed after confirmation using primers from IRRI.
- Distinct difference between lines differing in *Pup1* introgression were observed and subsequently the main experiments started using 3 different levels of P. Leaf and root samples will be collected for proteomics, real time PCR, and physiological measurements.

Preliminary data has been obtained using Northern blots and DNA arrays. These data is preliminary and will be confirm during 2005.

Most activities were delayed a little because it took little longer for the contracts to be in place and the budget to be dispatched.

3. Identifying the Physiological and Genetic Traits that Make Cassava One of the Most Drought Tolerant Crops

Principal Investigator:

Alfredo Alves, EMBRAPA/CNPMPF

Co-Principal Investigators:

Hernán Ceballos, CIAT

Martin Fregene, CIAT

Yvonne Lokko, IITA

Tim Setter, Cornell University

MID-YEAR REPORT

Based on the outcomes from the cassava breeding programs carried out by Embrapa (Brazil) and CIAT (Colombia), in the last 15 years, 40 cassava varieties were identified as contrasting for drought tolerance attributes, being 28 tolerant and 12 susceptibles. The identification data and other basic information of these selected genotypes are listed in the Appendix.

The selected contrasting varieties have been multiplied through *in vitro* micropropagation procedure at CIAT Headquarters. Around 50 individual copies of each contrasting genotype were produced and will be shipped to Embrapa and Cornell University for evaluation.

Tangible outputs delivered

- Drought tolerant contrasting cassava varieties identified
- *In vitro* plants of the drought tolerant contrasting varieties produced

The first six months of the project have been dedicated to solve some administrative constraints regarding the transfer of funds to the participant institutions. Considering that Embrapa (CoPI institution) could not be able to receive all the annual budget and then transfer to the CoPIs institutions, new agreement between CIMMYT and Embrapa were appropriately arranged in order to allow that the budget to the CoPIs (CIAT, IITA and Cornell) can be sent directly to them without sending first to Embrapa. By the end of May the year 1 budget was transferred from GCP to the participants institutions. Shortly, Embrapa and CoPIs will sign subcontracts which has the rules, rights and responsibilities to be undertaken by the participant institutions. The subcontract incorporates many of the rules that were already agreed upon when the institutions became a member of the Generation Consortium.

Another initial constraint to implement the project is the delayed bureaucracy to issue the import permit to ship the cassava genotypes (*in vitro* plants) from CIAT to Brazil and USA. These constraints have caused some delay of the original workplan and can be minimized by no cost extension of the project's activities.

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

Co-Principal Investigators:

Jean-Marcel Ribaut, CIMMYT
Matthew Reynolds, CIMMYT
Scott Chapman, CSIRO, Australia

Collaborating Scientists:

José Crossa, CIMMYT
Mateo Vargas, Universidad Autónoma Chapingo, Mexico.
Sergio Ceretta, INIA, Uruguay
Marco Bink, WUR

MID-YEAR REPORT

As the budget for this project was halved in relation to the original proposal, some extra time was asked for from the GCP committee to investigate how the original plan could as much as possible to be maintained with half of the budget. This implied searching for additional financial support to contract the original two foreseen post docs, one for maize and one for wheat. We have been successful in this, so that the original plan will be retained dropping a few of the original objectives, but keeping the two crops. The post docs will now both work half time on the GCP project, thus together forming a complete post doc. For the other half of their time the post docs will work on projects that are very similar in vein to the GCP project, i.e., developing and evaluating statistical methodology for the analysis of GxE and QTLxE, where also the crops in these complementary projects are the same as for the GCP project.

The first week of May 2005 a kick-off meeting was organized at CIMMYT to make an updated inventory of the available data and the objectives for the whole of the project. Also the program for the first year was detailed. Present were the post docs Marcos Malosetti and Ky Mathews, PI Fred van Eeuwijk, Co-PIs Scott Chapman, Jean Marcel Ribaut and Matthew Reynolds, and collaborators José Crossa and Mateo Vargas. I include the reports and planning for maize and wheat as appendices/ attachments. In principle, the post docs will work full time on the GCP project in the period July-December 2005, where the period January-June 2005 served for preparations for the project. Thus, at this moment no concrete advances can be reported for the methodology as the post docs actually just started to work on the GCP project.

Inventories of phenotypic, physiological and genotypic have been made. Based on this inventory, a planning has been made for the generation/ preparation of required additional data.

Detailed working schedules have been defined for the first year (see appendices/ attachments)

The first half year has been used for planning and making inventories. The second half year will be fully productive in the sense that two post docs will work full time on the project, thereby complying within the period July-December 2005 with the tasks for one post doc for a full year.

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, Universidade Católica de Brasília, Brazil

Serge Braconnier, Centre d'Etude Régional pour l'Amélioration de l'Adaptation à la Sécheresse, Senegal

Jonathan Crouch, CIMMYT

Pietro Piffanelli, CIRAD

Guillermo Seijo, IBONE

Jens Stougaard, University of Aarhus, Denmark

Vincent Vadez, ICRISAT

MID-YEAR REPORT

For the creation of genetic maps we have concentrated on four classes of molecular markers. Microsatellite markers, because they are highly informative, transferable between populations, and useful for plant breeders; single-copy genic anchor markers for genome comparisons with other legumes; resistance gene analogues (RGAs) as candidate markers for disease resistance; and for map saturation we have used a gene-rich sequence characterised AFLP. Building on previous work, computational programs for the efficient production of microsatellite markers, and anchor markers have been further refined. These programs have been made publicly available through the Internet. In total we have now developed 86 anchor markers (25 this year), and 346 microsatellite markers (75 this year) and 15 RGAs (10 this year).

Using an F2 population derived from a cross of *A.duranensis* and *A.stenosperma*, representing the AA genome of *Arachis*, a microsatellite-based gene-rich map has been produced. The map consists of 170 markers in 11 linkage groups covering an estimated 86.4% of the diploid genome. Because most markers used were derived from ESTs and genomic libraries made using methylation-sensitive restriction enzymes, about one-third of the markers are genic. To further build on this map, 70 anchor markers, 15 RGAs and 155 AFLP markers have been genotyped in the mapping population. For AFLP we used Pst I, a methylation sensitive enzyme that effectively enriches the markers for genes, and polymorphic bands were sequence characterized. A second generation map with an initial genome comparison to *Lotus japonicus* is expected soon.

Work for the production of a microsatellite-based map in a population representing the BB genome (*A.ipaensis* and *A.magna*) is in progress. A tetraploid mapping population of 150 F2 plants derived from a cross between *A.hypogaea* and the synthetic amphiploid (*A.ipaensis* x *A.duranensis*)^c has been produced. A tetraploid map is now under construction. Marker segregation in this population is notably free from segregation distortion, this further supports that the *A.ipaensis* and *A.duranensis* genomes are very similar to the AA and BB genomes of *A.hypogaea*.

F2 seed from the cross of the amphiploid (*A.aff.magna* V6389 x *A.aff.diogoi* V9401)^c and peanut is being generated.

Idiograms AA and BB genomes have been made using the localization of ribosomal DNAs, heterochromatic bands, and chromosome morphology, - a basis for physical mapping in *Arachis* is in place. For the generation of large genomic probes the project envisages the construction of bacterial artificial chromosomal (BAC) libraries. Plans for this are on schedule. Seeds of AA and BB genome representatives *A.ipaensis* and *A.duranensis* have been sent to CIRAD and are being grown under greenhouse conditions. One EMBRAPA researcher and one Brazilian PhD. student will visit CIRAD to carry out library construction, validation and clone selection in 2006.

Two synthetic amphiploids have been shown to produce highly fertile hybrids with cultivated peanut, (*A. ipaensis* KG30076 x *A. duranensis* V14167)^c and (*A. aff. magna* V6389 x *A. aff. diogoi* V9401)^c. The amphiploid (*A. ipaensis* x *A. duranensis*)^c, its crosses with peanut, and their progeny show resistance against foliar fungi, most notably against the rust *Puccinia arachidis*. Rust bioassays in the tetraploid mapping population for the identification of quantitative trait linked loci are underway. The amphiploid (*A. aff. magna* x *A. aff. diogoi*)^c also shows resistance against foliar fungi, and bioassays with its F1 hybrid with peanut and their progeny are underway.

Seeds of the amphiploid (*A. ipaensis* x *A. duranensis*)^c have been transferred to CERAAS and to ICRISAT-India.

Seed of wild *Arachis* accessions within the botanical section *Arachis* is being bulked in Brazil and India for drought resistance tests. Physiological tests for drought resistance in seven AA genome and two BB genome wild *Arachis* are underway. New collections of *A. duranensis* at the driest limits of its occurrence in the Chaco region of Argentina and *A. stenosperma* from coastal sand dunes in Brazil have been made. The construction of new amphiploids is ongoing.

Tangible outputs delivered:

1) A genetic map for the AA genome of *Arachis*:

A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae) M. C. Moretzsohn, L. Leoi, K. Proite, P. M. Guimarães, S. C. M. Leal-Bertioli, M. A. Gimenes, W. S. Martins, J. F. M. Valls, D. Grattapaglia & D. J. Bertioli. (2005) *Theoretical and Applied Genetics*. Online first.
<http://dx.doi.org/10.1007/s00122-005-0028-x>

2) New genetic markers for *Arachis*:

25 new anchor markers

75 new microsatellite markers

10 new RGA markers

3) A new mapping population for *Arachis*:

A tetraploid mapping population of 150 F2 plants derived from a cross between *A. hypogaea* and the synthetic amphiploid (*A. ipaensis* x *A. duranensis*)^c

4) New collections of *Arachis* directed towards finding drought resistance in wild *Arachis* from the Argentinean Chaco from Coastal sand dunes.

5) Transfer of the synthetic amphiploid (*A. ipaensis* x *A. duranensis*)^c from Brazil to Africa and Asia.

Setting up contracts, and transfer of money to partners took longer than anticipated. This has delayed work for the evaluation of drought resistance in wild *Arachis*.

Transfer of germplasm from Brazil to Africa and Asia took longer than anticipated delaying the work of partners in these regions. In order to safe-guard the project against future delays in germplasm exchange, modifications of the project have been proposed that envisage tests of wild *Arachis* for drought resistance, and the construction of amphiploids in both Brazil and India.

6. Marker Development and Marker-assisted Selection for Striga Resistance in Cowpea

Principal Investigator:

Festo Massawe, IITA

Co-Principal Investigators:

M.P. Timko, University of Virginia

B.B. Singh, IITA

V. Mahalakshmi, IITA

Ndiaga Cissé, CERAAS, Senegal

N'deye Ndack Diop, CERAAS, Senegal

MID-YEAR REPORT

For the first 6-months of the project efforts have been concentrated on developing molecular markers, mapping populations and breeding lines for *Striga gesnerioides* resistance in cowpea, testing of existing *S. gesnerioides* resistance markers, field evaluations of advanced populations segregating for drought tolerance and capacity building and training. At present two markers have been developed as SCARs and are being used for germplasm evaluation and efficacy testing on populations segregating for *S. gesnerioides* resistance/susceptibility in the field. One marker, designated 61R, is associated with resistance to *S. gesnerioides* races 1 and 3 (SG1 and SG3) mapping to Linkage Group 1 (LG1). The second SCAR is SEACTMCAC83/85 linked to SG3 on LG1. More than 48 different AFLPs have also been identified that are linked to *S. gesnerioides* resistance/susceptibility of cowpea SG1 and SG3 on LG1, and SG1 on LG6. In addition to the use of SCAR and AFLP markers, efforts have also been initiated aimed at generating SSR markers linked to *S. gesnerioides* resistant genes in cowpea. So far one SSR has been found that gave an ~ 500 bp fragment polymorphic between IT84S-2246 (*S. gesnerioides* susceptible line) and TVU14676 (*S. gesnerioides* resistant line). Efforts to develop molecular markers were also extended to the development of markers that could differentiate races of *S. gesnerioides* in the field and a number of primers have been identified, among these is one selective primer combination that produces approximately 120 bp fragment which distinguishes SG1 (Burkino Faso) from SG4 (Benin) and SG5 (Cameroon). Testing of the SCAR markers by IITA is underway and CERAAS will initiate testing in the second half of 2005. Blind MAS based upon the presence or absence of the markers for *S. gesnerioides* resistance genes will be carried out to validate the markers in different populations and MAS protocols will be developed to initiate the first MAS in cowpea.

In order to reduce cowpea yield losses caused by *S. gesnerioides* several breeding efforts designed to transfer, combine and pyramid *S. gesnerioides* resistance into popular and adapted varieties are underway. IITA and CERAAS are also developing advanced mapping populations segregating for *S. gesnerioides* resistance/susceptibility and drought tolerance. At IITA, three F2 populations have recently been advanced (each containing > 150 individuals) to F6 seed. In addition to *S. gesnerioides* resistance, the populations are apparently segregating for other traits that make these populations potential material for development of markers for multiple traits. Four RILs populations for drought tolerant studies are also available, three developed by IITA and one by CERAAS. These will be crucial in developing molecular markers for drought tolerant.

The planned activities have progressed at a good pace with more results expected after the first growing season. One of the critical activities of this project is the West African *Striga* hotspots trial that is being conducted in seven countries to evaluate and screen different cowpea varieties and breeding lines against known races of cowpea *Striga* and possible identification of new races. This trial is currently underway and results will be reported in the first year report.

7. Measuring Linkage Disequilibrium across Three Genomic Regions in Rice

Principal Investigator:

Susan McCouch, Cornell University

Co-Principal Investigators:

Michael Thomson, ICABIOGRAD, Indonesia

Endang Septiningsih, ICABIOGRAD, Indonesia

MID-YEAR REPORT

This is a one-year project with 3 primary activities:

- 1) SNP discovery and marker development across 3 genomic regions associated with resistance to bacterial blight (using 8 diverse Indonesian rice accessions)
- 2) Evaluation of SNP diversity in 96 diverse Indonesian accessions and measuring Linkage Disequilibrium (LD) in the 3 target regions
- 3) Technology transfer and capacity building

During the first 6 months of this project, we have addressed the first and third objectives of this project. A scientist from Bogor (Fatimah Suwardjo) spent 4 months at Cornell working closely with a Cornell research assistant (Nicholas Polato) to complete the work outlined below.

The first activity included the design of PCR primers across the 3 target regions of interest. These included a) a 1 Mb region around the *Xa7* resistance gene on chromosome 6, b) a 1 Mb region around *Xa13* on chromosome 8, and c) a 3 Mb region around the cluster *Xa4/Xa22/Xa26* on chromosome 11.

PCR primers were designed to amplify 700-800 bp DNA fragments across each region, using the publicly available genomic sequence for cv Nipponbare (*japonica*) as template. The PCR products were designed at evenly spaced intervals (averaging one PCR product every 20-30 kb) across each region, with particular attention to genes containing NBS-LRR motifs (candidate disease resistance genes). A majority of primer pairs were anchored in the conserved coding sequences of annotated genes, and oriented to amplify across the more variable introns or 5' or 3' untranslated regions (UTRs) of those genes. Gene models in the target regions were identified using the TIGR and Gramene Databases (<http://www.tigr.org> and <http://www.gramene.org>). PCR primers were designed using Primer3 input (<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi>).

A total of 300 PCR primers were designed, 60 in the *Xa7* region on chromosome 6, 60 around *Xa13* on chromosome 8, and 180 around the cluster *Xa4/Xa22/Xa26* on chromosome 11. The 300 primer pairs were tested for amplification on two accessions (Limar, a *tropical japonica*, classified as Bulu in Indonesia, and Sera, an *indica*, classified as Gundil in Indonesia, and 257 (86%) amplified successfully, and 250 were selected for sequencing. Of

these, 200 were anchored in exons and 50 were designed in non-genic regions. The reliability of amplification was similar between primer pairs anchored in genic and intergenic regions, but amplification was lower in genes containing the NBS-LRR motif.

The 250 primers were then used to amplify products (both forward and reverse sequences were obtained) from a subset of 8 diverse rice accessions collected from different geographical regions of Indonesia. The 8 accessions were: Ingsa Bondol, Ketan Siam, Limar, Rojolele (*tropical japonica*'s) and Sera, Sehan, Popot and Si Anak Bogor (*indica*'s). There are 4,000 amplified products (2,000 from each direction) that have been sent to the BioResource Center at Cornell for sequencing. This sequence will be used to identify SNPs among the 8 accessions, and we will compare the frequency and distribution of SNPs within and between the *tropical japonica* and *indica* sub-populations, as well as between the genic and intergenic amplified fragments.

A preliminary analysis from 56 primer pairs in the *Xa7*-containing region of chromosome 6 suggests that the frequency of SNPs among the 8 accessions is approximately 1 SNP per 150-350 bp. This frequency primarily reflects the divergence between *indica* and *tropical japonica*; most SNPs are detected between these two groups and the frequency within a sub-species is significantly lower. Only about half of the amplified fragments contained any SNPs at all, making it difficult to obtain uniform SNP coverage across the three target regions.

There have been no deviations from the work plan and we are on target for deliverables.

Fatimah Suwardjo, a researcher from ICABIOGRAD, came to Cornell for 4 months to work in the McCouch lab from April 6 to July 31, 2005. During this period, she trained in bioinformatics, PCR and sequencing protocols, as well as in SNP-calling and data analysis techniques. The group at ICABIOGRAD provided the DNA for the project.

Outputs delivered:

- 1) A total of 250 PCR primers have been designed, 60 in the *Xa7* region on chromosome 6, 60 around *Xa13* on chromosome 8, and 130 around the cluster *Xa4/Xa22/Xa26* on chromosome 11. The sequence information for these primers is available.
- 2) Sequences from the 8 diverse Indonesian rice accessions have been generated for 215 primer pairs described above and the 1720 sequences are available.
- 3) Twenty five of the sequences from the 8 rice accessions have been aligned and SNPs have been called for 12 of them. Information about SNP frequency and location is available for these regions.

Once we have sequence data for all 250 amplicons of interest, we will select 25 high frequency SNP targets for the *Xa7* region, 25 for the *Xa13* region and 75 SNP targets in the *Xa4/Xa22/Xa26* region and design SNP assays for each locus. The final density of confirmed SNP markers will be one SNP every 40-50 kb on average, for a total of 125 SNP markers for this project.

Over the next 6 months, high through-put SNP assays will be carried out at ICABIOGRAD based on a primer extension protocol with fluorescent detection and capillary electrophoresis using Beckman CEQ 8000 Genetic Analyzers. A set of 96 diverse rice accessions (selected from the larger collection of 250 Indonesian accessions) will be genotyped with the 125 SNPs.

Over the next 4 weeks while Fatimah is still at Cornell, we will optimize the multiplex SNP arrays (we aim to average 12 SNP markers in each multiplex reaction) to run on the CEQ 8000. The primers to be used for this phase of the project will be designed at Cornell based on the sequence data from the 250 amplicons. Primers will be synthesized and sent to ICABIOGRAD for testing on the CEQ 8000 machines. The resulting data will then be used to define the SNP haplotype structure across each region and determine the rate of LD decay.

After Fatimah returns to Bogor, the high-throughput SNP marker system will be used to genotype the SNP markers across the 96 Indonesian varieties. Once the data is generated, both groups will work together to analyze the data. In addition, upon Fatimah's return to Indonesia, a 2-day training course will be offered to the researchers in Bogor covering SNP discovery and genotyping protocols and offering hands-on training in the necessary bioinformatics and laboratory techniques learned by Fatimah during her visit to Cornell.

The objective of this scientific exchange is to transfer expertise in bioinformatics and laboratory protocols related to SNP discovery, SNP marker development and data analysis to ICABIOGRAD in Bogor, Indonesia as the foundation for future studies in allele mining in rice. It also provides Cornell researchers with an opportunity to learn more about Indonesian germplasm and to gain familiarity with Indonesian researchers.

8. Targeted Discovery of Superior Disease QTL Alleles in the Maize and Rice Genomes

Principal Investigator:

Rebecca Nelson, Cornell University

Co-Principal Investigators:

Casiana Vera Cruz, IRRI

Darshan Brar, IRRI

Hei Leung, IRRI

Margaret Smith, Cornell University

Peter Balint-Kurti, USDA

Jan Leach, Colorado State University

Jane Ininda, KARI, Kenya

Jedidah Danson, KARI, Kenya

James Gethi, KARI, Kenya

Masdiar Bustamam, ICABGRRD, Indonesia

Utut Widiyastuti Suharsono, Bogor Agriculture University, Indonesia

MID-YEAR REPORT

1. Subcontracts have been established with all partners (KARI, NCSU, IRRI and CSU) and work is in progress.
2. A planning meeting was held at KARI-Biotechnology Center and KARI-Kakamega between KARI and Cornell collaborators. A shuttle research program at IRRI was planned for an Indonesian team member. Reciprocal exchange visits were made between NCSU and Cornell.
3. At IRRI, analysis of 145 SSR markers and STS markers co-localizing with known candidate gene sequence has been completed for bulked DNA of 86 F₅ families of Vandana / Moroberekan intermated lines. The 86 F₅ families were advanced to F₆, and we have identified candidate genes in heterozygous condition in five F₆ families. Four hundred eighty F₆ seeds from F₅ derived families identified with candidate genes in

- heterozygous conditions will be grown and each individually analyzed for markers in the specific candidate gene locus to extract near-isogenic lines for candidate gene alleles.
4. At IRRI, 16 entries consisting of wild rice accessions of *O. ridleyi*, *O. longistaminata*, *O. minuta* and 6 accessions of *O. rufipogon*, 3 rice elite lines and 3 M4 lines of IR64 mutants, and IR64 as control were sown in staggered manner and planted in RCB in 4 replications to evaluate for sheath blight resistance in the screenhouse. They were inoculated at maximum tillering stage (45 DAS for rice lines and mutants, and about 50 days old for wild rice accessions) on 16 July. Assessment for sheath blight resistance will be conducted in 5-day interval from July 25 to August 14, 2005.
 5. Introgression lines derived from IR64 x Binam (99 lines) and Teqing x Binam (78 lines) that were identified for drought tolerance from the International Molecular Breeding Program were sown in two batches in RCB in 2 replications in the screenhouse, and will be evaluated for the fourth trial for sheath blight resistance. Inoculation was done on 16 and 23 July 2005 and will be scored at 14 days post inoculation.
 6. At IRRI, 177 entries of introgression lines derived from two mapping populations of superior germplasm identified from the International Molecular Breeding Program were sown in two batches in RCB in 2 replications in the screenhouse, and will be evaluated for sheath blight resistance. Inoculations were done on 16 and 23 July 2005 and will be scored at 14 days post inoculation.
 7. In Indonesia, progress was made in the molecular analysis of the ABQTL population Way Rarem // Oryzica llanos-5. Among 300 SSR markers tested, 52 were selected for use in segregation analysis and 22 were used to in segregation analysis of 186 lines of the BC₂F₃ population. Another 30 polymorphic SSR markers were used in segregation analysis of 30 lines, and we need to analyze 159 more line using those 30 markers. At this moment, 16 of those 30 SSR markers have been use to analyzed 159 lines. We are no waiting for shuttle research program to complete (saturated) the analysis.
 8. In Indonesia, phenotypic evaluations have been done in the greenhouse and field. Several isolates of *P. grisea* from two screening sites were tested. Two of those selected isolates had been used to see the segregation of BC₂F₄ lines. Phenotypic evaluation was recently completed in green house tests with five fungal isolates at the BC₂F₃ generation. Both monocyclic and polycyclic tests were also conducted at IRRI, and field test at IRRI screening sites were also carried out.
 9. Dr. Genia Diaz from UPLB joined the CSU lab in May, 2005. She is screening rice mutants induced by physical (Fast Neutron) and chemical (Diepoxy-butane) means for deletions in regions of chromosomes that house disease QTL. She aims to do a proof-of-concept study on strategies for screening the large library of available deletion mutants produced by IRRI (n=40,000) to detect mutants in regions harbouring disease QTL. Her salary is funded by UPLB, but the materials and supplies are funded through the GCP project. CSU graduate student Myron Bruce has made excellent progress on the use of rice oligonucleotide arrays to localize deleted regions.
 10. An offer has been made by CSU to a post-doc to begin analysis of the utility of oligo-arrays to identify deleted genome regions relevant to disease resistance. At NCSU, graduate student John Zwonitzer was recruited for the GCP project. At Cornell, graduate student Jesse Poland has started work on the project.
 11. MTA between Cornell and KARI was developed and sent out for signatures between the two parties involved. The MTAs have now been finalised and ready for use. This will allow easy exchange of materials.
 12. Maize germplasm panels including maize lines considered to have multiple disease resistance (MDR) were assembled at KARI and Cornell for use in the project from various sources such as KARI, IITA, CIMMYT and GRIN. Over 10 MDR lines are already available and about three are common between Cornell and KARI. At KARI and

- at Cornell, the MDR material was planted for seed increase and evaluation. At KARI, F₂ seed from 4 crosses with resistance to GLS were planted for further inbreeding and crossing to adapted lines at Kiboko.
13. At KARI, tissue from the MDR lines and F₂ plants have been harvested to extract DNA to screen them for GLS loci and identify whether it is the same or different resistance sources for the disease. A shuttle research visit has been planned for KARI scientist to visit Cornell for related marker studies and collaborative planning.
 14. Fifty published studies on QTL analysis of disease resistance in maize were synthesized by the NCSU and Cornell collaborators. A manuscript was accepted for publication on this synthesis. The consensus map is being used to identify chromosomal segments of interest for further analysis.
 15. F₅ seed was provided by NCSU to Cornell for several crosses of mutual interest. CML52 / B73 (n=19 F₅ families) was chosen for analysis at Cornell, as a first test-case for the HIF approach. For each family, 5 plants were grown and selfed, and the DNA for the 95 lines was extracted and analyzed for 29 SSR markers covering the 18 bins of interest. Several of the F₆ families were grown in the field at Aurora NY, assayed for resistance, selfed, and analyzed with SSRs to identify the different marker classes. At least one NIL pairs showing contrast for disease resistance was identified, and more are being extracted.
 16. A near-isogenic pair of lines, with the B73 background and an introgression from Tx303, was shown to differ for GLS resistance in one field trial at NCSU in 2004. The phenotypic contrast was confirmed by NCSU in two additional field trials in 2005, and a third 2005 dataset is awaited. The NILs were back-crossed to the recurrent parent (B73) in order to produce plants with recombination events in the region of the QTL so that it can be mapped more precisely.
 17. Nurseries were planted at NCSU for both grey leaf spot and southern corn leaf blight. An NCLB nursery was planted at Cornell. Several populations are being assessed for the evaluation of disease resistance QTL. In collaboration with Virginia Tech, a subset of the Cornell resistance panel was tested for GLS in Virginia in the 2005 field season. GLS pathogen samples were collected for isolation from this trial.
 18. Near isogenic line pairs are being compared for GLS, SCLB and NCLB. Preliminary results suggest that three pairs of near isogenic lines differing for SLB resistance have been identified. These lines will be assessed next year to confirm the results. For three new sources of GLS resistance, backcrosses to B73 were advanced to BC₂ (for two sources) and BC₁ (for third source) to generate NILs for testing trait-marker associations. At Cornell, NILs for the IBM population were advanced from the BC₄ to the BC_{F5} and BC_{4S1} generations. These lines will be used to confirm trait-marker association and for detailed phenotypic analysis. At NCSU, NILs differing for a GLS QTL were back-crossed to the recurrent parent (B73) in order to produce plants with recombination events in the region of the QTL so that the gene(s) involved can be mapped more precisely.
 19. Eight maize populations previously improved by CIMMYT through four cycles of recurrent selection for NCLB resistance are being used to conduct “recurrent selection (RS) mapping” at Cornell. This method is based on the idea that allele frequency shifts from the initial and final cycles of the RS population that significantly deviate from changes attributable to drift can be identified and interpreted as being associated with selection. We have developed a combined simulation-test statistic method to test the null hypothesis that a given locus is experiencing drift.
 20. Using RS mapping for one of the populations (Pool 30; n=90 individuals divided between the cycles 0 and 4) described above, ~130 mapped SSR loci have been examined, spanning the maize genome at ~20cM intervals and/or concentrated in specific chromosomal segments. We determined allele frequencies in cycles 0 and 4 by

- genotyping a random sample of 45 individuals from both cycles. At present, the null hypothesis of drift could be rejected for 47 of the SSR loci after correcting for multiple tests. Many of these loci correspond to previously published QTL for NCLB.
21. As mentioned above, saturation of selected chromosomal segments has been conducted. Specifically, a region on chromosome 8 has been examined with some 20 SSR loci, nine of which exhibited significant departures from drift. In this region the following resistance loci have been reported: (a) four QTL for NCLB, (b) two major genes for NCLB, (c) three QTL for GLS, (d) two QTL for common rust, (e) one QTL for common smut, and (f) one QTL for maize streak virus. We are now using the maize disease QTL consensus map to select additional SSR loci in the vicinity of previously reported QTL (typically where NCLB QTL co-localize with QTL for several other diseases) for study.
 22. Efforts are underway to determine if allelic differences can be associated with phenotypic differences in disease resistance. A random sample of individuals from intermediate cycles (n=40 for cycle 1; n=20 for cycle 3) from four (including Pool 30) of the eight populations were previously crossed with a common maize inbred line, B73. In the summer 2005 season in upstate NY, F₂ populations were derived from 10 random progeny of each F₁ line (n=2,400 F₂ families). Selected F₂ families will be used to conduct an association analysis of putatively selected alleles versus B73 alleles (similar to a bulk-segregant analysis), and to develop further derivatives of the material (e.g., NILs).

Tangible outputs delivered:

1. Panels of disease resistant maize lines and genetic stocks derived from them.
2. Synthesis of disease QTLs in maize submitted for publication.

We would like to include Southern Corn Leaf Blight as one of the target diseases for the project. An analysis of the literature indicates that this can be justified based on importance in the developing world.

9. Development of Low-Cost Technologies for Pyramiding Useful Genes from Wild Relatives of Cassava into Elite Progenitors

Principal Investigator:

Anthony Bellotti, CIAT

Co-Principal Investigators:

Martin Fregene, CIAT

Alfredo Alves, EMBRAPA-CNPMP

Collaborating Scientists:

Hernan Ceballos, CIAT

Elizabeth Alvarez, CIAT

Elizabeth Okay, CRI, Ghana

Chiedozie Egesi, NRCRI, Nigeria

Anton Bua, NAARI, Uganda

Titus Alicai, NAARI, Uganda

Yona Baguma, NAARI, Uganda

MID-YEAR REPORT

Wild *Manihot* germplasm are a wealth of useful genes for the cultivated species *M. esculenta* but their use in regular breeding programs is restricted due the long reproductive breeding cycle of cassava and linkage drag associated with the use of wild relatives in crop

improvement. This project seeks to identify useful genes for pest and disease resistance, and post-harvest deterioration in cassava and to develop low cost marker tools for their rapid introgression into cassava. During the first six months of the project the following outputs were obtained:

Previous work revealed that a RAPD marker RME1 and an SSR marker NS158 are the closest markers to the gene CMD2 that confers resistance to the cassava mosaic disease (CMD), they are located at distances of 9 and 4 cM respectively, and are being routinely used for marker-assisted selection (MAS) of CMD resistance at CIAT. To reduce the cost and time as well as accuracy of assaying the most important marker, RME1, the polymorphic RAPD fragment in the CMD resistant parent was eluted from an agarose gel, cloned into pGEMT-easy (Promega inc, Madison) and sequenced. Primers were designed from the sequences (Appendix 1) and the RAPD marker successfully converted into a SCAR marker, this marker is now routinely being used for MAS at CIAT and primer sequences has been sent to NARs partners in preparation for its use in MAS in their breeding programs.

Several previous reports have revealed moderate to high levels of resistance to many pests and diseases that attack cassava. Some of these species are being used in this project to introgress the resistance genes into cassava. Additional evaluations of 5 Wild *Manihot* species accessions, F1 Inter-specific hybrids, and BC1, derivatives growing in the field at CIAT were conducted to identify high levels of resistance to green mites, mealybugs, whiteflies, and cassava bacterial blight (CBB). Results reveal excellent sources of resistance to white flies, and moderate sources of resistance to mites and mealybugs (Appendix 2). Preliminary results of the evaluation of CBB resistance in BC2 derivatives of *M. esculenta* sub spp *flabellifolia* revealed moderate to high levels of resistance in some genotypes.

Sexual seeds of natural wild populations of many wild *Manihot* species and their inter-specific hybrids with cassava were distributed to NAS participants for field establishment and evaluation for pest and diseases endemic in their own environment. Seed lots of a total of 1740 sexual seeds from 175 families representing 5 wild *Manihot* species namely: *M. esculenta* sub spp *flabellifolia*, *M. esculenta* sub spp *peruviana*, *M. tristis*, *M. carthaginensis*, and *M. Fomentosa* were each shipped to Brazil, Uganda, Ghana, and Nigeria. Also sent to participating NARs were 1072 sexual seeds of F1 hybrids representing 171 inter-specific families obtained from crossing selections from accessions of the 5 species and elite cassava varieties

Wild relatives of cassava are important sources of genes for resistance to pests and diseases and longer shelf life. The only source of dramatically delayed PPD has been identified in an inter-specific hybrid between cassava and *Manihot walkerae*, a unique source of resistance to the cassava hornworm was also identified in 4th backcross derivatives of *M. glaziovii*. Moderate to high levels of resistance to white flies have been found in inter-specific hybrids of *M. esculenta* sub spp *flabellifolia*. BC1 and S1 mapping populations for the identification of molecular markers for the introgression of delayed PPD, resistance to the cassava hornworm and white were developed last year. They include a cross between CW429-1 (F1 hybrid of *M. walkerae*) and MTAI 8 (BC1), a total of 205 progenies a cross between MNG11 (BC4 derivative of *M. glaziovii*) and MTAI8, 157 individuals, and a cross between CW67-7 (F1 hybrid of *M. esculenta* sub spp *flabellifolia*) and MTAI 8, 230 genotypes. The above-mentioned crosses were established *in vitro* from embryo axes and are currently being multiplied, at least 8 plants per genotype, for transfer to the screen house for hardening and eventually to the field during next year's planting season.

An advanced field-based and molecular marker-assisted selection (MAS) breeding course in cassava was held at CIAT from April 11 to May for NARs partners in the GCP competitive grant project from Uganda, Ghana, Nigeria, and Brazil. The purpose of the course was to expose the NARs cassava breeders to methodologies being used at CIAT for MAS and to update them on current methods in scientific field-based breeding of cassava. Specific objectives of the course were to teach participants the theory and practice of every aspect of cassava breeding and to expose them to new approaches, for example, molecular markers in cassava breeding, doubled haploid technology, tissue culture, and genetic transformation. Molecular marker labs have also been established at CRI, Kumasi, Ghana and NRCRI, Umudike, Nigeria, the lab in NAARI, Namulonge, Uganda is still under construction.

Tangible outputs delivered:

- Development of a low-cost SCAR marker for MAS for breeding resistance to the cassava mosaic disease (CMD)
- Evaluations of several natural populations of 5 *Manihot* species, their F1s and BC1s for resistance to whiteflies and green mites
- Shipment of sexual seeds of several natural populations of 5 *Manihot* species to NARs collaborators in Brazil, Nigeria, Ghana, and Uganda for establishment in the field and eventual evaluations
- *In vitro* establishment of BC1, and S1 gene mapping populations for delayed post-harvest deterioration, resistance to Horn Worm, and whiteflies
- Training of NARs partners from Brazil, Nigeria, Uganda, and Ghana in the theory and practice of field-based and molecular breeding during a one month intensive course at CIAT

A delay in the shipment of BC2 populations with CMD resistance and tolerance to Mites to NARs partners for Molecular breeding, the plants will now be shipped in early October, the delay has been due to the large volume of in vitro culture work involved in establishing the mapping populations for PPD, whiteflies, and hornworm.

Primer3 Output

No mispriming library specified

Using 1-based sequence positions

OLIGO	start	len	tm	gc%	any	3' seq
LEFT PRIMER	115	22	53.07	40.91	2.00	0.00 AGAAGAGGGTAGGAGTTATGTT
RIGHT PRIMER	797	21	53.00	42.86	3.00	2.00 TAGTATGTGCTTGACCCCTATG

SEQUENCE SIZE: 987
INCLUDED REGION SIZE: 987

PRODUCT SIZE: 683, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 0.00

```
1 AAANNACGNNNCCCACNCTTGCATCCACGCGTTGGGAGCTCTCCCATATGGTCGACCTG
61 CAGGCGGCCGCGNAATTCAGTAGNTGNATTGGGAATTCGGGGTGTCCGACTTCTAGAAGA
   >>>>>
121 GGGTAGGAGTTATGTTTACTCTGTGTGATATGTGTTTACTAATTTAACTAGGTTTGA
   >>>>>>>>>>>>>>>>
181 GTTGGAATGACCCAAGGCTGAGGAGCCAAGTTCGGCCGCCGAAGGTAAGTTCGGCCGCCG
241 AAAGTGCTCAATGTTTCGGCTTCCGAATTCAGGTTTCGGCCCCGAACGTTGCATGGTTGTG
301 CATGCGCTTCGGCAGCCGAAGCTGAGTTGCTGAGCCAGCAGTTTTGTATCCCTTAGTCAG
361 CATGTTGGACAGGTGTTATCACCAGATGATGTCCAGAAGATTGGCCATGTCTCCCACT
421 TTATTTGCTGAGTTTGGCAGCTCATGCAGAGTGTTTTTGTAAACGCCGGAAATCCGAC
481 CCGTTACCGGCCTAGGATCCAGATCGGCTTAAGGCTGCCGGGACCCGTAGCAAGCCAAA
541 CGTACTTCTGTAACTGTAATATCCCATACATGATCAACAACATACATGAAACTGTAAA
601 CTTTTCTTTCTTTTCCGAGCTTAACCTGAACATGTGCACATCTAATATCATGACCACC
661 ACTGGAGTCTCATCAAATGCTCCAATGGGGTATTTTCATCATTCGATAAGCTTGGACTAT
721 CAAAATACATCATAAAACATTTCTATAGATCATGTGTCAAGGGGATAACCATAAACATA
   <<<<<
781 GGGTCAAGCACATACTAGTCTCAATCATCTCATCANNACATAACTATACATTGCTCTT
   <<<<<<<<<<<<<<<<<<<<<
841 TCATTACATTAACATAGTCATAAATACTCATTACATCATATTCATGTCCACTACTATCTA
901 TTACATCATAACATGACTTTACNTCCCTCTANCCCAATTNCCCAATCGAATTTCCCNCCGG
961 CCNCCATGGCGGCCNGGAAACATGCNA
```

KEYS (in order of precedence):

```
>>>>> left primer
<<<<<< right primer
```

ADDITIONAL OLIGOS

	start	len	tm	gc%	any	3' seq
1 LEFT PRIMER	115	22	53.07	40.91	2.00	0.00 AGAAGAGGGTAGGAGTTATGTT
RIGHT PRIMER	664	21	53.01	42.86	6.00	2.00 CAGTGGTGGTCATGATATAG

PRODUCT SIZE: 550, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00


```

2 LEFT PRIMER      116  21  51.80  42.86  2.00  0.00  GAAGAGGGTAGGAGTTATGTT
  RIGHT PRIMER     797  21  53.00  42.86  3.00  2.00  TAGTATGTGCTTGACCC2ATG
  PRODUCT SIZE: 682, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 0.00

3 LEFT PRIMER      116  21  51.80  42.86  2.00  0.00  GAAGAGGGTAGGAGTTATGTT
  RIGHT PRIMER     664  21  53.01  42.86  6.00  2.00  CAGTGGTGGTCATGATATTAG
  PRODUCT SIZE: 549, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00

4 LEFT PRIMER      115  22  53.07  40.91  2.00  0.00  AGAAGAGGGTAGGAGTTATGTT
  RIGHT PRIMER     808  21  52.85  42.86  6.00  3.00  TGATTGAGGACTAGTATGTGC
  PRODUCT SIZE: 694, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00

```

Statistics

	con	too	in	in	no	tm	tm	high	high	high			
	sid	many	tar	excl	bad	GC	too	too	any	3'	poly	end	
	ered	Ns	get	reg	GC%	clamp	low	high	compl	compl	X	stab	ok
Left	3964	67	0	0	666	0	32	3057	4	0	0	0	138
Right	3684	102	0	0	2094	0	35	1292	0	0	0	0	161

Pair Stats:

considered 79, unacceptable product size 71, ok 8

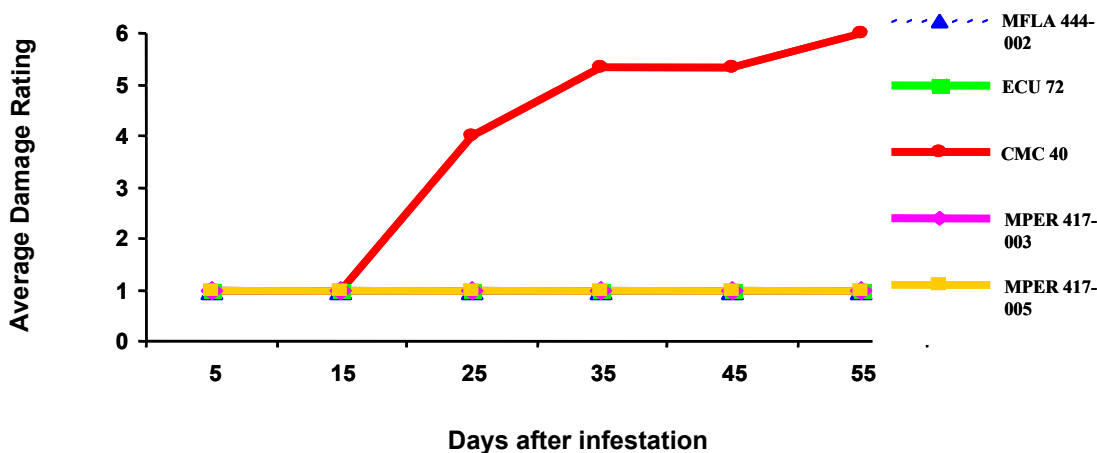


Figure 1. Whitefly (*Aleurotrachelus socialis*) damage ratings on the wild *Manihot* genotypes (MFLA444-002, MPER 417-003) and the *M. esculenta* cultivars CMC40 and MECU72 during a 55 day infestation.

10. Exploring Natural Genetic Variation: Developing genomic resources and introgression lines for four AA genome rice relatives

Principal Investigators:

Joe Tohme, CIAT

Mathias Lorieux, CIAT/IRD

Co-Principal Investigators:

Susan R. McCouch, Cornell University

Claudio Brondani, CNPAF-EMBRAPA

Howard Gridley, WARDA

César P. Martinez, CIAT

Miguel Diago Ramirez, Fedearroz, Colombia

MID-YEAR REPORT

A meeting was held 21-24 February/05 in CNPAF-Embrapa headquarters in Goiania, Brazil. Protocols and work plan for the development of these populations were defined. Criteria for choosing recurrent parents were agreed upon, mainly yield under stress conditions, tolerance

to drought stress under field conditions, tolerance to main diseases, seedling vigor, performance in farmers fields, stay green or late senescence, and improved grain quality.

Representatives from Cornell University, Fedearroz and WARDA could not attend the meeting but were given the opportunity to comment and present their views.

The interspecific cross *O. sativa* x *O. glaberrima*, was done with the varieties IR64 and TOG5681 respectively. Two sub-populations of BC4F2 and BC3F3 lines obtained from IRD were grown at CIAT HQs. The BC3F3 had 233 lines and the BC4F2 had 120 lines. Tissue for DNA isolation were harvested in bulks of 3 to 10 plants by line.

A set of 144 SSR rice primers well distributed along the genome were screened for polymorphism. We selected 76 primers pairs that showed polymorphism between the varieties TOG5681 (*O. glaberrima*) and IR64 (*O. sativa*).

The two sub-populations were evaluated using the 76 primers selected. The PCR products were separated on two kinds of gels. One kind was agarose gels and the other one was denaturing polyacrylamide gels. The agarose ones were used when the polymorphism was greater than 8 bp, and the acrylamide gels were used when the polymorphism was lower than 8 bp. As a result, 50 markers were visualized in agarose gels and 26 were visualized in polyacrylamide gels.

The data were analyzed with the program CSSL Finder (M. Lorieux, unpublished). The program choosed a set of lines that represented the *O. glaberrima* genome as well as possible and gave a graphical genotyping of the lines. From the 353 lines evaluated, 54 were selected to be part of the introgression lines population. The size of the introgression was measured based on the position of the markers in base pairs the rice genome (TIGR release v. 2). The majority of the *O. glaberrima* genome was conserved in the two sub-populations, but it appears that some parts of the genome of *O. glaberrima*, especially on chromosomes 3, 4, 6, 10 and 11 were lost. Anticipating this result, we develop new BC2F1 lines from the same cross that we will check for the lost fragments. The analysis also revealed that for 42 lines from the 54 selected, the bulks must be analyzed individually in order to find the plants that are in homozygous state for the target fragments. The remaining lines ewre already fixed for the target fragments.

In order to facilitate the genetic analysis of quantitative traits and advancing in the improvement of rice, a set of 74 interspecific introgression lines of rice has been developed. A cross between Caiapo and *Oryza glaberrima* (IRGC 103544) was made, using the last one as the male parent. The F1 was backcrossed with Caiapo in the subsequent 2 generations until taking the population to the third backcross (BC₃F₁). From these lines, anthers were collected and through *in vitro* culture of anthers a population of 695 lines BC₃F₁DH was obtained. From this population, 312 lines were selected and genotyped using 150 polymorphic SSRs markers located in the 12 chromosomes at an average distance of 10 cM. In addition, each line was phenotypically evaluated for yield traits and components yield traits. The lines were selected according to the presence of contiguous chromosome segments that covered all the genome of *Oryza glaberrima*, for this purpose the software CSSL Finder specially designed for the search of these lines was used.

Microsatellites markers were selected according to their unique position in the genome (data taken from the Gramene database). Following this criteria, 143 SSRs markers were used for

analysis and search of best lines, the remaining 7 markers were located in different positions and different chromosomes, therefore they were replaced by others. In order to fill the gaps (mainly in chromosomes 2, 4 and 7) and the ends of each chromosome 76 SSRs markers were added, of which 32 have already been evaluated in 312 BC₃DH lines and the remaining 44 are in the standardization phase. All the SSRs has been amplified following the same standard conditions for PCR and gel migration (PAGE/silver staining and agarose gels at 4% stained with ethidium bromide when polymorphism was >10pb).

In each line, one or few different chromosomal segments of *Oryza glaberrima* (IRGC 103544), was substituted in the genetic background of cultivar Caiapo (*Oryza sativa* ssp tropical japonica). The substituted chromosome segments in the 74 CSSLs represent the complete genome of *Oryza glaberrima*, except for small regions of chromosomes 2, 4 and 7. At the end, the software CSSLs Finder will be ran again with all the SSRs data to redefine the best set of lines suitable for developing CSSLs.

Choice of recurrent parents: A scale from 1 to 5 was used during the launching meeting to assess the merits of candidates presented by participants from CIAT and CNPAF. The following cultivars were selected: Lideranza, CNA 8557, Bonanza, Linea 30 and CNA 9025. Some NERICA lines were also suggested by WARDA. Five accessions from each of the following wild rice species will be used as donor parents: *O. glumaepatula*, *O. meridionalis*, *O. rufipogon* and *O. barthii*. CNPAF will be responsible for developing populations with *O. glumaepatula* whilst CIAT, Fedearroz and WARDA will take care of the crosses with the remaining wild species. A flow chart for the development of these populations was also established and shared with participants.

Seed of recurrent parents was exchanged among participants following quarantine regulations from each country. Dr. Susan McCouch for Cornell University provided data and seeds of about 100 wild rice accessions to facilitate the identification of the best accessions for the development of the populations.

CIAT and CNPAF got started in their crossing program. At CIAT, we are currently collecting F1 seeds from single crosses between recurrent parents and *O. barthii*. F1 seeds will be planted shortly to produce BC1F1 by the end of this year.

Also, during the period of time from July 11th to July 19th, pictures were taken for forty-seven wild types of the gender *Oryza* at CIAT HQs. These types are distributed in eight types of *Oryza nivara*, seven types of *Oryza rufipogon*, five types of *Oryza barthii*, twenty-four types of *Oryza meridionalis* and five types of *Oryza glumaepatula*. This activity was due to describe some phenotypic characteristics (qualitative) of the wild parental that will be used in the interspecific crosses, in order to keep a data record, and compare them during the production of the F1 crosses and the backcrosses.

Nasu et. al (2002) found 2800 SNPs in 417 regions distributed throughout the genome of three *Oryza sativa* ssp. *japonica*, two *indica* cultivars and a wild rice (*O. rufipogon*). We decided to evaluated the polymorphism rate obtained with a subset of these sequences between parent species that will be used in this project.

Sixteen SNPs were chosen (at least one per chromosome) for standardization of single base extension methodology detection. SNP detection was carried out using the single base extension (SBE) methodology described by Chen et al. (2000). Briefly, PCR products containing the SNP were obtained from amplification of genomic DNA. Excess dNTPs and

primers of the PCR product were removed with shrimp alkaline phosphatase and exonuclease I. Then the amplified fragment was annealed with an SBE primer with a 20-22 oligonucleotide (ZIP code or tag) attached to its 5' end. The single base extension was carried out using ddNTPs (one labeled with biotin) and thermosequenase so that the 3' end of the SBE primer anneals to the base that immediately precedes the SNP. Streptavidin-phycoerythrin was conjugated to the biotin labeled oligonucleotide, which was hybridized to polystyrene micro spheres (5.6 μ M diameter) bounded to a ssDNA sequence complementary to the ZIP code. The reaction is read in a flow cytometer (Luminex 100), which detects each micro sphere by its unique fluorescent signal and the presence or absence of the SNP (streptavidin-SBE product). Data were analyzed with Masterplex GT (Miraibio Inc.) package in which the mean fluorescence intensity emitted by each of the samples is analyzed and used to define the SNP alleles belonging to each genotype.

Sixteen rice genotypes including *O. sativa* (*indica* and *japonica* cultivars) and related species *O. glaberrima*, *O. barthii*, *O. longistaminata* and *O. rufipogon* were genotyped. Single base extension products were read for 7 out of the 16 SNPs. Although results are preliminary due to the small sample size, we observed that SNPs S0155 and S0168 were useful to identify polymorphism between the *japonica* and *indica* accessions, and between the *O. barthii* and *O. glaberrima* genotypes; S0323 distinguished between *indica* and *japonica*, *O. barthii* and *O. glaberrima*

There is experimental evidence for 22,057 of the 3,931,108 SNPs detected by means of computational comparisons between the *indica* (cv. 93-11) and *japonica* (cv. Nipponbare) genomes (NCBI, dbSNPs July 2005, http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi). Given the abundance of these markers and their potential for specific gene tagging, we aimed at automating the mining and annotation of the SNP variation in genes of agronomical importance prior to the validation of these markers in actual crosses. The careful annotation of SNPs will permit us to make high resolution associative mapping between the variation at single nucleotide positions in genes belonging to key regulatory networks and the relevant phenotypes they presumably control.

In this order of ideas, we've developed an automated tool, SNPspipe v0.8, written in PERL language that can be applied to the general problem of finding *indica/japonica* SNPs in genomic sequences spanning known full length cDNAs provided by the user. In addition to the initial SNP detection and annotation of the chromosomal coordinate, this application is intended to assign coding/non-coding nature of the polymorphism and specifically determine its location in an exon, intron, or 5' regulatory sequence (promoter region). Finally, SNPspipe can automatically generate PCR primers for all the polymorphisms found, along with the files required for high throughput Single Base Extension primer design.

All the information produced would be stored both as sequences with the fasta format (according to the specifications of the NCBI dbSNPs), as well as organized in a relational database implemented in MySQL. The SNPspipe program, still in developing phase, has been tested in 5 genic families involved in iron homeostasis finding 410 SNPs in a total of 39 genes.

Due to its general design SNPspipe would be used in the automated discovery and annotation in 169 rice gene candidates involved in responses to drought-stress.

To help at the genotyping of the all six populations, we designed and started to develop an *universal rice core genetic map*. Although this activity was not initially planned in the

framework of this project, we decided to include it in because it will help grandly the comparison of results between populations.

More details about the principle of this universal map will be presented in the GCP Annual Meeting.

A program, CSSL Finder, was developed to help at choosing the best candidate lines as candidate for CSSLs according to their genome content.

Three B Sc students were trained at CIAT HQs. Two Phd D and two Master Sc students have been identified (one in each partner). Two of them have already started (Fedearroz – Embrapa), and two will join us on early September (CIAT – WARDA).

11. Functional Genomics of Cross-species Resistance to Fungal Diseases in Rice and Wheat (CEREALIMMUNITY)

Principal Investigator:

Pietro Piffanelli, AGROPOLIS

Co-Principal Investigators:

J-L Notteghem, AGROPOLIS

M. E. Ferreira, EMBRAPA

R. Singh, CIMMYT

P. Ronald, University of California-Davis

M. William, CIMMYT

J-B Morel, AGROPOLIS

D. Tharreau, AGROPOLIS

S. Kikuchi, NIAS

F. Dedryver, University of Rennes, France

L. Boyd, JIC

S. Brammer, EMBRAPA

AS. Prabhu, EMBRAPA

M.C. Chaves, EMBRAPA

E. Guiderdoni, AGROPOLIS

MID-YEAR REPORT

AGROPOLIS experimental work in the first six months focused on the cytological characterization of the rice-*Puccinia triticina* interaction. Nine rice genotypes were tested with the *Puccinia triticina* strain BOOM and analysed at both macroscopic and microscopic levels. The cytological analysis was carried out using a DAB + aniline blue staining technology which enabled to visualize both fungal structures (appressorium, infection hyphae) and production of ROIs (reactive oxygen intermediates) in plant cells. The work was carried out by a graduate student – Miss Emilie Callizo – under the supervision of Dr. P. Piffanelli and Dr JL Verdeil (head of the Imaging Platform) at CIRAD Montpellier. The results of this work were published in a University Report.

Coordination work focused on the organization of the First Year Cerealimmunity Meeting and planning of the experimental work to be carried out at Agropolis, Embrapa an JIC.

In addition to the work that was planned to be carried out at JIC we have undertaken a cytological study of the interaction between wheat and barley isolates of stripe rust (*Puccinia*

striiformis) and six of the nine rice genotypes tested at Agropolis (P. Piffanelli). Four wheat and four barley stripe rust isolates were screened. Wheat controls included cultivars Renan, Recital, Lemhi and Chinese 166. Samples were taken at 24 and 48 hours after inoculation and at 15 days. Examination of the 24 and 48 hour samples has been completed, but the 15 day samples still remain to be scored. To complement the work at Agropolis three rice genotypes have been inoculated with two wheat leaf rust (*P. triticina*) isolates and one barley leaf rust isolate, and these are currently being examined. This work has been carried out by Ruth MacCormack, the Research Assistant of L. Boyd.

Artificial inoculations were made, under controlled greenhouse conditions, to identify non-host and host specific isolates of *M. grisea* of rice and wheat. Fourteen wheat cultivars, one rice cultivar and one barley cultivar (Table 1) were utilized for inoculation tests. Inoculations were made with aqueous spore suspension (3×10^5 conidia per ml) on 21-day old plants, using fourteen isolates of *M. grisea* collected from wheat, two from grasses (*Digitaria horizontalis* and *Eleusine indica*) one each from barley and rice. The isolates retrieved from wheat represent wide genetic diversity and geographical distribution in Brazil (Table 2). Analysis of variance showed significant differences among cultivars and isolates (Table 3). Even though the isolate x cultivar interaction was significant the percentage participation in explaining total variation was small (7%). All test isolates were virulent to wheat cultivars, including the rice, barley and grass isolates, but showed differences in aggressiveness (Figure 1). The grass isolate from *D. horizontalis* was least aggressive on wheat cultivars. None of the isolates were virulent to the rice cultivar Bonança excepting the rice isolate (race IB-9). The most aggressive isolate Py 5996 on wheat cultivars was selected for further studies on cytological characterization of non-host interactions.

These 14 wheat cultivars will be inoculated with 40 races of *Puccinia recondita* at Passo Fundo to select two virulent races and three wheat cultivars. Also, it is proposed to test initially 8 rice cultivars including IR64 with *P. recondita* races. The work is underway.

Table 1 Disease severity on wheat cultivars in inoculation tests with 18 isolates of *Maganaporthe grisea*.

Cultivars	Mean of severity(%)
Ágata	95,41
Aliança	89,37
Brilhante	76,59
BR 18	56,11
Embrapa 22	85,34
BR 33	86,73
Embrapa 42	90,13
BRS 207	68,33
BRS 208	89,86
BRS 210	81,52
BRS 234	68,75
BRS 254	81,87
BRS 264	87,98
Pioneiro	88,33
Barley	95,48
Rice (Bonança)	0,00

Table 2 Isolates utilized in the inoculation tests, their origin, collection site and year of collection

Isolates	Host/Cultivar	Location/State	Year
Py 182	Wheat /Anahuac	Mato Grosso do Sul	1988
Py 5994	Wheat /-	Rio Verde/Goiás	2002
Py 7596	Wheat / PF 89375	Montividiu / Goiás	2004
Py 7608	Wheat /BR 17	Costa Rica / Mato Grosso do Sul	2004
Py 183	Wheat /Anahuac	Mato Grosso do Sul	1988
Py 5996	Wheat /-	Rio Verde/Goiás	2002
Py 7601	Wheat /BR 17	Costa Rica / Mato Grosso do Sul	2004
Py 201	Wheat / Anahuac	Mato Grosso do Sul	1995
Py 7612	Wheat / BH 1146	Alto Taquari / Mato Grosso	2004
Py 7618	Wheat / BH 1146	Alto Taquari / Mato Grosso	2004
Py 204	Wheat / Anahuac	Mato Grosso do Sul	1988
Py 7599	Wheat / PF 89375	Montividiu / Goiás	2004
Py 7606	Wheat / BR 17	Costa Rica / Mato Grosso do Sul	2004
Py 7600	Wheat / BR 17	Minas Gerais	2004
Py 212	Grass / <i>Digitaria horizontalis</i>	Goiás	1989
Py 195	Grass / <i>Eluesine indica</i>	Goiás	1989
Py 5990	Barley /-	Rio Verde/Goiás	-
Py 3970	Rice / Bonança	Goiás	2002

Figure 1. Differences in aggressiveness of isolates *M. grisea* virulent on wheat cultivars in inoculation tests (Goiania, 2005) (See author for figure)

AGROPOLIS inoculation tests of rice plants with *Puccinia striiformis* were not successful due to high temperature conditions in the growth room (>20 degrees C) – a new growth room will be tested in August 2005.

The microscopic analysis of *P. striiformis* and *P. triticina* development on rice at JIC is in addition to the proposed work plan and has been undertaken to complement and support the work carried out at Agropolis.

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, CAAS

Co-Principal Investigators:

Gary Atlin, IRRI

Jian-Min Wan, CAAS

MID-YEAR REPORT

Progeny testing of selected drought tolerant introgression lines (ILs) under stress and non-stress conditions and development of intercross populations for QTL pyramiding
A total of 630 ILs from 74 BC populations, including 630 C418 ILs and 193 Liaojing 454 ILs were progeny tested under both stress (no irrigation) and normal irrigated conditions in Beijing and Liaoning. Preliminary observations based on visual scoring allowed us to identify 26 promising ILs which performed significantly better than the recurrent parents under both mild stress (no irrigation) and normal conditions. A total of 46 crosses have been made between these promising ILs.

DNA has been extracted from 582 DT ILs from 23 BC populations and the 25 parental lines. A total of 600 anchor rice simple sequence repeat (SSR) markers have been used to screen the polymorphisms among the parental lines (350 SSR markers have been completed). Genotyping of these ILs has been started will be completed by before March 2006. The F₁ plants of the 46 crosses were planted into the field and each was examined with 5 polymorphic SSR markers to eliminated false F₁ plants. F₂ populations will be obtained from these F₁ plants and screened under severe stress during the dryseason of 2005-2006 to select DT lines with pyramided DT QTLs from 2 different donors.

A total of 21 crosses were made between the 7 selected DT IR64 lines each with multiple QTLs from two donors. The F₁ plants of the 21 crosses were planted into the field and each was examined with at least 5 polymorphic SSR markers to eliminated false F₁ plants. F₂ populations will be obtained from these F₁ plants and screened under very severe stress during the dryseason of 2005-2006 to select DT lines with pyramided DT QTLs from 4 different donors and for development of DT and high yielding lines for the rainfed areas of South/Southeast Asia.

A total of 30 crosses were made between a high yielding cultivar, Swarna, and DT IR64 ILs selected from 16 populations (donors) and 4 populations were advanced to BC₁ generation.

An elite C418 ILs with good drought tolerance and restoring ability was identified, which was testcrossed to a CMS line. The F₁ hybrid showed a high level of heterosis in grain yield in two locations under mild stress and normal irrigated conditions (picture 1). This hybrid will be recommended to be tested in multilocal yield trials under stress and non-stress conditions of North China next year.

Deviations from the work plan:
None.

13. Development of Informative DNA Markers through Association Mapping in Maize to Improve Drought Tolerance in Cereals

Principal Investigator:

Jean-Marcel Ribaut, CIMMYT

Co-Principal Investigators:

Edward Buckler, Cornell University

Alain Charcosset, INRA

James Gethi, KARI

Grudloyma Pichet, NSFCRC, Thailand

Luke Mehlo, SIRDC

Mark Sawkins, CIMMYT

Tim Setter, Cornell University

Wanchen Li, Sichuan Agriculture University

Collaborating Scientists:

Marianne Bänziger, CIMMYT

Javier Betran, Texas A&M

Jose Crossa, CIMMYT

Luz George, CIMMYT

Philippe Monneveux, CIMMYT

MID-YEAR REPORT

Out of the 600 entries evaluated last 2004 summer cycle, mainly for adaptation to our experimental station in Mexico, 460 genotypes were selected and classified in 3 precocity groups. Those 460 entries were planted in 5 meter rows trial under optimal conditions. Half of the plants were used to make a seed increase of the different genotypes and the other half was used to produce hybrid seeds by crossing plants with CML312. CML312 is one of the best CIMMYT tester presenting a broad adaptability in Africa and Asia. On this trial, leaf samples were harvested for DNA extraction by bulking tissue from 10 plants.

In parallel to the seed increase, three randomized trials (for the three precocity groups respectively) with 2 replications per trial, were conducted under water limited conditions. As stated in our proposal leaf samples, silks and ear tips were harvested at mild and severe stress stages for metabolite analysis. A total of about 6000 samples were harvested in 80% methanol. Other morphological data such as precocity, plant size, chlorophyll content and root conductivity were collected for all entries.

Because this issue of precocity it a major one, a set of 20 entries from each group, both lines and corresponding hybrids, has been sent to Co-Pis in China, Kenya, Thailand and Zimbabwe. Each group will evaluate under well-watered conditions the entire set of 60 genotypes and 60 hybrids with just one row per entry to look at flowering time and adaptation under your local conditions. The output of this small experiment should help a lot to prepare the final list of genotypes that will be evaluated under stress and well-watered conditions in 2006 and 2007. A set of 430 hybrids, 30 entries did not have enough seeds, was sent to James Gethi in Kenya to conduct a full trial under both stress and well-watered conditions. The trials have been planted early July and phenotypic data will be available for our meeting at the end of the year, early November in Kenya. This trial is a pilot study in

preparation to the large phenotyping effort that will be conducted next year in five different countries.

Samples of leaves, silk and ear tip have been stored at -20 during 4 weeks to let the tissues exodiffuse in the medium. Samples have been processed and dry aliquots of the methanol have been sent (96 well plates) in a stepwise manner to Cornell for analysis. Leaf samples were shipped about a couple of months ago, silk a couple of weeks ago, and ear tips will be sent those days. All tissues have been weighted for Tim Setter to adjust the volume of extraction medium based on the amount of tissue harvested. An aliquot of the dry tissue will also be sent soon to allow the cellulose quantification.

Assays of glucose and sucrose, as well as ABA and ABA-glucose ester have been completed on leaf samples and silk samples are now under analysis. Tim is finalizing the large-scale protocol for both the phaseic acid and dehydrin.

DNA has been extracted from the 460 genotypes and is ready to be used to determine the structure of the population using SSR neutral markers as well as the gene sequencing.

A first list of candidate genes has been circulated and discussion to select most suitable is ongoing considering: 1) the relevance of the gene pathway, 2) potential regulatory role of the candidate genes on the metabolites quantified in this study, and 3) what has already been done by Ed Buckler and Alain Charcosset in their respective association projects conducted on temperate germplasm. A final first set of candidate genes to be sequenced this year should be identified soon.

So far we are pretty much on track looking at the timing projected in the proposal. We are perhaps a little bit late in identifying the candidate genes and making the contact with Genassiance, but this should not affect the outputs on this year. However, due to some change in CIMMYT staff involved in the project, some delay in some activities might be expected.

14. Characterisation of Genetic Diversity of Maize Populations: Documenting global maize migration from the centre of origin

Principal Investigator:

Marilyn Warburton, CIMMYT

Collaborating Scientists:

Luz George, CIMMYT

S. Taba, CIMMYT

V. Mahalakshmi, IITA

Abebe Mentir, IITA

Alain Charcosset, INRA

Zachary Muthamia, KARI

S.H. Zhang, CAAS

B. M. Prasanna, Indian Agriculture Research Institute

Sutrisno, Indonesian Department of Agriculture

Pichet Grudloyma, Nakhon Sawon Field Crops Research Centre

Phan Xuan Hao, National Maize Research Institute, Vietnam

Artemio Salazar, Philippine Department of Agriculture

MID-YEAR REPORT

The organizational meeting for the partners in this project was held on April 4 – 6, 2005 in Nairobi, Kenya. The meeting was attended by at least one person from every institution, excluding INRA, who sent regrets and a PowerPoint presentation on association analysis. At the meeting, the following actions were completed:

1. Protocols for SSR amplification and electrophoresis and data analysis were distributed
2. Contracts were sent to all partners and clarifications made
3. The list of accessions to be genotyped was discussed (and has since been finalized). Numbers chosen per country/region and institute responsible for DNA extraction: IITA – 60 populations from western and central Africa; KARI – 60 populations from eastern and southern Africa; CIMMYT – 15 populations of teosinte, 15 populations of controls (populations run in previous studies); India – 25 populations from India; China – 25 populations from China; Indonesia – 22 populations from Indonesia; Vietnam – 22 populations from Vietnam; Thailand – 22 populations from Thailand; Philippines – 22 populations from the Philippines (all Asian countries were asked to include populations from non-represented countries, if they have landraces in their collections. 288 accessions were chosen as they will fill three 96 well plates. African accessions to be chosen also possibly from the list from Dr. Taba of African accessions in the CIMMYT genebank (already distributed) and possibly from the Genetic Resources Information Network, GRIN, of the United States. They may have older sources of seeds. KARI was also recently given seeds of a collection survey of maize landraces done in Kenya by CIMMYT (led by Hugo deGroot). This list to be checked as well).
4. Because all money budgeted for the first meeting was not spent, each NARs was invited to submit a \$1500 capacity building grants. Two countries (China, Thailand) are sending people to CIMMYT for training; one (Kenya) will buy computers. Other countries have not yet submitted grants.
5. Work plans were created by each institution including detailed activities and timelines.

In addition to the meeting, a virtual work space has been created at <http://www.generationcp.org/vw/modules.php?name=News&file=article&sid=39> where the workshop schedule and participants (and photos!) have been posted, as well as protocols agreed upon by partner institutions. Finally, a CD of relevant materials was distributed to workshop participants, which included germplasm data collection sheets, reporting procedures and expectations for this project, laboratory protocols, data analyses programs, database programs, and the full proposal for this project. The contents of the CD have also been posted to the virtual web space.

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

The second workshop for this project has been scheduled for December 5 – 9 in Beijing, China. At this workshop, participants will update progress, report problems, exchange DNA if it has not yet been done, and learn the data analysis programs that will be used to analyze the bulked maize populations for this project. Discussions and presentations on Association Genetics will also be given by collaborators working in that area.

Tangible outputs delivered:

- CD of workshop materials (described above)
- List of germplasm to be genotyped in this project

- Virtual web site working space
- Capacity building grants of \$1500 for each of the 7 NARs partners in the project.

Deviations from the work plan:

None to date; all proceeding as planned.

15. Determination of a Common Genetic Basis for Tissue Growth Rate Under Water-limited Conditions across Plant Organs and Genomes

Principal Investigator:

Mark Sawkins, CIMMYT

Co-Principal Investigators:

Jean-Marcel Ribaut, CIMMYT

Francois Tardieu, INRA

Peter Stamp, ETH, Switzerland

Matthew Reynolds, CIMMYT

Peter Langridge, ACPFG, Australia

Renee Lafitte, IRRI

Ravindra Kumar, IGAU, India

Luke Mehlo, SIRDC, Zimbabwe

Collaborating Scientists:

John Bennett, IRRI

Marianne Bänziger, CIMMYT

Claude Welcker, INRA

Yvan Fracheboud, ETH, Switzerland

Richard Trethowan, CIMMYT

MID-YEAR REPORT

- QTL detection of maize leaf growth to water deficit (P1xP2 segregating population) (INRA).

Elongation rate of the sixth leaf of 120 RILs and environmental conditions was recorded in greenhouse and growth chamber over a period of drying. QTLs were detected for maximum growth under non stress conditions and response to soil water potential. These QTL colocalized with previous QTL for ASI under stress in the field, suggesting a common genetic mechanism for growth of silks and leaves.

- Characterization of genetic variability for leaf elongation, leaf development under water stress, yields and related traits in maize in field conditions (CIMMYT - Mexico)

A population of RILs (P1xP2 - 220 genotypes) was planted December 2004. Leaf 5 and 10 were marked to permit leaf counting over the cycle. At the 10-12 leaf stage water was withheld. In February 2005, the number of leaves (ligulated and non-ligulated) were recorded once a week until all leaves had emerged. After 4 weeks, irrigation was re-applied before flowering and then applied on a regular basis. During grain filling, leaf size (length and width) were measured for 6 leaves of 5 plants per plot. In addition flowering traits, plant morphology, chlorophyll content and senescence as well as yield components were also recorded. A parallel trial, to provide material for large scale profiling, of two parental lines and six contrasting genotypes (three good and three bad) from the same segregating population was also planted and evaluated under well watered (2 replicates) and stress (three

replicates) conditions. The whorl was dissected from 30 plants and 5cms of the base of the youngest leaf that had just emerged taken and pooled. This was done for stressed and well-watered treatments and leaf emergence also recorded weekly (from leaf 10). The number of leaves (ligulated and non-ligulated) were counted as well as chlorophyll content. Predawn leaf water potential was measured on parental lines (20 leaf samples each) under both conditions.

- Characterization of genetic variability for ear and silk development, yields and related traits in maize under drought field conditions (CIMMYT - Zimbabwe)

This will be conducted this summer in Zimbabwe.

- QTL maize root growth under stress (ETH)

This activity was split into two phases for evaluating root growth: i) Evaluation of desiccation tolerance in growth pouches and ii) Evaluation of desiccation avoidance in growth columns.

i) Pouch system experiments were conducted. Plants were grown in growth pouches on filter paper for 10 days until the one-leaf (V1) stage. Desiccation stress was applied using polyethylene glycol in two concentrations (15 and 25 % w/w). Root development under stress and non-stress conditions was measured by means of digital image analysis. Target trait was the desiccation tolerance indicated by the response of root growth to desiccation stress. This forms a diploma project that will end in September 2005. The results will be published in form of the diploma thesis of Samuel Trachsel.

ii) Experiments in the sand columns were first outlined as a poster at the 25th Anniversary of cooperation between Kasetsart University (KU) and Swiss Federal Institute of Technology (ETH), International Conference on Maize Adaptation to Marginal Environments. Plants were grown in 80 cm growth columns until the five-leaf (V5) stage. Water content was initially adjusted to 25 or 100% of the maximal water holding capacity. Target trait was the vertical root-length distribution as indicator for differences in desiccation avoidance due to different rooting depths among the genotypes. This research is being undertaken by a PhD student Nathinee Pa-In.

The results of both experiments will be available soon. In a second phase, QTL mapping for relevant root traits will be conducted in one or both systems, depending on the trait where the parents differ most.

- Characterisation of leaf emergence and elongation rate in rice (IRRI)

Leaf emergence and elongation rate is currently being evaluated using 150 BC lines (Vandana x Moroberekan), under well-watered conditions in the field, and NILs (IR64 x Azucena), under well-watered and stressed conditions in the greenhouse. (Los Banos, rainy season). These two populations also be evaluated in the field during the dry season (January).

- Characterization of leaf growth of a set of wheat lines in field under stress and well-watered conditions in the field (CIMMYT - Obregon) and controlled conditions (CIMMYT – El Batan).

One hundred lines were evaluated in the field under stress conditions and trait data collected. A subset was selected based on 2004 commissioned research and data for leaf and stem extension rate under stress and irrigated conditions collected. A subset of six pot grown contrasting lines is currently being evaluated for leaf extension rate at known soil water potentials in controlled conditions.

- Adapt the silk growth Model developed at INRA to tropical maize (INRA)

Tropical maize was grown and subjected to different environmental conditions in a growth chamber. Silk elongation rate was continuously monitored (non-destructive). Growth responded in a linear way to temperature. Growth changes also paralleled ear leaf water potential. Results suggest that silk growth rate is partly controlled by plant water status and could be driven by turgor in the growing zone.

- Characterization and modeling the variability of rice leaf growth to environmental conditions (INRA/IRRI)

This activity begins in September 2005 when a new PhD student has joined the team.

- Large scale expression profiling on contrasting lines using leaf growth zone RNA under stress and optimal conditions (CIMMYT, IRRI, ACPFG).

In maize, material has been collected for use in profiling experiments. These activities span years 1-2. In rice, material will be collected in the second half of the year (see rice phenotyping above).

Results presented at international conferences (International Conference on Maize Adaptation to Marginal Environments) and the upcoming Interdrought II conference, Rome. Capacity building in terms of offering training to students at diploma and PhD level. Presentation of IRRI results at Rice Genetics 5 (Manila, November 2005).

One activity where more serious attention needs to be placed is the development of the database and collation of QTL data into a consensus map. Considerable efforts have been made over the six months to standardize the maize data from CIMMYT and this is now been brought together into a consensus map. Now, other QTL data from other sources (maize and other species) needs to be systematically brought together.

16. Isolation and Characterization of Aluminum Tolerance Genes in the Cereals: An Integrated Functional Genomic, Molecular Genetic and Physiological Analysis

Principal Investigator:

Leon Kochian, Cornell University

Co-Principal Investigators:

Ed Buckler, Cornell University

Owen Hoekenga, Cornell University

Jurandir Magalhaes, EMBRAPA

Claudia Guimarães, EMBRAPA

Vera Alves, EMBRAPA

Newton Carneiro, EMBRAPA

Robert Schaffert, EMBRAPA

Sandra Brammer, EMBRAPA
Pericles Neves, EMBRAPA
Rosangela Bevitori, EMBRAPA
Samuel Gudu, Moi University, Kenya

MID-YEAR REPORT

High-resolution mapping of *Alt_{SB}* was undertaken with two markers very closely linked to the Al tolerance (*Alt_{SB}*) locus with a mapping population 2200 F2 individuals derived from the cross of an Al sensitive by Al tolerant parental line, allowing us to identify 27 single recombinant individuals. Progeny testing of selected recombinants for Al tolerance and BAC end mapping identified a single BAC that harbors *Alt_{SB}*. By the time the proposal was submitted to the GCP, this BAC was being prepared for sequencing. The BAC was subsequently sequenced and annotated. Sub-BAC high-resolution mapping was done by identifying sequenced-tagged site (STS) markers within open-reading frames (ORFs) contained within the BAC. Genotyping of the STS markers on the selected recombinants allowed for the identification of two markers that flanked *Alt_{SB}* and defined a 27 KB interval in which only three candidate ORFs lay. One of these ORFs encodes for a transporter-like protein that has been implicated in organic acid efflux, and thus is a strong candidate for *Alt_{SB}*. The other two ORFs appear to encode housekeeping genes that shouldn't have a role in Al tolerance.

The expression profile of the ORF that is the best candidate for *Alt_{SB}* was obtained by quantitative real time PCR analysis using mRNA from root tips of near-isogenic lines differing at the *Alt_{SB}* locus. The candidate ORF was strongly expressed in an Al-inducible manner only in the root tips of the Al tolerant isogenic line and not the sensitive NIL; the two additional ORFs were not expressed in roots at all, which strongly support the hypothesis that this orf is the major sorghum Al tolerance gene. We are now conducting genetic association tests to define polymorphisms related to the Al tolerance phenotype. This is being done at this moment with a panel of 47 sorghum lines that have been phenotyped for Al tolerance. If necessary, formal association analysis will be conducted in the sorghum diversity panel that was made available to Embrapa Maize and Sorghum by CIRAD (France). We are also at this moment transforming both Al sensitive maize and sorghum with our candidate for *Alt_{SB}* as well as with *ALMT-1*, the wheat Al tolerance gene cloned by Dr. Matsumoto's lab (Okayama University). Similar to *Alt_{SB}* in sorghum, *ALMT-1* also encodes for a transporter-like protein. However, *ALMT-1* provides tolerance to wheat via Al-induced malate release whereas *Alt_{SB}* mediates root citrate release in sorghum. Having *ALMT-1* and *Alt_{SB}* in the same genetic background should give insights into the relative degree of tolerance provided by each gene. We are pursuing Al tolerance QTL mapping in two RIL mapping populations, one Brazilian (AI237 x L53) and one North American (Intermated B73 x Mo17, or IBM). In the Brazilian population, we are currently landing markers found to be closely linked to either *Alt_{SB}* or *Alt_{BH}* (the wheat Al tolerance locus) on the Al tolerance QTL map. In addition, we are closing gaps in the Al tolerance QTL map published by Ninamango-Cárdenas et al. (Euphytica 130:223-232, 2003) using SSRs. This map should provide for a solid framework for sequence-based verifications of orthology hypotheses that can be formulated based on the synteny between grass chromosomes. For this purpose, both *ALMT-1* and the *Alt_{SB}* candidates are being anchored to the maize map. Aluminum-induced citrate release is also being scored in the maize mapping population to identify QTLs associated with this trait. Al tolerant recombinant inbred lines that exhibit lower rates of citrate release, if identified, will be used for isogenic line generation. As our current knowledge suggests that Al tolerance mechanisms other than organic acid release take place in maize, these isogenic lines will be instrumental to the elucidation of these additional mechanisms. Similarly, QTLs with major

phenotypic effects are being selected and will be used for isogenic line generation by marker-based backcross breeding.

In the IBM population, five QTL were identified using composite interval mapping analysis on a 1,000+-marker genetic map. Mo17 donated three loci Mo17 and B73 two. Backcross lines were constructed in order to verify the QTL models and initial NIL construction. At the BC3 generation, the QTL model was confirmed; additionally, the three loci donated by Mo17 were found to act epistatically. In the summer 2005 research field, BC3S2 families were generated, from which we will derive tolerant and sensitive NIL pairs. The NIL pairs will be the objects of physiological, genetic and genomic profiling in 2006. In addition to working with RIL populations, we have also utilized diversity-based approaches. First, we have used microarray analysis to characterize Al responsive genes in a number of maize genotypes differing in Al tolerance and have identified several candidate Al tolerance genes. One these genes may be involved in cell wall composition and thus may describe a novel tolerance mechanism, while a second is involved in organic acid metabolism. These candidate genes were subjected to association analysis using 288-variety association panel assembled by Dr. Buckler and this analysis identified significant associations between Al tolerance levels and particular SNPs in these genes. Thus, they are currently being characterized further. Finally, microarray analysis identified a gene strongly induced by Al only in the tolerant lines that is a close of our candidate for *Alt_{SB}* in sorghum. This gene is being analyzed for its role in maize Al tolerance.

Tangible outputs delivered:

- Elite Al tolerant sorghum hybrids developed from the breeding program
- Identification of an Al tolerance candidate gene in sorghum

There have been no major deviations from the work plan. However, the setting up of subcontracts between Cornell University and the three Embrapa labs and Moi University in Kenya is taking more time than initially expected. Dr. Kochian has informed Jennifer Nelson of the GCP about this and was told this is also happening with some other of the CGP projects, particularly for subcontracts between host CG institutions and other institutions where no pre-existing agreement has already been in place. Dr. Kochian has informed Jennifer that the delay may require the requesting of a no-cost extension at the end of the project. Nonetheless, excellent progress has been made in the first 7 months of the grant.

17. Allele Mining Based on Non-Coding Regulatory SNPs in Barley Germplasm

Principal Investigator:

Michael Baum, ICARDA

Collaborating Scientists:

W. Powell, University of Adelaide, Australia

P. Langridge, Australian Centre for Plant Functional Genomics Pty Ltd

Mark Tester, Australian Centre for Plant Functional Genomics Pty Ltd

J. K. Eglinton, University of Adelaide, Australia

M. Morgante, Universita' di Udine Via delle Scienze, Italy

Salvatore Ceccarelli

Stefania Grando, ICARDA

Sripada Udupa, ICARDA

Wafaa Choumane, Tishreen University, Syria.

MID-YEAR REPORT

Administrative issues: with the move of Dr. W. Powell from Australia to NIAB, UK we added in agreement with GCP management NIAB as a collaborating institute. An appropriate MoU was prepared and signed.

The *cis*-regulatory regions have been hypothesized to facilitate adaptive innovations because slight nucleotide changes may generate novel phenotypes while preserving existing functions of genes. Promoters and other *cis*-acting regions form a protein/DNA complex with trans-regulatory proteins, thereby promoting integrative control of expression. These regions consist of short and often redundant transcription factor binding sites. In some cases, a point mutation may effect binding property of trans-acting factors, results in altered expression/regulation. In order to detect such regulatory polymorphism in barley the following activities were performed during Jan-June, 2005:

Twelve barley lines (including its wild relative *H. vulgare ssp. spontaneum*) was used to make forward and reverse hybrids. The crossing schemes are presented below. The numbers indicated in the table are the number of seeds available at this moment. Parental lines were AFLP fingerprinted to assure purity.

	Tadmor	Alexis	H.spont. 41-1	CPI71284- 48	Arta	Sloop	WI3408.
	98	99	100	106	101	102	103
Tadmor	98	300	350	106	350	350	400
Alexis	99	105	74	X	220	230	138
H. spontaneum 41-1	100	427	325	X	372	540	220
CPI71284-48	106	X	X	X	X	X	X
Arta	101	349	165	110	X	39	66
Sloop	102	400	250	130	X	31	465
WI3408.	103	200	500	49	X	120	450

Forward cross
Reverse cross
Under preparation

		Barque-73	CPI71284-48
		107	106
Barque-73	107		X
CPI71284-48	106	X	
		Morex	Cebada
		95	Capa
			96
Morex	95		X
Cebada Capa	96	120	
Haruna Nijo	97	112	95
			Haruna
			Nijo
			97
			58
			200

This part of the work was conducted in co-operation with Michele Morgante, Udine University, Italy. Prof. Morgante's laboratory has already standardized the technique to detect allelic imbalance expression in maize. Two scientists from ICARDA and one from Tishreen University were trained in his laboratory in this technique. In order to apply this technique to barley, we had selected dehydrin 12 gene of barley (located on 6H) as an example. Dehydrins (DHNs; LEA D11) are water-soluble lipid-associating proteins that accumulate in response to dehydration, low temperature, osmotic stress or ABA, or during seed maturation, and are thought to play a role in freezing and drought tolerance in plants. Upon sequencing of *Dhn 12* genomic regions of Arta and Tadmor, 3 SNPs were detected in coding regions of the gene. One was located in exon1 (Figure 1) and other two were on exon 2. Two pairs of primers were designed to amplify the two exons. These primers were used to amplify the exons using genomic DNA (the barley hybrids Arta x Tadmor and Tadmor x Arta) and also cDNAs prepared using mRNA extracted from shoots and roots of barley hybrids (Arta x Tadmor and Tadmor x Arta) after giving 30 min dehydration stress. The genomic DNA templates were prepared for estimating a standard curve by mixing Arta and Tadmor at 25:75, 50:50 and 75:25 ratios. Two forward and reverse single base extension (SBE) primers (~20-25 bases long) were designed for individual SNPs, in such a way that the primer annealing ends just one base before the SNP site, so that these primers can be extended with fluorescence labelled dd-nucleotides. The amplified genomic DNAs and cDNAs corresponding to the two exons were used as template for SBE (Fig. 2). The SBE was performed using Snap Shot multiplex kit supplied by ABI. The SBE products were further treated with SAP (Shrimp Alkaline Phosphatase) to remove unincorporated dNTPs. The purified products were denatured and electrophoresed on ABI 3730 DNA analyzer. The results were analysed using GeneMapper 3.5 software. The peak height ratios of Arta and Tadmor alleles estimated with SBE products derived from Arta and Tadmor DNA mixes (25:75; 50:50 and 75:25) were plotted and standard curves were estimated. The reliability of standard curve was tested by comparing the peak height ratios obtained from the genomic DNA of the two alleles in the two hybrids.

The ratio of peak heights of two alleles of the SBE products derived from cDNAs of shoots and roots of the forward and reciprocal hybrids were compared. The result shows that there was a difference in expression levels of two alleles in hybrids. This difference in allele specific expression was not same between forward and reverse hybrid, indicating presence genomic imprinting at *Dhn12* locus in barley.

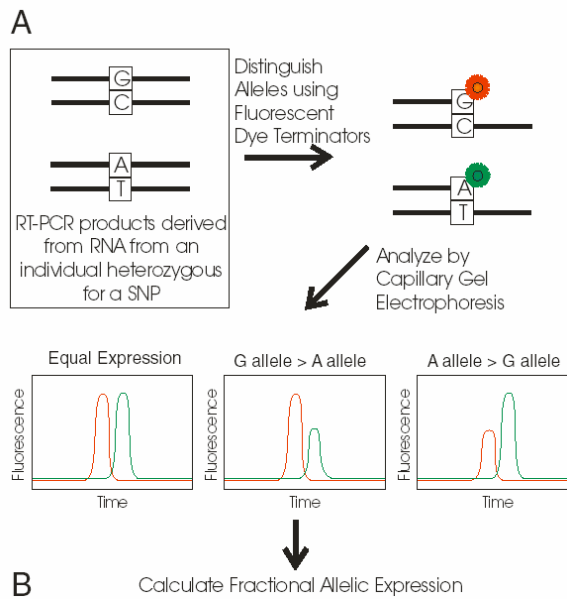


Fig. 2. Overview of detection of cis-acting allelic expression variation.

Tangible outputs delivered:

1. Forward and reverse crosses between barley lines were constructed
2. Scientists from ICARDA were trained in the technology and technology transferred to ICARDA Syria.
3. Cloning and sequencing of candidate drought genes in parental lines (dehydrin 12) were initiated.

SP1 COMMISSIONED GRANTS

1a. Completing Genotyping of Composite Germplasm Set of Barley

Principal Investigator:

Michael Baum, ICARDA

Collaborating Scientists:

J. Valkoun, ICARDA

S. Grando, ICARDA

Zhang Jing, CAAS

Joanne Russell, SCRI, Scotland

MID-YEAR REPORT

The seeds of 2692 accessions, representing a barley collection collected from 84 countries, were used for the project activity. Seeds were grown in pots in a greenhouse. Leaves were cut from seedlings of 3-4 weeks and collected in Liquid N₂, then lyophilized in a freeze-dryer for DNA extraction. The whole plants were kept to harvest the seeds at the end of the season.

CTAB protocol was used for DNA extraction from individual plants. Quantity and quality of DNA was evaluated. The DNA samples were diluted to the concentration of 25ng/ul. Up to

now, 2550 DNA samples were extracted and were prepared for analysis. 1000 DNA samples were sent to CAAS in China, for genotyping with 35 barley SSRs.

At ICARDA, 15 barley SSR primer pairs (Bmac 18, Bmac 316, Bmag 382, Bmag 211, Hvltppb, HvhvaI, SCSSR 5939, SCSSR 10148, SCSSR 07970, SCSSR 02748, SCSSR 15864, SCSSR 25691, SCSSR 02306, SCSSR 08447, SCSSR 03907), localized on 6 different chromosomes, were selected for genotyping. 1200 DNA samples were amplified with 15 primer pairs, separated on ABI sequencer and analyzed with Genotyper program. An example of the results obtained after the analysis of Syrian Jordan accessions with 40 SSR primers is presented in Table 1 where the number of alleles detected per locus was calculated for each country. 900 Samples were amplified and they are ready for separation and analysis. The remaining DNA samples will be amplified and prepared for running. The missing DNA samples (142 accessions are now grown in the greenhouse) will be extracted soon.

The preliminary results showed variation for the analyzed loci. Some country specific alleles were detected. The genetic diversity will be determined once the analysis with 50 SSR primers has been completed.

Table 1: Number of alleles detected in a Syrian and Jordanian Barley collection.

Loci/ Primers	Total No. of alleles	No. of alleles in Jordan	No. of alleles in Syria
Bmac 18	5	2	5
Bmac 134	18	10	18
Bmac 211	11	4	11
Bmac 316	12	6	12
Bmac 399	2	2	2
Bmag 9	8	6	8
Bmag125	9	6	9
Bmag 353	14	11	13
Bmag 382	6	4	6
Bmag 419	4	2	4
Ebmac415	2	2	2
Ebmac 701	10	5	10
Hvcma	3	2	3
HVHVAI	3	2	3
Hvltppb	8 alleles / 2 loci	4	8
Hvm 40	10	7	10
Hvm62	9	4	9
Hwaxy-4	4	3	3
Scind 5281	5	1	5
Scind 6002	12	5	12
Scind 16991	5	2	4
Scssr 15864	9	6	9
Scssr 25691	3	3	3
Scssr 5599	8	5	7
Scssr 581	8	2	8
Scssr 02306	5	5	4
Scssr 05939	4	3	4
Scssr 07106	5	4	5
Scssr 07970	18	10	16
Scssr 08477	6	3	5
Scssr 10148	4	3	4
Scssr 10477	13	3	12
Scssr 10559	5	3	5
Scssr 14079	4	4	4
Scssr 18005	3	1	3
Scssr 20569	4	3	4
Scssr 7759	6	3	6
Scssr 9398	32	16	32

Tangible outputs delivered:

1. DNA from 2692 accessions, representing a barley collection collected from 84 countries, has been extracted.
2. DNA from 1200 DNA was PCR amplified at ICARDA with 15 SSR primer pairs
3. DNA from 1000 accessions were send to CAAS China.
4. Seeds from 2550 accessions were harvested.

Deviations from the work plan:
Until now ICARDA has not received meaningful genotyping data from CAAS.

1b. Completing Genotyping of Composite Germplasm Set of Wheat

Principal Investigator:

Marilyn Warburton, CIMMYT

Participating Institutions:

CIMMYT

ICARDA

MID-YEAR REPORT

At CIMMYT, 400 Mexican landraces, 2300 Iranian landraces and 400 new synthetics were grown under drought conditions in trials in Ciudad Obregon, Sonora, Mexico. They were grown during the dry season of 2004 – 2005, but unfortunately, this was a year of record rainfall during this time. No useful data was gathered from this drought trial, as essentially every entry was well watered by excess rainfall.

The 200 most drought resistant landraces that had been analyzed in drought trials in previous years were chosen for the genotyping at CIMMYT instead of the ones mentioned in the paragraph above. 250 lines were identified by Nachit Miloudi of ICARDA for genotyping as well. The DNA from the 450 drought resistant wheat populations is currently being extracted and DNA will be exchanged for genotyping as soon as it is ready.

Tangible outputs delivered:

List of 450 drought resistant wheat accessions.

Deviations from the work plan:

Because of unexpectedly high rainfall in the winter of 2004-2005, no drought characterization data were obtained on the 3100 landraces that were planted for characterization. Furthermore, the collaborator in charge of this project has since left CIMMYT. If the GCP wishes to pay for this activity again, and if the new head of the CIMMYT wheat gene bank wishes to run the drought trials again, we could try to get the data again, although it would be very late. If not, this data will not be available.

1c. Completing Genotyping of Composite Germplasm Set of Sorghum

Principal Investigator:

C T Hash, ICRISAT

Collaborating Scientists:

Claire Billot, Agropolis

Monica Deu, Agropolis

Jean-François Rami, Agropolis

Jacques Chantreau, Agropolis

P Ramu, ICRISAT

HD Upadhyaya, ICRISAT

RT Folkertsma, ICRISAT

Yu Li, CAAS

MID-YEAR REPORT

At ICRISAT-Patancheru, seed samples of single selfed plants of 2300 sorghum accessions (the balance of the 3000 entries of the GCP composite germplasm set for sorghum over and above the nearly 700 entries that were genotyped during 2004) were harvested from a 2004/05 postrainy season crop, divided into portions to be provided to each of the collaborating institutions (ICRISAT, Cirad and CAAS), and a small portion of the seeds of each lot sown to produce seedlings from which DNA was isolated. DNA samples were prepared in 96-well format, with 95 experimental samples and one BTx623 control entry on each plate. DNA quality of each sample was assessed on agarose gels, and the samples then split four ways (one each for ICRISAT, Cirad and CAAS, and one backup). After drying, sets of 25 plates each will be ready for dispatch to Cirad and CAAS by the end of July.

Refinement of data generated in 2004 at ICRISAT-Patancheru for the initial 48 entries x 104 SSR primer pairs, and full 670 entries x 11 SSR primer pairs was completed, and the resulting cleaned up SSR genotype data sets passed to Cirad for comparison with data sets from Cirad and CAAS (48-entry data set) required for choice of the remaining 20 marker loci that are to be used in genotyping the complete 3000 entry sorghum composite collection, and merger (670-entry data set) for initial diversity analysis based on this 2004-generated SSR data set.

Seed samples of 210 entries contributed by Cirad to the 670-entry subset genotyped during 2004 were received at ICRISAT-Patancheru in 2005. Export of samples contributed by CAAS to Cirad and ICRISAT is still pending, as the necessary export permits have not been issued yet. Seed samples of entries contributed by ICRISAT were harvested and are now available for distribution to Cirad and CAAS.

Tangible outputs delivered:

Sorghum seed samples from Cirad received by ICRISAT; sorghum seed samples contributed by ICRISAT multiplied and prepared for dispatch to Cirad and CAAS.

Deviations from the work plan:

Freshly harvested single selfed plant seed samples were not available from the for DNA isolation until late in May, resulting in a 3-month delay in DNA sample preparation and distribution. Sample isolation was completed in early July and distribution should be completed by July end. A no-cost extension of a month or two may be required to complete the SSR marker genotyping and analysis for these 2300 samples as a result of this delay, but we will first attempt to make up the lost time by improving our use of multi-loading of dye-labeled PCR products using information gained during the 2004 marker data generation for this project.

The proposed June meeting (to have been held in conjunction with the SP1 genotyping meeting in Chennai) was delayed as a consequence of the postponement of the Chennai meeting. In place of this meeting, Dr Claire Billot visited Dr Yu Li in Beijing for a week in June, and provided some useful pointers to the CAAS team that should improve both the rate of marker data generation and the quality of this data.

1d. Completing Genotyping of Composite Set of Chickpea

Principal Investigator:

HD Upadhyaya, ICRISAT

Collaborating Scientists:

PM Gaur, ICRISAT
CLL Gowda, ICRISAT
S Chandra, ICRISAT
SM Udupa, ICARDA
BJ Furman, ICARDA
M Baum, ICARDA

MID-YEAR REPORT

- The global chickpea composite set consists of 3000 accessions from 58 countries, ICRISAT, ICARDA and some germplasm of unknown origin.
- Genotyping of 286 accessions (211 ICRISAT mini core and 75 accessions from ICARDA) was completed in 2004 using 50 SSR markers at ICRISAT (35 SSR markers) and ICARDA (15 SSR markers).
- For the remaining 2714 accessions, seed was assembled at ICRISAT and representative plant of each accession grown.
- We have completed extraction of high quality DNA, using a high-throughput DNA extraction protocol and sent 5 µl DNA of all the 2714 accessions to ICARDA.
- At ICRISAT we have started generating marker data using high throughput ABI Prism 3700 DNA sequencer and have generated marker data on 1956 accessions using 35 SSRs.
- At ICARDA 1694 samples have been PCR amplified for 6 SSR loci and electrophoresed in ABI 3100 DNA analyzer.

Tangible outputs delivered:

- High quality DNA of the 2714 accessions extracted and shared with ICARDA.

Deviations from the work plan: Nil.

***Note: 1e-1n are projects from 2004**

1e. Genotyping of composite germplasm set, Tier 1, maize

Principal Investigator:

Marilyn Warburton, CIMMYT

Collaborating Scientists:

Alain Charcosset, INRA

Mahalakshmi, Visvanathan, IITA

Yu Li, CAAS

MID-YEAR REPORT

The project was divided into two parts: maize inbred line fingerprinting and maize population fingerprinting. Labs fingerprinting the inbred lines included CIMMYT and CAAS, and labs fingerprinting the populations included INRA and CIMMYT. All DNA had previously been extracted and shipped to the cooperating labs, so between January and June of 2005, all labs completed fingerprinting of the DNA. However, all labs experienced some problems which did not allow any of us to finish all markers on all germplasm, see details below. All data is currently being transformed into the data templates for eventual submission to the GCP Central Repository. CIMMYT will try to complete an initial analysis of the data before the meeting in Chennai, India, on August 22, and data will be ready and available for the

workshop. Thus, when the workshop provides the guidance on how the data will be further analyzed, we can proceed immediately.

Dr. Yu Li of CAAS submitted SSR data for 7 SSRs on 1215 inbred maize lines and 150 populations. However, some problems were noted by M. Warburton: 1). no data were submitted for the control genotypes (CML051 and CML292), for which each lab was provided with DNA. 2). Every line had only ONE allele for every SSR; as it has been noted in previous SSR characterization studies that generally maize inbred lines retain approximately 2 – 15% heterozygosity, this was seen as potentially problematic. 3). The germplasm list sent by Dr. Li did not match the Composite Genotype Set agreed upon for fingerprinting (there should have only been 987 inbreds). 4). The populations were fingerprinted with the protocols for the inbred lines, so may not be able to be used in the population study. Dr. Li is working to fix the first two problems as quickly as possible, and is working with M. Warburton to fix the third problem. We will determine what can be done about the fourth problem; in any event, these 150 populations were done in addition to the lines that Dr. Li had committed to do, so he has done more than the work necessary for his part.

Dr. M. Warburton of CIMMYT submitted SSR data for 43 SSR markers on 987 inbred lines. Some problems were noted with this data set as well: 1). There are some SSR markers missing for many of the inbred lines, bringing the total missing data to 21%. Although we will continue to work to fill in the missing data, four SSR markers would not amplify despite repeated attempts with new primers. If we remove these 4 SSRs from the analysis, the missing data drops to just under 12%, and within a month we expect to have re-run enough of the remaining SSRs with missing data to bring the total below 5% missing data, which we believe to be acceptable. Dr. M. Warburton also submitted SSR data for 15 primers run on 467 populations. Missing data is also a problem with this data that we are working to correct. CIMMYT was originally committed to run 24 SSR markers, but optimization of markers for population analysis took much longer than expected because the postdoc hired for this project started in 2005 and not in 2004 as expected and it took 9 months to install the ABI3100 bought by CIMMYT, due to problems with the company in Mexico. We do not expect to be able to run all 24 markers on the populations, but we will continue to run as many markers as we can until the deadline of the project.

Dr. Alain Charcosset of INRA submitted SSR data for 19 SSRs for 100 populations. INRA was originally committed to run 24 SSRs, but no reliable amplification was obtained with 5 of the SSRs. Dr. Charcosset does not intend to run more markers on the populations because the funding provided by the GCP for this project has all been spent. He also reports that data for an additional 217 populations has been run but not completely analyzed, so will be submitted very soon. He does not expect to run any SSRs on the final 150 populations, which were added after he had agreed to do the fingerprinting and therefore he has no budget for that. No problems with this data were noted, except that the controls were not run with these populations.

A data set consisting of 46 SSR markers run on 987 inbred lines is available for analysis. A subset of the data consisting of 7 SSRs run on 228 inbred lines is also available. A data set containing 34 SSR markers run on 317 populations will very soon be ready. An additional data set containing 15 SSR markers run on 150 populations (and, possibly 22 SSR markers if we can reconcile the analysis from CAAS with the analysis from CIMMYT) is available.

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

As noted, only 34 SSRs total were run on the maize populations, and possibly only 46 SSRs run on the maize inbred lines, instead of the 50 originally planned. We believe that this number of SSRs is sufficient to complete the analysis on polymorphism, diversity, and structure of the germplasm we have analyzed; therefore, no negative consequences should be noted.

1f. Genotyping of composite germplasm set, Tier 1, wheat

Principal Investigator:

Marilyn Warburton, CIMMYT

Collaborating Scientists:

Xueyong, CAAS

François BALFOURIER, INRA, Agropolis

Nachit Miloudi, ICARDA

MID-YEAR REPORT

All DNA had previously been extracted and shipped to the cooperating labs, so between January and June of 2005, all labs completed fingerprinting of the DNA. However, all labs experienced some problems which did not allow any of us to finish all markers on all germplasm, see details below. All data is currently being transformed into the data templates for eventual submission to the GCP Central Repository. CIMMYT will try to complete an initial analysis of the data before the meeting in Chennai, India, on August 22, and data will be ready and available for the workshop. Thus, when the workshop provides the guidance on how the data will be further analyzed, we can proceed immediately.

Dr. Xueyong submitted data for 12 SSRs run on 2379 populations, lines, and wild species of wheat. However, a few problems were noted: 1). The data was sent in a format incompatible with the templates for the GCP. 2). No control data was included. 3). M. Warburton noted some potential problems with the SSR data, for example, some markers showed only one allele for every individual (highly improbable) and others showed two alleles for every individual (equally improbable). 4). There was not an exact match between the germplasm submitted by Dr. Xueyong and the wheat Composite Genotype Set (although it was very close; some lines are in addition to the composite set and approximately 200 accessions may be missing) and some of the lines were apparently duplicated for some, but not all, SSR markers. For the ones that were duplicated, the same allele data was not always seen. M. Warburton has asked for Dr. Xueyong's help in solving these problems.

Dr. Balfourier of Agropolis submitted SSR data for 8 SSRs run on 2356 populations, lines, and wild species of wheat. There was a misnumbering of accessions that occurred in the DNA from AGROPOLIS after the composite genotype set was defined, but we think we have corrected the list and Dr. Balfourier feels that the DNA sent from his lab can be renamed according to the numbers and his new list. However, the data from Agropolis also showed the problem that only one allele was recorded for every accession for every SSR. Finally, 145

wheat accessions are missing from this data set compared to the composite genotype set. M. Warburton will be in contact with Dr. Balfourier to look again at the data.

Dr. Marilyn Warburton has submitted SSR data for 26 SSRs on 806 populations, lines, and wild species of wheat. 12 SSR markers have been completed on an additional 1695 accessions of wheat. The final 14 SSRs should be finished on the entire set of 2501 accessions by the August 22 meeting in Chennai, but the preliminary analysis prepared for that meeting may not contain all the SSR data yet.

Dr. Nachit Miloudi has not yet submitted data for the 3 SSR primers his lab is in charge of. In December, 2004, he reported that they had finished running 3 SSR markers on 1176 accessions of wheat, but the data was not sent to M. Warburton as of the writing of this report.

A data set containing at least 2300 wheat populations, lines, and wild species (the exact number depending on how the four labs reconcile their lists of germplasm) analyzed with at least 47 (and 50, if ICARDA submits soon) SSR markers will be available very soon. A data set containing the 2300 and 33 SSR markers is available now.

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

No major deviations are expected, although it is possible that the entire data set of 2300 – 2500 accessions and 50 SSRs may not be ready before August, 2005.

1g. Genotyping of composite germplasm set, Tier 1, rice

Principal Investigator:

Kenneth McNally, IRRI

Collaborating Scientists:

N. Ruraidh Sackville Hamilton, IRRI

Claire Billot, CIRAD

Brigitte Courtois, CIRAD

Laure Benoit, CIRAD

Cesar Martinez, CIAT

Matthias Lorieux, CIAT

Claudio Brondani, EMBRAPA

Long-zhi Han, CAAS

Marie-Noelle Ndjiondjop, WARDA

Susan McCouch, CORNELL

MID-YEAR REPORT

During the June 2004 workshop, the decision was made to partition the set of 3000 accessions into gene pools consisting of 2150 traditional *O. sativa* (selected from 7 geographical regions: E Asia, SE Asia, S Asia, W Asia, Europe, Oceania, Africa & America), 500 improved *O. sativa* (selected from varieties, breeders' materials, genetic stocks and other reference materials), 300 African cultivated rice *O. glaberrima*, and 50 wild inbreeding AA genome accessions (from *O. barthii*, *O. glumaepatula*, *O. meridionalis*, *O.*

nivara, *O. rufipogon*). Since 1536 accessions had already been chosen from the IRGC, the 1464 remaining accessions were chosen so that the target number was achieved wherever possible. These accessions were selected by the MaxiMin approach of N. Ruaraidh Sackville Hamilton using Genstat during early October 2004. Accessions were chosen from the nominated materials outside of IRRI as follows: 231 from CAAS (31 improved and 200 traditional), 106 accessions from CIAT (97 improved, 7 traditional, and 2 *glaberrima*), 61 accessions from CIRAD (47 improved, 10 traditional, and 4 *glaberrima*), and 98 accessions from WARDA (93 improved and 5 *glaberrima*). Some of the selected nominations were available at IRRI (e.g. those selected from CIRAD's nominations), so the IRRI source was used.

Seed from the 106 CIAT entries were obtained in 2005, and these DNAs have been prepared. No materials have been obtained from WARDA or CAAS. The recent situation of WARDA has made it difficult to provide the materials. Recent communication has indicated their willingness to send the lines. In the case of those lines from CAAS, even though DNAs from these have been prepared at CAAS, permission from the Ministry of Agriculture has not yet been given for their shipment to IRRI and distribution among the partners.

The initial set of 1536 entries was shipped in two lots during 2004 to CIRAD, CAAS, CIAT and EMBRAPA. The first was in May/June for 48 samples, and the second of 1488 samples was shipped to CIRAD, CIAT, and CAAS in October and to EMBRAPA in December (due to importation requirements). The 1536 samples were shipped to Cornell (for WARDA) in March 2005 (timed to precede the arrival of a WARDA technician at Cornell for assisting in the genotyping).

Of the remaining 1464 samples needed, DNAs are available for 1232 entries. Some of these samples replace part of those that were to have been obtained from outside of IRRI. The majority of those missing are nominations from CAAS. Plates of lyophilized samples have been prepared and are awaiting clearance for shipment (by August 10, 2005). The final shipment was delayed on the hope that all nominated entries could have been included.

The SSR panel was partitioned across the partners with 24 to IRRI, 7 each to CIRAD, CIAT, and CAAS, and 6 to WARDA. CIAT is genotyping with EMBRAPA, and WARDA is genotyping at Cornell.

Genotyping of 1536 samples has been accomplished for 3 SSRs and is almost complete for 9 additional SSRs. Data for these will be available for the Chennai workshop. IRRI's genotyping platform is the MJ Research BaseStation BST51 (48 lanes) and BST100 (96 lanes). This genotyping platform has severely hampered our genotyping commitment: MJR was acquired Bio-Rad in 2004 without the inclusion of the BaseStation which MJR had already decided they would no longer sell. Hence, Bio-Rad has not been able to facilitate resolution of the following problems. Sometime in the last half of 2004, MJR decided to no longer provide technical support. This timing was especially problematic since the BST100 was in need of repair. Initially, the laser power unit was replaced by the local technician in September 2004. Furthermore, attempts to procure consumables from MJR from March 2004 until March 2005 were unsuccessful with only a partial shipment of supplies obtained in January 2005. In April 2005, we were informed by the local supplier that MJR would no longer be selling consumables, and then we were directed to an alternative supplier (previously unknown to us or the local supplier). This supplier has been able to provide orders on a timely basis. In January 2005, the BST100 was found to have focus failures and collected images did not contain any signals in the four channels. The lack of technical

support by MJR precluded resolving this problem until the local supplier made contact with a prior MJR technician (in May 2005) who had left MJR the previous fall. A special arrangement was made for him to visit IRRI at the beginning of July 2005. This visit led to the conclusion that the photomultiplier tube assembly and possibly the amplifier assembly would need replacement.

Hence, we are operating at a third of what was our expected capacity. This was further exacerbated by insufficient supplies and competition with 3 other projects for time on the BST51. We are considering options other than the use of the BST. The easiest is to shift SSR genotyping to our LiCors that have been devoted to ecotilling. This will require the purchase of the SAGA software for SSRs.

Genotyping of the 1536 samples for 7 SSRs has been accomplished including positive controls using LiCor genotypers. Unfortunately, during the course of updating the Saga software for the analysis of the Licor genotyping gels on a new server, half of the analyses that had been performed on rice were recently found to have been lost. LiCor has attributed this to incompatibilities between the Oracle database and the new version of Saga. Consequently, these analyses will need to be repeated over the next few weeks.

The 7 SSR markers were split between CIAT (4) and EMBRAPA (3). CIAT has accomplished genotyping on 1536 samples for 2 markers. Genotyping on the other 2 markers has been accomplished for 5 of the 16 plates and is ongoing for the remaining plate. Although labeled primers from IRRI stocks were shipped in May and October of 2004 with the DNA shipments, these primers failed to work at CIAT (although they work at IRRI). Subsequently, CIAT ordered their own primers delaying the start of their genotyping. Samples on plate 5 failed to give any amplification. This plate will be sent to CIAT with the remaining shipment. EMBRAPA has finished the genotyping on 1536 accessions with 3 SSRs using directly labeled primers on an ABI genotyper, and this data is available for analysis. EMBRAPA managed to complete this work using funds from other sources; their GCP funds for 2004 were just recently received at the beginning of July 2005.

The 1536 samples that were shipped have been genotyped on an ABI 3730 for 6 of the 7 SSRs. One primer pair failed to amplify. This primer will be resynthesized. CAAS is reliant on the technical support of their centralized genotyping facility for the analysis. Since this assistance has not been forthcoming, the data remain to be analyzed.

Genotyping at Cornell is being accomplished by Mr. Mamado Cissoko, a research assistant from WARDA, with the assistance of Adam Famoso, a Cornell graduate student. Mr. Cissoko arrived at Cornell on July 23 for 8 weeks (his arrival was delayed due to a visa problem). Thus far, 5 plates of DNA have been successfully amplified with all 6 markers, confirming the optimization of the PCR and that the DNA is of good quality. Positive and negative controls have been included on all runs. Based on a preliminary analysis of the data from these 5 plates, the number of alleles per marker is as follows: RM 124 (3), RM316 (9), RM413 (14), RM536 (9), RM431 (8), and RM19 (14). Each marker detects heterozygosity for a few accessions. The genotyping on the remaining samples (including those in the remaining shipment) will be accomplished by the end of September 2005.

Genotyping data has been obtained for part of the panel of SSRs for 1536 accessions of the rice collection.

The intended rice collection was to comprise 3000 accessions from the holdings worldwide. As yet, no permission has been given by the Chinese government to include their materials.

Hence, the current rice collection of 2768 accessions does not include these materials. Shipment of the remaining the remaining 1232 samples was delayed while waiting for the DNAs from the Chinese nominated varieties. While those nominations from WARDA have now been promised, the importation, growth, and preparation of DNAs on these 98 accessions will require at the least 3 months. Hence, the WARDA set should be genotyped and analyzed as a separate activity.

Genotyping at several locations met with unexpected problems due to equipment failure (IRRI), lack of sufficient consumables for the genotyping platforms (IRRI), technical support for data analysis (CAAS), working primers (CIAT), delayed receipt of funds (EMBRAPA), and corruption of analyses during a software upgrade (CIRAD). Genotyping by WARDA was delayed until the visit of a research assistant from WARDA could be arranged at Cornell. The throughput needed for genotyping at IRRI will be met by splitting the effort across the BST51 and LiCor 4300 genotypers (on the receipt of SAGA software for the marker analysis).

These difficulties have largely been overcome, and genotyping is finished or well underway for the 1536 samples so far received. Preliminary analysis on partial data may be possible for presentation at the Chennai workshop “Molecular Markers for Allele Mining.” The remaining samples will be shipped in the near future, so that the rest of the genotyping can be done within 2005.

1h. Genotyping of composite germplasm set, Tier 1, sorghum

Principal Investigator:

Dr Claire BILLOT, Agropolis-CIRAD

Collaborating Scientists:

Dr Yu LI, CAAS

Dr Tom HASH, ICRISAT

Dr Rolf FOLKERSMA, ICRISAT

Dr Monique DEU, CIRAD

Dr Jean-François RAMI, CIRAD

MID-YEAR REPORT

Work was divided into three tasks from which task2 and 3 were followed during this period

1. Task 1: Choice of 50 markers from a subset of 100 markers based on (i) their reproducibility in the different partners laboratory, using different technologies (CIRAD: M13-tail, Licor IR2; ICRISAT: direct primer labelling, ABI 3700; CAAS: direct primer labelling, ABI 3700); (ii) their information content; and (iii) their location on sorghum genetic map. From these 50 markers subset, only 30 will be used during the first year.
2. Task 2: 700 accessions genotyped with 30 markers.
670 accessions chosen in SP1-C1 were genotyped with 33 markers shared between the three laboratories (11 in each lab). Controls were used in Agropolis-Cirad and ICRISAT experiments. Verified dataset includes Agropolis-Cirad and ICRISAT data (i.e 22 loci). CAAS data still need to be verified carefully before being included in the final dataset. Data analysis has been performed and will be presented in the Chennai meeting, August 2005.

3. Task 3: Development of new markers, EST-derived (ICRISAT) or neutral (Agropolis-CIRAD and CAAS).
600 EST-derived SSR markers have been developed by ICRISAT and were tested for polymorphism on 16 accessions (parental lines of six mapping populations, *S. propinquum*, one *S. bicolor var margaretiferum* accession, BTx623, R16 and recurrent parents used for marker-assisted backcrossing of the stay-green trait).

In parallel, new neutral SSR markers were developed by Agropolis-Cirad: around 250 new markers have been designed, from which 144 were tested for polymorphism on 15 different accessions. 40 were mapped on a Sariaso10xSSM249 progeny.

This will give opportunities to complete the final 50 marker set.
Complete results will be available for the Data Analysis meeting in June 2005.

Deviations from the work plan

Task 1:

- 33 markers (instead of 30) were chosen to genotype the overall accessions due to laboratory facilities and interest.
- Remaining markers to complete the final set of 50 need to be refined with the introduction of new markers (during Chennai's meeting, August 2005).

Task 2:

- 670 accessions (instead of 700) were genotyped.
- Complete dataset including CAAS data (11 SSR markers) will not be ready for the Chennai meeting in August 2005 due to experiment adjustments in CAAS. Raw data are however available, were collected during PI's stay at CAAS in June 2005, but need lot of refinement. This will be finished in September 2005.

Task 3:

- No development of new markers performed by the CAAS, and no remedy to that is proposed.

1i. in barley genotyping of CGIAR Challenge Program

Principal Investigator:

Zhang Jing, CAAS

MID-YEAR REPORT

Barley is one of the most important crops in the world with the longest cultivation history among all Human's crops. It not only still is a staple crop in most dry and plateau areas and provides foods for most poor people of the world even by now but also is main raw materials both for beer and for feed-processing industries in rich countries. Nevertheless with warming of the global climate, natural calamities such as drought, salinity, strong-wind and water-logging take place more frequently than before, productivity and sustainability of barley production are limited in great extent. Exploration and utilization of the related new genes in barley gene pool will give resolution to the problem, which must be based on barley genetic diversity research.

To genotype 500 barley accessions with 49 SSRs was commissioned to Institute of Crop Science, CAAS, and International Center for Agricultural Research in the Dry Areas (ICARDA) in 2004.

Through hard work last year, 500 barley accessions were selected, out of which 250 were from 23 provinces of China, including 20 wild barley and 10 bred cultivars; another 250 provided by Dr. Baum, Michael and Dr. Jan Valkoun (ICARDA), from 22 different countries, which are in the Middle East, the barley origination place mostly. 500 DNA samples were abstracted from fresh leaf tissues of the 500 barley accessions. 49 pairs of SSR primers were selected averagely on 7 chromosomes of barley, based on amplifying efficiency and polymorphism by silver-staining combined with screening on ABI 3700. All of the 500 barley accessions have been genotyped with 49 SSRs. PCRs were performed on the thermal cycler PTC 200. The PCR products were analyzed on ABI 3700. For the fluorescent labeling of PCR fragments, an economic method of M13 tail described by Markus Schuelke (2000) was used. The data in a peak value were transformed into a o,1 format and analyzed preliminarily. Alleles at each locus were counted and the difference was simply compared among chromosomes, countries and provinces and between wild and native barley.

In total, 3377 alleles were found at 49 loci (differentiated by 1 bp, may not correct and need to recount). On the average there were 69 alleles at each locus. The locus scind1691 had 124 alleles with the highest number, the next was Bamc 0018 with 115 alleles. They were situated at the chromosomes 5H and 6H respectively. Scssr00334 had only 18 alleles with the lowest number. There is the same trend in the changes in allele number among 7 chromosomes between world and Chinese barley. The chromosomes 3H, 5H and 6H have more alleles than the others, 7H has the lowest allele number. China has the most, Iran, Afghanistan, Ethiopia, Jordan, Syria and Turkmenistan have more alleles. But the number of alleles may be related to sample number. In China Tibet is the highest, Yunnan is the next in allele number. Zhejiang, Jiangsu, Qinghai, Shan'xi and Heilongjiang have more alleles also. The allelic diversity is in a line with that of phenotypes basically. Bred barley has the lowest allele number, but with almost the same to the wild barley collected in Tibet, Qinghai, Si-chuan and Yunnan. The landraces hold much more alleles than the wild and the bred barley. This result reveals that alleles may exist mainly in barley landraces, but not in the wild in China.

1j. Genotyping of composite germplasm set, Tier 1, common bean

Principal Investigator:

Matthew Blair, CIAT

Collaborating Scientists:

Maria Jose Peloso, EMBRAPA

Rosana Brondani, EMBRAPA

Shumin Wang, CAAS

Teresa Avila, CFP, Bolivia

Gloria Santana, CORPOICA, Colombia

Sandra Lorigados, INCA, Cuba

Steve Kresovich, Cornell University

Sharon Mitchell, Cornell University

MID-YEAR REPORT

The following research achievements were wrapped up in the period for this report but were begun in the proceeding year:

- 1) Analysis of a microcore (44 genotypes) with all available microsatellites (150 markers) to decide on best SSR markers to use in full genotyping experiments. Heterozygosity was determined for 57 gene-based and 72 genomic microsatellites. Analysis was completed and an article prepared and submitted for this genotype x marker set.
- 2) Design and implementation of silver stain duplex combinations for 30 microsatellites to increase the efficiency of manual sequencing gel use. This set was tested with a number of trainees from various countries (Bolivia, China, Colombia, Cuba) who found silver staining to be highly reproducible and allowed cross gel comparisons.
- 3) Design and implementation of fluorescent microsatellite panels for three automated genotyping platforms (ABI377; ABI3100 and ABI3730) using 56 microsatellites. This task was shared between collaborators at CIAT, EMBRAPA and Cornell University.
- 4) Analysis of a set of race standards for overall species population structure (120 genotypes). A decision was made to establish race structure with an additional minicore of genotypes representing all the morphological races of common bean using 30 markers. Analysis shows race structure that broadly agrees with morphological classification however new patterns of diversity were uncovered. Microsatellite analysis was very useful for distinguishing Nueva Granada and Peru races from each other and the Guatemala race from other Mesoamerican races. Misclassified genotypes based on morphological analysis or varietal name could be placed into their correct race based on microsatellite genotyping.
- 5) Analysis of selected primary center genotypes from a full core collection (180 genotypes). Genotyping of core collection was begun but DNA quality was problematic and new extractions were made for a preliminary test of the full core. These were run exclusively on an ABI3730 and were used to refine the fluorescent microsatellite panels originally designed for the ABI377. A total of nine 4-color panels were designed and implemented.
- 6) Analysis of national germplasm sets: Bolivia (171 genotypes), Brasil (560 genotypes); Caribbean (310 genotypes); Colombia (90 genotypes); China (237 genotypes) and Cuba (210 genotypes). Mass genotyping was carried out with 20 to 30 SSR markers on accessions from the partner institutions EMBRAPA - Brazil and CAAS – China and from the CIAT germplasm collection for primary centers of diversity in Bolivia and Colombia as well as a secondary center of diversity in Cuba and other nations from the Caribbean. Accessions included landraces, modern varieties and some breeding lines chosen from each of these national and FAO-designated collections.
- 7) Comparison of silver stain and fluorescent microsatellite results. Preliminary comparisons were made for microsatellite markers run on both systems and between ABI platforms.

Perspective: This is the first large scale genotyping using microsatellite markers in common bean and sets a precedent for genotyping in the species as previous studies used RAPD or AFLP fingerprinting rather than single-copy markers. This is the first integrated analysis of genotypes from primary and secondary centers of diversity and from both major gene pools of common bean using SSR markers. The single-copy, co-dominant microsatellite markers will be useful for feeding into marker assisted selection programs and for tracing lineages.

The uncovering of introgression between the gene pools will prove invaluable for association mapping studies.

A large number of bean researchers (3 – Bolivia, 1 – Brazil, 1 - China, 4 – Colombia, 1 – Cuba) were trained in evaluation of genetic diversity of common bean.

Articles submitted on microcore evaluation (Theor Appl Genet) and analysis of the Colombian national collection (Fitotecnia Colombiana).

Talks and posters presented in Congreso Nacional de Fitomejoramiento (Colombia) and posters submitted for CRSP meeting in September.

Delays were encountered in signing sub-contracts and organizing researcher/student training. Phytosanitary and IP related issues for shipping seed between partners also caused delays. Technical problems in establishing fluorescent microsatellite marker panels in common bean arose due to the limited number of markers of different size ranges and to requirement of higher quality DNA for fluorescent marker analysis than the miniprep DNA originally contemplated for the analysis. Total number of genotypes is expected to be slightly reduced given these problems and higher than expected cost per datapoint. Present genotyping is of 2000 accessions with 30 to 50 SSR markers.

1k. Cowpea Genotyping

Principal Investigator:

Morag Ferguson, IITA

Collaborating Scientist:

Sarah Hearne, IITA

MID-YEAR REPORT

A series of cowpea and bean derived genomic SSR markers have been screened for polymorphism against a set of 24 diverse accessions to ascertain polymorphism. We had concerns regarding the variability of the amplifications across accessions. We ascertained this was due to problems resulting from inaccurate quantification of DNA using the spectrophotometer in the ILRI facility. All DNA samples were re-quantified using agarose gels, new dilutions of DNA were made and consistent results are now obtained. A series of 47 SSRs were screened (originally identified as those SSRs that amplified by PCR when visualized using agarose gels), of these 17 are polymorphic across the core set of accessions. The amplification of these SSRs has been optimized individually and co-loading groups have been developed. The intra-accession variation of 100 cowpea accessions is being assessed using these 17 SSR markers. Alongside this, efforts are being placed in developing further SSR markers for diversity assessment. A Dutch funded APO, Ms. Manjeet Kaur is working on the generation of further SSRs derived from genomic DNA. In addition, a SP2C3 project is ongoing which focuses on marker development from ESTs.

Once the intra-accession assessment is complete we will start to run the diversity set of accessions using the 17 SSRs identified during the first half of 2005. Further markers will be tested for polymorphism concurrently and will be utilized when optimized.

Tangible outputs delivered:

The preliminary data from the intra-accession variation study will be presented in Chennai

There have been no significant deviations from the intended work or research outputs. However there have been significant deviations from the timeframe. These deviations resulted from the installation of equipment within the BecA facility in Nairobi. The installation of the ABI 3730 and 3100 capillary sequencers at BECA was delayed by five months during 2004 impacting significantly on the timeframe for the proposed work. In addition the machines have encountered a number of technical problems including a faulty laser. Engineers attended the machines in November 2004 and January 2005 and the problems have been rectified. Both sequencers are now functioning normally. We are now continuing with the work as planned.

11. Genotyping of composite germplasm set, Tier 1, cassava

Principal Investigator:

Martin Fregene (CIAT)

Collaborating Scientists:

Morag Ferguson, IITA

Sarah Hearne, IITA

Alfredo Alves, EMBRAPA-CNPMP

Claudia Ferreria, EMBRAPA-CNPMP

Paula Hurtado, CIAT

Carmen de Vicente, IPGRI

MID-YEAR REPORT

1. Passport data and phenotypic information completed for the whole set of 3000 accessions SP1 proposes to examine the genetic structure of a large and representative sample of crop with molecular markers and information from this analysis would be collated and stored. A sub- set of 3000 cassava accession from collections held at CIAT, IITA, and EMBRAPA was selected based on criteria that emphasizes location, to capture the broadest genetic diversity, and key agronomic traits such as Drought tolerance, resistance to major pests and diseases, adaptation to different ecologies, etc. The complete set of criteria used to select the germplasm set is listed in Table 1. Passport data of all 3000 cassava accessions was collated and sent to the responsible scientist in Sub Program 5 for processing and storage.

2. The genotyping of the remaining group of 3000 samples with 36 SSR markers completed at CIAT and IITA.

DNA samples from the 3000 accessions were assembled at CIAT from the respective collections and aliquots of each sample sent to the genotyping teams at IITA (Nairobi) and CIAT. SSR marker analysis at CIAT was by silver stained polyacrylamide gel electrophoresis of PCR amplification product of all accessions amplified with 22 SSR primer pairs completed early in May. At IITA the 14 SSR markers were analyzed by automated fluorescent capillary electrophoresis, data will be sent to CIAT at the end of July.

3. Data compilation and analysis of information generated at CIAT is being conducted at Cassava Genetics (CIAT).

Preliminary data analyses from the completed 22 SSR markers include assessment of genetic structure using principal coordinate analysis (PCoA) and multiple correspondence analyses (MCA), using the computer package NTSYS-PC, and estimation of genetic diversity and differentiation were calculated using the computer package. Analysis is ongoing and would be completed by mid August.

Tangible outputs delivered:

1. Passport data template for IITA, EMBRAPA and CIAT cassava accessions was sent to Guy Davenport (Bioinformatics SP5) in April 2005
2. Three different molecular marker data files from the Pilot study information were sent to SP5:
 - Molecular weight information / locus / genotype
 - Allele / locus / genotype
 - Binary data / genotype
3. Data from 26 SSR marker loci are ready and already being analyzed using the relevant software like NTSYS, SAS, Gensurvey, Popgene

Deviations from the work plan:

1. Due to technical problems with the genotyping facility at IITA-Nairobi, there has been a delay in completion of SSR analysis at IITA. Results will now be sent end of July for analysis at CIAT, the earlier time frame was for all SSR data to be in by May.
2. No DNA samples were sent to CNPMF because of delays in obtaining a permit from the Brazilian genetic resources (GR) council to exchange cassava germplasm with CNPMF. Consequently, 390 CNPMF accessions held at CIAT world cassava collection were included in CIAT genotyping work and the 6 SSR markers corresponding to CNPMF were evaluated at CIAT with the participation of a CNPMF scientist (Ms Claudia Fortes Ferrerira). CNPMF will evaluate with 12 markers an additional 200 accessions held at CNPMF, there has been a delay in the initiation of this work due to administrative issues on drawing up a contract.

1m. Gene pool structure of cultivated potatoes assessed by SSR marker analyses

Principal Investigator:

Marc Ghislain, CIP

Collaborating Scientist:

Jorge Núñez, CIP

Maria del Rosario Herrera, CIP

Guillermo Trujillo, CIP

Reinhard Simon, CIP

Edwin Rojas, CIP

MID-YEAR REPORT

The high throughput genotyping facility established in 2004 has produced a large dataset of microsatellite (SSR) markers for cultivated potato. The objectives of this research were to analyze the gene pool structure of the cultivated potato as well as increase the number of SSR markers for potato germplasm analyzes. Potato landraces from all cultivated species or taxonomic groups* have been included in the potato composite genotyping set. So far, 716 genotypes by 53 SSR markers constitute the potato SSR data set. The analysis of this large data set was performed by cluster analyses using Darwin software kindly provided to us by CIRAD. These results revealed both expected and unexpected structure of the cultivated potato germplasm.

The Chilean *Chilotanum* group forms a clearly separated cluster from the other tetraploid *Andigenum* group which agrees with the hypothesis *Chilotanum* deriving from *Andigenum*

through the hybridization with another species, possibly a wild species. The diploid species form a well separate structure from the tetraploid potatoes. Within diploid species, the *Phureja* group form a well separated cluster which was unexpected based on previous studies at CIP using RAPD markers. The *Stenotomum* group form a wide and highly diverse group with one major cluster composed of both *goniocalyx* and *stenotomum* accessions. This finding was expected as the species *S. goniocalyx* was recognised by some potato taxonomists to be a sub-species of *S. stenotomum*. The triploid *Chaucha* group does not form well separated clusters, instead *Chaucha* landraces are found intermixed with all other groups which is an evidence of its hybrid origin between diploid and tetraploid potatoes. Finally the diploid *Ajanhuiri* group, the triploid *Juzepsuki* group, and the pentaploid *Curtilobum* group appeared to separate markedly with all other taxonomic groups. This may indicate an independent domestication process from other wild ancestors. Many of potatoes are indeed refer to as the bitter potatoes and are usually found in the high Andes. Hence, the native potato germplasm appears to be highly structured.

*We adopt hereafter the taxonomic group nomenclature of Huamán and Spooner, 2001.

Tangible outputs delivered:

Poster presentation at the Plant Animal Genome conference, San Diego, USA, January 2005

A number of landraces were omitted from the 2004-2005 analysis due to delays in obtaining leaf tissue from in vitro plants or acquisition status which is now being clarified. Since phenotyping data are available for these, their genotyping would allow them to be included with germplasm that will be used for association mapping and other activities of the GCP. In addition, no advanced hybrids or improved varieties have been analyzed yet, beyond a small start made for identification purposes in our breeding program. CIP and NARS breeders have their preferred gene pool which has not been genetically characterized. This germplasm was identified by potato breeders in 2004 and to a large extent included in the composite genotyping set. However, due to limited experience with genotyping large collection and with the new HTG facility, we did not reach our target of genotyping 1083 genotypes with 50 SSR markers. We decided to leave out breeder's material in order to produce a valuable data set which has intrinsic interest. We have now gained significant experience and see that it is essential to genotype the germplasm used by breeders. The objective is to understand what germplasm has been exploited by breeders as well as engage this community in a common research strategy making better use of the tools derived from molecular genetics. We are currently seeking additional funds to complete the Composite Genotyping Set.

1n. Genotyping of composite germplasm set, Tier 1, Musa

Principal Investigator:

Nicolas Roux, IPGRI-INIBAP

Collaborating Scientist:

Isabelle, Hippolyte, CIRAD

Maria Kolesnikova-Allen, IITA

MID-YEAR REPORT

Since January 2005 a third set of accessions have been defined from the collection of CIRAD Guadeloupe. 207 accessions have been harvested and extracted at CIRAD. DNA extracts from these accessions were sent to IITA. From this third set, 189 accessions have been

partially analysed : 20 ssr markers have been amplified and amplification products revealed by Licor gel migration. Analysis of these gels is still to be performed.

In the same time, analysis of 93 accessions of the second set coming from IITA have been performed on 24 ssr markers.

As many accessions were triploids, the classical softwares used to analyze the results were not suitable for *Musa* germplasm. Some efforts have been carried on to automate reading and recording of the migration results. Then Quantar Pro is now used as software. According with other species involved in the project, new experimental conditions have been developed to get more robust and reliable results.

Consequently new reading with the software of the first set of 48 accessions genotyped with 100 ssr markers revealed the need to make again some experiments to ensure the first results.

The work done in the frame of the project is summarized on table 1.

Supporting distribution of reference germplasm will be done for the GS1 (48 accessions= mini-core collection) + 3 other reference *Musa* clones. In addition DNA of plants which served to prepare the BAC libraries will be extracted and distributed to GCP partners according to the needs.

In IITA, due to the changes in personnel, the workflow was affected. Starting from May 2005, the work resumed and the PCR optimization for the set of 25 SSR markers, assigned to IITA, was performed. The conditions for running PCR samples on ABI 3100 were optimized. The bulk screening of IITA set (192 DNAs) and first set of samples received from CIRAD (207 DNAs) had been initiated. The analysis of raw data is being performed using GeneMapper v. 3.5 software which allows easy identification of triploids and tetraploids.

In order to uniform the data across the laboratories the DNAs were first screened with set of 4 common SSRs. The resulting data is being analysed.

Tangible outputs delivered:

- An efficient method to genotype and analyzed polyploidy accessions with Licor genotyper.
- A set of 93 accessions genotyped with 24 ssr
- The PCR and running conditions on ABI 3100 had been optimized in IITA

A problem, evocated in the previous report and still existing, is the inability to compare the results between the two laboratories involved in the project (i.e. IITA and Cirad).

The third set of DNA samples should be provided by other institution, and was at last devoted to Cirad. This change in the task assignment to our laboratory delayed the recovery of results.

Table 1: Work done on *Musa* germplasm genotyping since 2004.

	DNA extraction	PCR	Gel migration	Analysis	Verification
GS1: 48 accessions (CIRAD field collection)	x	x	x	x	x → Some experiments to be made again
GS2: 192 accessions (IITA field Collection) 93 accessions	X**	26 ssr	26 ssr	24 ssr	-
99 accessions	X**	-	-	-	-
GS3: 347 accessions (CIRAD + CARBAP field collection) 207 (CIRAD collection) 189 accessions	X	20ssr	20ssr	-	-
18 accessions	X	-	-	-	-
140 (CARBAP field collection)	-	-	-	-	-

*GS# = Germplasm set

** Done by IITA

2. Supporting Distribution of Reference Germplasm

Principal Investigator:

Germplasm bank curators

Participating Institutions:

Any CG centre mandated for germplasm conservation for a particular crop studied in year one

MID-YEAR REPORT

Activities to be reported at the end of 2005.

3a. Development of Composite Collection of Tier 2 (Orphan) Crops—Finger Millet

Principal Investigator:

HD Upadhyaya, ICRISAT

Collaborating Scientists:

C.T. Hash, ICRISAT
Subhash Chandra, ICRISAT

MID-YEAR REPORT

- Composite collection of finger millet consisting of 1000 accessions using phenotypic characterization and evaluation data, geographical origin, and taxonomy was developed.
- Composite collection includes 508 accessions of core developed at ICRISAT, 114 accessions from the entire collection with at least one accession representing each cluster, 50 accessions from core collection developed by All India Coordinated Small Millets Improvement Project (AICSMIP) available with ICRISAT and trait specific accessions (high and low) for different traits.
- Composite collection includes accessions from 21 countries and represents all the four cultivated and two wild races.
- Composite collection will be planted in field during the third week of July, and representative plants selected.

Tangible Outputs delivered

- Composite collection of 1000 accessions established.
- Seed of representative plant of each accession will be collected.

Deviations from the work plan:

Nil.

3b. Molecular Characterisation of Tier 2 (Orphan) Crops--Pigeonpea

Principal Investigator:

HD Upadhyaya, ICRISAT

Collaborating Scientists:

Subhash Chandra, ICRISAT

KB Saxena, ICRISAT

MID-YEAR REPORT

- Composite collection of pigeonpea comprising 1,000 accessions was established using the phenotypic characterization and evaluation data.
- This collection consists accessions of pigeonpea mini-core, comparator mini-core, and representative accessions of landraces, breeding lines, genetic stocks and wild species
- The composite collection will be planted in field during the first week of August, and leaf material will be used for DNA extraction for genotyping studies.
- Thirty SSR primer pairs were selected to pre-screen the mini-core accessions to identify 20 polymorphic primer pairs, which will be used for genotyping the composite collection.

Tangible Outputs delivered:

- Pigeonpea composite collection comprising 1,000 accessions was established.

Deviations from the work plan:

- Nil

3c. Application of Molecular Markers for Gene Pool Division and Heterosis Estimation under Drought Stress Conditions in Sweet Potato

Principal Investigator:

Marc Ghislan, CIP

Collaborating Scientists:

Wolfgang Grüneberg, CIP

Jorge Benavides, CIP

Robert Mwangi, NAARI, Uganda

MID-YEAR REPORT

All the activities currently undertaken with the molecular characterization of sweetpotato are aiming to meet 3 objectives: (i) SSR fingerprints for diversity assessment, (ii) intra- and inter gene pool and crossings, and (iii) field evaluations, to investigate to which extend heterosis and genetic distances in sweetpotato are associated [this activity will be funded by core budget or other projects (details of workplan see Table 1)]

1250 clones of the CIP germplasm have been multiplied under greenhouse and field conditions. 500 clones of better agronomic performance and diverse origin will be used to estimate genetic distances.

In August we started to isolate DNA from the first 50 clones to start the estimation of genetic distances with 43 SSR markers [including 20 old and 23 new SSRs developed in this project (see 1.2)]

In addition to the 20 available SSR markers (Table 2), we developed 23 new SSR markers (Table 3). For those markers with the initials IB-JB the library of the potato Research Center Canada / Fredericktown (contact: Xiu-Qing Li) was used and for those with the initial IB-JS the library of ARC Seibersdorf Research GmbH (contact: M. Berenyi / Department of Biotechnology), the library of Korea University / Seoul (contact: Jung Myung Bae / Plant Molecular Breeding) and the library of USDA/ARS (contact R.L. Jarret / Plant Genetic Resources) were used.

In total 43 SSR markers tested for their value (Table 2, Table 3) are currently available for this project.

So far the SSR primers published by Padmavathi Nimmakayals et al. at Plant and Animal Genome XII Program (January 10 - 14, 2004) are not available for this project, but we are in contact with the author to achieve this.

Moreover, we expect that an additional source of SSR markers will be 10,000 ESTs that are currently sequenced by CIP within a co-operation with ARC Seibersdorf Research GmbH.

Activity ii: Intra- and inter gene pool and crossings:

According to work plan (Table 1) Not yet started.

Activity iii: Field evaluations, heterosis and genetic distances:

According to work plan (Table 1) Not yet started.

Table 1. Work plan and Activities.

<i>Component / Activity description</i>	2005			2006		
	Apr. – Jun.	Jul. – Sept.	Oct. – Dec.	Jan. – Mar.	Apr. – Jun.	Jul. – Oct.
Activity (i) SSR fingerprints for diversity assessment						
• Estimation of Genetic Distances on basis of validated markers	X	X	X			
• Validation of new SSR markers	X	X				
• Gene Pool Formation			X			
• Publication			X	X		
Activity (ii) Intra- and inter gene pool and crossings						
• 64 Inter- & 128 Intra- Gene Pool Crossings				X		
Activity (iii) Field evaluations, Heterosis and Genetic Distances (funded by core budget or other projects)						
• Field Experiments Peru					X	X
• Field Experiments Uganda					X	X
• Publication						X

Table 2. Names, origin and characterization of SSR Primers available at 1. April 2005.

Name	Primer Origin	Primer forward	Primer reverse	Tmp (C)	value
IB242	Uni. Lousia.	GCGGAACGGACGAGAAAA	ATGGCAGAGTGAAAATGGAACA	58	Good
IB255	Uni. Lousia.	TGGGCATTCTCATATTTTCT	AAGGACCACCGTAAATCCAA	57	Low
IB286	Uni. Lousia.	AGCCACTCCAACAGCACATA	GGTTTCCCAATCAGCAATTC	57	Medium
IB297	Uni. Lousia.	GCAATTTACACACAAAACACG	CCCTTCTTCCACCACTTTCA	58	Medium
IB316	Uni. Lousia.	CAAACGCACAACGCTGTC	CGCGTCCCCTTATTTAAC	58	Good
IB324	Uni. Lousia.	TTTGGCATGGGCCTGTATT	GTTCTTCTGCACTGCCTGATTC	56	Good
IB-R03	CIP Lima	GTAGAGTTGAAGAGCGAGCA	CCATAGACCCATTGATGAAG	73	Good
IB-S07	CIP Lima	GCTTGCTTGTGGTTTCGAT	CAAGTGAAGTGATGGCGTTT	69	Good
IB-R08	CIP Lima	GGCGACACCTTAGAGTAT	CACCCCCTATTCACAA	78	Good
IB-S09	CIP Lima	GCTGCTCAATCCCTCTCTT	GGAACTCGATACAGCGTGGT	90	Good
IB-S10	CIP Lima	CTACGATCTCTCGGTGACG	CAGCTTCTCCACTCCCTAC	87	Good
IB-S11	CIP Lima	CCCTGCGAAATCGAAATCT	GGAATTCCTCTGCCTTGTTG	82	Good
IB-R12	CIP Lima	GATCGAGGAGAAGCTCCACA	GCCGGCAAATTAAGTCCATC	71	Good
IB-R13	CIP Lima	GTACCGAGCCAGACAGGATG	CCTTGGGATTGGAACACAC	83	Good
IB-R14	CIP Lima	CCTATGGCAATTCGGTCACT	GGAACATTGCCTACATCTG	50	Low
IB-R16	CIP Lima	GACTTCCTTGGTGTAGTTGC	AGGGTTAAGCGGGAGACT	76	Good
IB-S18	CIP Lima	CTGAACCCGACAGCACAAG	GGGAAGTGACCGGACAAGA	42	Médium
IB-R19	CIP Lima	GGCTAGTGGAGAAGGTCAA	AGAAGTAGAACTCCGTCACC	76	Good
IB-R20	CIP Lima	CTTCACTCTGCTCGCCATTA	GTAAGTGGACGGGAGGATGA	68	Medium
IB-R21	CIP Lima	GACAGTCTCCTTCTCCATA	CTGAAGCTCGTCGTCAAC	58	Good

Table 3. New SSR Primers - Names, origin and characterization (31. August 2005)

Name	Primer Origin	Primer forward	Primer reverse	Tmp (C)	Value
IB-JB-S01	CIP	TCCTCCACCAGCTCTGATTC	CCATTGCAGAGCCATACTTG	56	Medium
IB-JB-S02	CIP	CTGGGAATGAAGTATCC	CTCACAGATTTGGTAGCAC	52	Medium
IB-JB-S04	CIP	CTTCCATGACTAAGGGACAA	CCACCACCAATAACAACAAG	54	Medium
			GAACAACAACAAAAGGTAGAG	63	Good
IB-OY-CIP1	CIP	CCCACCCCTTCATTCCATTACT	CAG		
IB-OY-CIP2	CIP	GTAACCTGTCAGCCATCTGT	CCTAGTGGGTATTTGCAGAG	63	Good
IB-OY-CIP5	CIP	CCTCAACGAATTTGACCTC	GATGACGGTGTGTCTGAAG	65	Good
IB-OY-CIP7	CIP	GGTTTGACCGTGGAGTTGTT	GGACGAACTTTCCCAAATCA	48	Medium
			TGGAGAGATAACCCTAGCAGTC	48	Medium
IB-OY-CIP8	CIP	ATAACCCACCCCTTCATTCC	T		
		AGACTGCTAGGGTTATCTCTCC		48	Good
IB-OY-CIP9	CIP	A	GACATTGCCAAGGACACTGA		
IB-OY-CIP12	CIP	CGTAAAGCAAAAGCGAGGAC	GACGGGAATTTGGAAGGAAG	63	Low
		CGTGCTTGAGGTCTGAGTAGA		53	Medium
IB-OY-CIP13	CIP	A	TTCCCTAGAAGCTGCGTGAT		
		ATCTATGAAATCCATCACTCTC		57.6	M.-Good
IB-JS-175	CIP	G	ACTCAATTGTAAGCCAACCCTC		
		TCTAGAGGGGAAAGTGCTTCC	AGTGCAAAACAAAGCCCCAAG	50	M.-Good
IB_JS-27	CIP	T	G		
IB-JS-522a	CIP	ACCCGCATAGACACTCACCT	TGACCGAAGTGTATCTAGTGG	57	M.-Good
IB-JS-10a	CIP	TCAACCACTTTCATTCACTCC	GTAATTCACCTTGCGAAGC	58.6	M.-Good
IB-JS-67a	CIP	CACCCATTTGATCATCTCAACC	GGCTCTGAGCTTCCATTGTTAG	58	M.-Good
		GAGAAATCATAGCATAACGGAG		56.5	M.-Good
IB-JS-162E	CIP	TAA	ACTACTACTACTACTACTACT		
IB-JS-315E	CIP	TAGGTGTGTTTATGGGAGATT	GGGACTTGACTTTCATTATTAC	56.5	M.-Good
IB-JS-1809E	CIP	CTTCTCTGCTCGCCTGTTT	GATAGTCGGAGGCATCTCCA	60.8	M.-Good
IB-JS-116a	CIP	TCTTTTGCATCAAAGAAATCCA	CCTCAGCTTCTGGGAAACAG	57-58	M.-Good
IB-JS-130	CIP	CCAGTTGTGGCGGAGAAAG	GGATTCTGTGAGGGACAACA	60.7	M.-Good
IB-JS-664	CIP	CACATGCCATGGACGCTCCAA	GATTCTTCTCCTTCCACTCCT	55-57	M.-Good
IB-JS-263	CIP	CTCTGCTTCTCCTGCTGCTT	GTGCGGCACTTGTCTTTGATA	55	M.-Good

3d. Assessment of Genetic Diversity of West African *Dioscorea* sp Collection

Principal Investigator:

R.Asiedu, IITA

Co-Principal Investigators:

V.Mahalakshmi, IITA

H. Chair, Cirad

MID-YEAR REPORT

Passport and characterization data for 3200 accessions of *Dioscorea* sp was compiled and collated. A preliminary set of 48 accessions (five from each group except *abysinnica*) will be assembled to optimize the 20 SSRs from CIRAD. This study will give both the level of amplification and polymorphism within each species. A core collection of about 350 accessions based characterization data would be made stratified by the species. Based on the optimization results strategy to fingerprint relevant species would be decided. As the ploidy nature of the accessions with in species also vary the ploidy of the core collection will also be determined.

Table: List of available accessions of *Dioscorea* at IITA genebank.

Groupby_Genus	Groupby_species	Total
Dioscorea	abysinnica	1
Dioscorea	alata	816
Dioscorea	bulbifera	70
Dioscorea	bulkiliana	6
Dioscorea	cayenensis	59
Dioscorea	dumentorum	53
Dioscorea	esculenta	20
Dioscorea	manganotiana	8
Dioscorea	preusii	10
Dioscorea	rotundata	2157

There is a delay in starting of the project – but a doctoral student will be joining and the work is expected to progress well.

3e. Genotyping a Composite Germplasm Set of Lentil

Principal Investigators:

Bonnie J. Furman, ICARDA; Michael Baum, ICARDA

Collaborating Scientist:

Christian Jung, Universität Kiel, Germany

MID-YEAR REPORT

The International Centre for Agriculture in Dry Areas (ICARDA) has a global mandate for research on lentil improvement. As such, ICARDA houses the world collection of *Lens*, which includes 8789 accessions of cultivated lentil from 70 different countries, 1146 ICARDA breeding lines, and 574 accessions of 4 wild *Lens* species representing 23 countries. From this collection, a composite germplasm set of 1000 accessions was identified consisting of landraces, wild relatives, and elite germplasm and cultivars. The methodology combined classical hierarchical cluster analyses using agronomic traits and two-step cluster analyses using agro-climatological data linked to the geographical coordinates of the accessions' collection sites, ensuring that the resulting composite collection contains representative genetic diversity and that the agro-climatological range of the species is represented. In addition, researchers at ICARDA included 64 accessions of breeding material and landraces important to lentil improvement for their resistances to biotic and abiotic stresses. A total of 28 wild *Lens* accessions were also included, 16 of which have been identified as having resistance to biotic and abiotic stress.

Individual plants of each accession was planted in spring 2005 in a plastic house in Tel Hadya, ICARDA, Syria in order to collect seed at the end of the growing season. Unfortunately, an infestation of thrips during the growing season made it impossible to collect the DNA. However, the population was planted again in July 2005 in growth rooms at ICARDA. DNA extraction will start in September. SSR analysis will then be done afterwards.

Deviations from the work plan:

DNA extraction has been delayed. Accessions have been replanted and DNA extraction will be done in September.

3f. Molecular Characterisation of Tier 2 (Orphan) Crops

Principal Investigator:

HD Upadhyaya, ICRISAT

Collaborating Scientists:

Subhash Chandra, ICRISAT

JFM Valls, EMBRAPA

MC Moretzsohn, EMBRAPA

S Leal-Bertioli, EMBRAPA

Patricia Guimarães, EMBRAPA

David Bertioli, Universidade Catolica de Brasilia, Brazil

MID-YEAR REPORT

- Groundnut composite collection with 1,000 accessions (about 850 accessions from ICRISAT, about 150 from EMBRAPA) was established using the available phenotypic characterization, evaluation data, geographic origin and taxonomy.
- The composite collection consisted of 184 mini-core and 184 accessions of comparator mini core, landraces, breeding lines, genetic stocks, and wild *Arachis* species.
- Accessions of the composite collection were planted in June 2005 and leaf material used for DNA extraction for genotyping.
- Twenty SSR primer pairs were selected for pre-screening groundnut mini-core accessions, and from this data 10 polymorphic primer sets will be selected to evaluate remaining accessions of composite collection
- Ten polymorphic primers selected by EMBRAPA will also be used for fingerprinting the composite collection

Tangible outputs delivered:

- The groundnut composite collection consisting of 1,000 accessions was established.
- DNA extraction from leaf material of 1000 accessions and quantification completed.
- Twenty SSR primer pairs were selected for genotyping studies

Deviations from the work plan:

None.

3g. Molecular Characterisation of Tier 2 Crops: Coconut

Principal Investigator:

Jean Christophe Glaszmann, Cirad-Agropolis

MID-YEAR REPORT

The objective of the project is to characterize a set of a thousand genotypes with at least twenty SSR loci.

The materials have been chosen by scientists at Cirad on the basis of the COGENT database, on order to best represent the coconut diversity. The thousand genotypes represent close to 70

populations. They include the 544 trees currently described with 13 SSRs within the COGENT initiative.

The molecular characterization consists in adding 9 new loci and complementing the data with 456 new accessions.

By the end of July, all accessions have been surveyed with all SSRs. Results are being scored using Saga. A first examination of the results shows that few data are missing.

The second semester will be used to score the markers, to conduct limited experiments to compensate the data still missing and to analyse the global structure of coconut diversity worldwide.

Tangible outputs delivered:
None yet.

Deviations from the work plan:
No deviation.

4. Validation of Diversity Arrays Technology (DArT) as a Platform for Whole Genome Profiling in Orphan Crops

Principal Investigators:

Andrzej Kilian, DArT P/L, Australia
Carmen de Vicente, IPGRI
Jean Christophe Glaszmann, Cirad-Agropolis

Co-Principal Investigators:

Eric Huttner, DArT P/L, Australia
Peter Wenzl DArT P/L, Australia
Ange-Marie Risterucci, Cirad-Agropolis

Collaborating Scientists:

Ken McNally, IRRI
Claire Billot, Cirad-Agropolis
Michael Baum, ICARDA
M Fregene, CIAT
Nicolas Roux, IPGRI-INIBAP
Patricia Lebrun, Cirad-Agropolis
Everard Jayamanne, Coconut Research Institute, Sri Lanka
Prapit Wongtiem, Rayong Field Research Station, Thailand

MID-YEAR REPORT

The DArT technology has been proven efficient to reveal numerous polymorphic markers in cassava using cultivars and wild relatives (Xia et al, 2005). A scientist from Thailand (Prapit Wongtiem, Rayong Field Research Station, Rayong, Thailand) has joined DArT Pty Ltd in Canberra, Australia, on 23rd April 2005 for nine months (planned departure at the end of January 2005). Ms Wongtiem is preparing new libraries and will extend the analysis to a large number of cultivars, focusing on accessions with large variation for dry matter content (DMC). DNA was extracted from 122 cultivated and 19 wild accessions and targets are prepared for the first array hybridization. The selection of polymorphic clones from the initial array has been completed and these clones will be included in the analysis of new accessions.

The work on coconut has started. The group at Cirad extracted DNA samples from 96 accessions and sent it to DArT P L for array development. A scientist from Sri Lanka (Chandrika Perera, Coconut Research Institute, Lunuwila, Sri Lanka) has arrived in Canberra on July 13th 2005 for a six month stay. She extracted DNA from c 20 Sri Lankan accessions which will be genotyped together with over 100 accessions selected by CIRAD scientists with arrays developed by Dr Perera. Several complexity reduction methods were tested and one promising method has been identified (PstI/BstNI). The first library is currently being developed.

Arrangements have been made for a scientist from France (Ange-Marie Risterucci, Cirad) to join DArT in late August for a three-month stay which will serve for developing the work on banana. DNA sources were identified and will be delivered to DArT PL by Dr Risterucci.

The selection process for the set of genotypes continues. The arrays will be hybridized within the next 1-3 months.

The comparison of the diversity revealed with DArT to that revealed with SSRs will be performed in September/October using representative samples of rice, sorghum and wheat (see deviations).

Tangible outputs delivered:

- First cassava array enriched for polymorphic clones created
- Targets prepared for all cassava samples ready for hybridization with the arrays

The comparison of DArTs vs other types of markers was planned for all six crops, including rice, sorghum and barley using samples of germplasm extracted from the reference sample identified by the GCP. For all three crops the GCP genotyping work has been delayed and the reference samples will not be available before months. In order not to delay the DArT work, some samples have been constituted by Agropolis based on previous work for rice and sorghum, using materials that are being genotyped by the GCP. The work on barley being far from completion, it has been decided to move to bread wheat (for which DArT PL also has commercial arrays), for which a representative sample and DNA extracts were readily available from INRA Clermont-Ferrand, Dr F Balfourier.

For personal reasons, Dr Jayamanne was replaced by Dr Chandrika Perera from the same Coconut Research Institute, Sri Lanka.

The list of collaborators will be modified accordingly.

5. Assessing EcoTILLING as a Methodology for Targeted Genotyping and SNP Discovery

Principal Investigator:

Kenneth L. McNally, IRRI

Co-Principal Investigator:

Claire Billot, Cirad-Agropolis

Collaborating Scientists:

Luca Comai, University of Washington

Jeff Harford, Li-Cor, Inc., USA

Abdelhafid Bendahmane, Unité de Recherche en Génomique Végétale, INRA, and CNRS, France

MID-YEAR REPORT

A workshop on EcoTILLING dedicated to the NARES titled “SNP discovery through ECoTILLING” was held February 28 – March 3, 2005 at IRRI. 16 external and 3 internal participants attended this training course. The external trainees were from Bangladesh [*1 with partial support*], China [*4 with full (1), partial (2), or no (1) support*], France [*1 with support from IRRI and CIRAD*], India [*2 with full (1) or partial (1) support*], Indonesia [*1 with full support*], Korea [*2 self-supporting*], Philippines [*2 with full support*], Taiwan [*1 self-supporting*], Thailand [*1 with full support*], Vietnam [*1 with full support*]. The internal participants were from Germany, India, and Australia. A second similar workshop will be organized at the end of the year 2005 or beginning of year 2006 at CIRAD.

The training was largely hands on with most of the time devoted to lab sessions which included demonstration of the preparation of celery juice extract (CJE) for the nicking assay, performance of PCR amplification, CJE digestion, PAGE analysis on the LiCor 4300, design of primers, and analysis of EcoTILLING images.

Training will be followed up at Cirad by the presence of a PhD student from Ceraas, Senegal funded by a SP5 fellowship, using Ecotilling to stay-green related genes in sorghum from July to September 2005.

A set of tentative candidate genes was given in the proposal: protein phosphatase 2a-4 (CH10), DREB2 (CH1, annotated by TIGR as DREB1), a putative AP2 domain TF (CH1), trehalose 6-phosphatase (TPP, two loci on CH2, one on CH7), viviparous-14 or NCED (vp14, CH12), 14-3-3 protein (CH2), MAPK (CH7), extensin (CH10), TRAB1 ABA responsive TF (CH10), Ca-dependent protein kinase or CDPK (CH2), Fiery1 for inositol polyphosphate 1-phosphatase (CH3), sucrose synthase (CH7), BZIP protein (CH1), and actin depolymerizing factor (ADF, two loci on CH2 and CH7). Primers were designed for these genes and checked by e-PCR using BLAT to the TIGR assembly version 2 for rice. Unlabelled primers were obtained for all of the genes and the amplification efficiency was checked. Good amplification products with the expected size were obtained for most target primer pairs except those for one of the TPP loci on CH2, the ADF locus on CH7, TRAB1, and Fiery1. New primers for these loci are being designed. Labeled primers for use on the LiCor have been obtained and verified for DREB2, putative AP2 domain, one of the TPP loci on CH2 and the one on CH7, MAPK, vp14, and the ADF on CH2. These primers need to be tested for their efficiency of amplification on Sorghum and Musa. Furthermore, the sequence homology of these loci to known Sorghum and Musa sequences needs to be defined.

Meanwhile, tests have been performed with the Waxy gene on sorghum and simulations of polyploidy were performed with bulks of genetically diverse DNA.

A modified ecotilling procedure has been adopted at IRRI that allows detection of cleavage products on agarose gels. This was effort was based on reports by several authors indicating it was possible to detect CclI cleavage products on denaturing PAGE with SYBR gold detection for tomato (Yang et al 2004 Mol. Breeding 14:21-34) and on agarose gels for mouse mutants (Greber et al 2005, Human Mut. 25:483-490) and human mitochondrial mutants (Bannwarth et al 2005, Human Mut 25:575-582). For this modified technique, we have compared the efficiency of detection on agarose for putative SNPs detected on the LiCor using the GCP microcore panel of 48 accessions for one of the TPP loci on CH2, the ADF locus, and 14-3-3 protein. Comparable results have been obtained. This modification will allow quicker and less costly screening than on the LiCor platform since unlabeled primers are used and shorter gel runs are possible.

Another modified procedure has been developed at CIRAD with the use of SP6 and T7 tailed primers. It reduces significantly the cost of labeled primers, especially for low number of individuals. Tests comparing the two methods have shown no significant difference. It is thus used extensively for sorghum at Cirad.

EcoTILLING on the LiCor has been accomplished on from 48 (for all available labeled primers) to 1000 (for the putative AP2 domain TF) GCP rice accessions at IRRI. We have accomplished ecotilling on agarose for 48 GCP accessions for BZIP, CDPK1, sucrose synthase, 14-3-3, one of the TPP loci on CH2, MAPK. We have sequence data on selected, representative haplotypes for the putative AP2 domain TF and on DREB2.

Options for sequencing of the 10 genes on the 48 samples each for rice, sorghum and musa have been explored with three companies. The offered cost per read from the commercial section of Beijing Genomics Institute is \$2. Agencourt (USA) was offering \$2 per read during June. Whether or nor this pricing could be available at the time samples are ready for sequencing will be explored. At Genaissance (USA), the cost might be on the order of \$2.50 per read.

Tangible outputs delivered:

- Workshop on ecotilling delivered to 19 NARES and ARI participants at IRRI.
- Verified primers available for ecotilling for 13 loci of 14 candidate genes in rice.
- Ecotilling data for rice on from 48 to 1000 accessions generated on 7 loci for 6 target genes using labeled primers on the LiCor and for three additional genes using unlabeled primers on agarose.
- Two modified procedures were developed:
 - Use of unlabeled primers and agarose gels. This technique promises to accelerate initial screening.
 - Use of tailed primers to minimize cost of the screening.
- Limits of the technique accessed when mutations concern indel of large number of bp (above 8 bp).

Identification of homologous loci in Sorghum and Musa is behind schedule. This should be prioritized to facilitate the final choice of target loci for use across the three species.

The adoption of screening on agarose gels and with tailed primers were not part of the work plan. Yet, these techniques promise to ease the step of screening large numbers of germplasm and large number of loci. Lines carrying putative SNPs can then be confirmed on the LiCor platform as well as by sequencing.

6. Supporting Emergence of Reference Drought Tolerance Phenotyping Centre

Principal Investigator:

Frederico Ozanan Machado Durães, EMBRAPA

Co-Principal Investigators and Collaborating Scientists (EMBRAPA):

Antonio Carlos de Oliveira

Antonio Marcos Coelho

Camilo de Lélis Teixeira Andrade

Elto Eugenio Gomes e Gama

Fredolino Giacomini dos Santos

Paulo Emílio P. de Albuquerque

Manoel Xavier dos Santos

Reinaldo Lúcio Comide

Beatriz da Silveira Pinheiro

Cleber Moraes Guimarães

Orlando Peixoto de Moraes

Nataniel Franklin de Melo

Luiz Balbino Morgado

Hélio Wilson Lemos de Carvalho

Luciana Marques de Carvalho

Milton José Cardoso

Edson Alves Bastos

Francisco Rodrigues Freire Filho

Maria da Glória Trindade

Walter Quadros Ribeiro Jr.

MID-YEAR REPORT

This first year (Semester 1, 2005) of CPG DPNetwork Project operations has been dedicating great attention on team organization and improving the sites-specific qualification for the cereals and legumes phenotyping network aiming drought tolerance.

We are working on CPG DPN Project thematic including five components project (CP), with co-PIs, as follow:

- CP0- Management of Project/Dr. Frederico Ozanan Machado Durães;
- CP1- Precision Site-Specific Experimental and Farming to Water Dynamics and Cereals and Legumes Phenotyping /Dr. Paulo Emílio Pereira de Albuquerque;
- CP2- Genetic Material per Crop Specie/Dr. Elto Eugenio Gomes e Gama;
- CP3- Traits for each Crop Specie under Evaluation for Drought Tolerance/Dr. Frederico Ozanan Machado Durães;
- CP4- Protocols of Methods and Techniques of Water Stress Control and Monitoring for Cereals and Vegetables/Dr. Reinaldo Lúcio Gomide; and,
- CP5- Structure, Maintenance and Management of a Data Bank, and Modelling/Dr. Camilo de Lélis Teixeira Andrade.

Also, during Year 1 (2005) the activities operationalization of CPG DPN Project, has been realized using the sites-specific experimentais with the following task technical sponsors:

- Site Sete Lagoas-MG/Dr. Reinaldo Lúcio Gomide, Embrapa Maize and Sorghum
- Site Janaúba-MG/Dr. Paulo Emílio Pereira de Albuquerque, Embrapa Maize and Sorghum
- Site Santo Antônio de Goiás-GO, and, Sítio Porangatu-GO, Dr. Cleber Morais Guimaraes, Embrapa Maize and Sorghum
- Site Teresina-PI, and, Site Parnaíba, PI, Dr. Edson Alves Bastos, Embrapa Mid-North Agriculture
- Site Planaltina-DF, Site Santo Antonio de Goiás-GO, Site Passo Fundo-RS, Dr. Walter Quadros Ribeiro Jr., Embrapa Wheat/CPAC, including activities on Embrapa-CNPAF and CNPT.

(1) Sites for Drought Phenotyping Network Project (CPGeneration , Semester 1, 2005		
Sites/Local	Crops	Genetic Material⁽⁴⁾/Phenotyping Strategies
Sete Lagoas, MG ⁽²⁾	maize, sorghum	Access and elite/ Preliminar
Janaúba, North region of MG ⁽³⁾	maize, sorghum	Access/ Preliminar
Santo Antônio de Goiás, GO ⁽²⁾	rice, common bean, wheat	Access, elite, segregation/ Preliminar
Porangatu, Nort region of GO ⁽³⁾	rice	Access/ Preliminar
Teresina, PI ⁽³⁾	maize (after Aug 05) sorghum (after Jan 06)	Access/ Preliminar
Parnaíba, PI ⁽³⁾	maize (after Aug 05) sorghum (after Jan 06)	Access/ Preliminar
Planaltina, DF ⁽³⁾	wheat	Access/ Preliminar
⁽¹⁾ Sites for Drought Phenotyping Network, ⁽²⁾ Sites of Excellence, ⁽³⁾ Sites of Reference		

⁽¹⁾Site-Specific for Drought Tolerance Phenotyping in Brazil Regions–Semester 1 2005

Brazil - UF	Site-Specific/Local	Longitude W	Latitude S	Altitude m
MG	Sete Lagoas	-44.2467	-19.4658	761
MG	Janaúba	-43.3089	-15.8025	533
GO	Planaltina	-47.6142	-15.4528	944
GO	Santo Antonio de Goiás	-49.1711	-16.2811	823
GO	Porangatu	-49.1486	-13.4408	396
PI	Teresina	-42.8019	-5.0892	72
PI	Parnaíba	-41.7767	-2.9047	5

⁽¹⁾ Sites for Drought Phenotyping Network, ⁽²⁾ Sites of Excellence, ⁽³⁾ Sites of Reference

⁽⁴⁾ Total per Year 1 and Semester 1, 2005 of Genetic Material (genotypes per crop specie) for all sites-specific in Brazil Regions								
Genotypes ⁽⁴⁾		Actions Strategies of Phenotyping ⁽⁵⁾						Notes ⁽⁶⁾
Crop Specie (total Year 1)	Type	Preliminar		Intermediate		Advanced		
		Year 1	Sem 1	Year 1	Sem 1	Year 1	Sem 1	
Maize (201)	Access	100	0 (221)					(After July 2005)
	Selected material	49	36					
	Elite	2	0 (2)					(After August 2005)
	CNPMS Maize Core Collection	25+25	50					
	CPG- Reference Sample	in definition	0					Partnership CPG and WPM Project
Sorghum (84)	Access	64	64					
	Selected material	0	0					
	Elite	20	0					
	CNPMS Sorghum Core Collection	0	0					
	CPG- Reference Sample	in definition	0					Partnership CPG and WPM Project
Rice (80)	Access	0	0					
	Selected material	40+ 40	80					
	Elite	0	0					
	CNPAF Rice Core Collection	0	0					
	CPG- Reference Sample	in definition	0					Partnership CPG and WPM Project
Wheat (95)	Access	200 fixed homozigotic	0					
	Selected material	95 greenhouse with 80 (field)	0 (80)					(After Aug 2005)
	Elite	0	0					
	CNPAF Rice Core Collection	0	0					
	CPG- Reference Sample	0	0					-
Common bean (144)	Access	0	0					
	Selected material	144	0					
	Elite	0	0					
	CNPAF Rice Core Collection	0	0					
	CPG- Reference Sample	0	0					-
Cowpea (120)	Access	60 (field)	0 (80)					(After August 2005)
	Selected material	60 (field)	0 (40)					(After August 2005)
	Elite	0	0					
	CPAMN Core Collection	0	0					
	CPG- Reference Sample	0	0					-

⁽⁴⁾ **Genetic Material** (access: germplasm, entries, etc.; or, **selected material**: access, variety, selected progenies per previous experiments under drought; or, **elite material**: characterized inbred lines (under intermediate-advanced phenotyping and/or commercial pre-released material-cultivar, etc.); or, **Core-collection material/crop specie**: selection controlled per water stress “per se” into each Core-collection; or, **CPG- Reference sample**: genetic material/crop specie selected from CPGeneration Collection (negotiable and according to leader SP1 and SP3).

⁽⁵⁾ **Phenotyping strategies** – genetic material characterization under selection for drought, with soil and plant drought tipification, and during critical phenological phase, per crop specie (**Preliminar phenotyping**: grouping per stress index based on phenotypic parameter and controlled environments and with water stress, typical to conventional breeding and with emphasis to access; Intermediate phenotyping: characterization using auxiliary descriptors techniques, and emphasis on selected material; **Advanced phenotyping**: specific characterization using especial plant organs (e.g., grain, root, leaf, etc.), based on mechanisms studies, and with emphasis to elite genetic material).

⁽⁶⁾ Blank space into this Table means data not yet worked during Semester 1, 2005.

Tangible outputs delivered

- In improvement of 02 Embrapa’s Phenotyping Center of Excellence for Drought Tolerance Studies (Sete Lagoas-MG, and Santo Antônio de Goiás-GO).
- In improvement of 04 Embrapa’s Phenotyping Sites of Reference (Sete Lagoas and Janaúba-MG; Santo Antônio de Goiás and Porangatu-GO). ⁽²⁾ **Note:** The sites Teresina and Parnaíba, PI are already been worked to carry out experiments after August 2005 (including maize, sorghum, rice and cowpea species).
- Mega- and micro-environment (regions and sites) are been evaluated and/or managed according to climatic, soil and crop species data (primary and/or secondary). ⁽³⁾
- In course, the definition of data base (climatic, soil and plant data set) and modelling (in partnership with CPG Whole Plant Modelling Project – CIRAD)

Technically, the project is later 2-3 months, because we had the financial resources delayed until last June 6, 2005. Because of this our acquisitions to installation of each site, as well as acquisitions of equipments and materials to work the factor water in soil and plant, it will be been adjusted after right now. Besides of this, the preliminary actions of our prepare of genetic material for each crop, and some experimental schedules were made during the last first semester of 2005. We are working fast now to buy the equipment and material, as well to carry out our experiments planned.

During next September will be planned the quantity and the genetic background for each crop specie (maize, sorghum and rice) to introduce from CPG reference sample aiming the partnership between CPG DPNetwork (Embrapa) and CPG WPModelling (Cirad) Projects, which will be carry out after Nov/2005-Fev/2006 (depending of each Site-Local), and legislation and procedures about Brazilian quarentine rules (3-6 months).

The improvement of 02 Center of Excellence (Sete Lagoas-MG and Santo Antonio de Goiás-GO) and 04 Reference Sites (Janaúba-MG, Porangatu-GO, Teresina-PI, and Parnaíba-PI) to Phenotyping for Drought Tolerance has been adequately worked, according the CPG DPNetwork Project. As reference to describe these protocols for “precision site-specific experimental and farming” is been prepared a technical boletin which will be concluded until November 2005. In this first year has been used the preliminary phenotyping strategy to water in different genotypes per each crop specie. In 2005, the CPG DPNetwork Project is evaluating genotypes, as follow: - Janaúba and Sete Lagoas sites [maize: (100+221 progenies; 36 inbred lines; 50 “maize core collection”; 02 contrasting inbred elite], and, [sorghum: 64 selected material]; - Teresina and Parnaíba sites – all after August [maize: (50 + 50 “maize core collection”); cowpea: 80 selected material]; - Porangatu and Santo Antonio de Goiás sites [rice: 80; common bean: 80, after August]; - Planaltina site – [wheat: 80, after August].

The financial resources were available on June 6, 2005. Since that date until today, we are preparing an adjustment of our workplan 2005 and also the new proforma invoice of each equipment and material, according planned in original project budget and letter of agreement between CIMMYT (acting on behalf of the donors to the Generation Challenge Program); Embrapa, and Supporting Research and Development Foundation (FAPED).

Description of Significant Travel:

- In: Whole Plant Modelling Project Workshop: Interaction with Drought Phenotyping Network Project, em Goiânia-GO, Brazil, 23-25 May 2005
- In: Drought Phenotyping Network Project Workshop, Brazilian Northeast Region - Teresina-PI, Brazil, 5 e 6/Julho/2005.
- In: Drought Phenotyping Network Project Workshop, Brazilian Northeast Region - Aracajú-SE, Brazil, 7/Julho/2005.
- Field evaluation and previous work to carry out experiments on Site of Janaúba-MG, Brazil, Junho/2005.

The Embrapa' sites constitute an net of research and development to the tropical agribusiness, including the small-scale farmers. Also, this network of R&D has provided condition of phenotyping genotypes for each crop specie, including cereals and legumes phenotyping for drought tolerance. Based on this the Embrapa's breeding programs have released a hundred different genotypes per specie well adapted to environment stress, specially to abiotic stress, like acid soils, and with tolerance to Al-toxicity, drought stress, N or P efficiency, and others environmental factors "per se".

For long term, the soil-water-atmosfera-plant relationships make our routine of R&D, and the Embrapa's teams have accumulated good experience in plant breeding and environmental factors management. However, this CPG DPNetwork Project is creating the opportunity of get scientific and technical gains to breeding programs and to genomics studies. At the moment, we are an enthusiastic and well trained team focusing on new approaches and good goals. Also, the current partnership with CPG-leadership and CPG Projects, e.g., like WPModelling Project lead our work to better possibility of scientific and technical goals.

The Embrapa's experimental sites to water are been prepared, after now, also to support the WPModelling Project partnership, including experiments after Nov-Dez/2005, and Jan-Fev/2006, in function of the climatic condition per each local. It is suggested to CPGeneration – SP1 leader to get some actions to previous choose of the genotypes per crop specie from CPG reference sample, aiming to provide seeds enough and available to quarentine period in Brazil (at least 3-6 mounths after come to Brazil). If necessary, the CPG DPNetwork Project leader can provide the technical procedures together the Embrapa Genetic Resources team, as well as together the Brazilian Ministry of Agriculture and Food Supply.

Milestones Completed (CGP Drought Phenotyping Network Project – 2005, First Semester)

Objectives	Activity	Semester per Year 1		Outputs Expected	Intermediate Report (Semester 1, 2005)
		Sem1	Sem2		
1) Phenotyping Center of Excellence for Drought Tolerance; and, Reference Centers.	1) Installation/enlargement of basic infrastructure in different locations for the forthcoming drought phenotyping research network ⁽¹⁾	x	x (to be continued)	1) To establish Embrapa's Phenotyping Center of Excellence and Reference for Drought Tolerance Studies, composed of a phenotyping network in six-eight experiment stations strategically located and equipped field and greenhouses at Embrapa Maize and Sorghum, Embrapa Rice and Bean, and, Embrapa Semi-Arid Tropics to coordinate and support drought research studies in Brazil.	1) In improvement of 02 Embrapa's Phenotyping Center of Excellence for Drought Tolerance Studies (Sete Lagoas-MG, and Santo Antônio de Goiás-GO). 2) In improvement of 04 Embrapa's Phenotyping Sites of Reference (Sete Lagoas and Janaúba-MG; Santo Antônio de Goiás and Porangatu-GO). ⁽²⁾ Note: The sites Teresina and Parnaíba, PI are already been worked to carry out experiments after August 2005 (including maize, sorghum, rice and cowpea species).
2) Protocols for drought tolerance evaluation in maize, sorghum, rice, wheat, common bean, and cowpea	2) Definition of the ideal conditions for soil-water, climate and plant to induce water stress for cereals (maize, sorghum, rice, wheat) and legumes (common bean and cowpea).	x	x (to be continued)	2) To generate knowledge, in terms of protocols, data collection and phenotypic characterization for drought tolerance, to be useful to breeding programs and also to be transferred to tropical Latin America and African countries.	1) Mega- and micro-environment (regions and sites) are been evaluated and/or managed according to climatic, soil and crop species data (primary and/or secondary). ⁽³⁾
3) Evaluation and characterization genotypic response to drought conducted under adequate environment control for each specific crop involved.	5) Creation, organization and management of a data bank will all genotypes and environment conditions, from the very begin of the project, aiming the characterization and the genotypes development. 6) Monitoring the environment: climatic and soil data set.	x	x (to be continued)	2) Generated knowledge, in terms of protocols	1) In course, the definition of data base (climatic, soil and plant data set) and modelling (in partnership with CPG Whole Plant Modelling Project – CIRAD)
⁽¹⁾ Sites for Drought Phenotyping Network, ⁽²⁾ Sites of Excellence, ⁽³⁾ Sites of Reference					

7. Whole Plant Physiology Modeling of Drought Tolerance in Cereals

Principal Investigator:

Marcel de Raïssac, Cirad

Collaborating Scientists:

Delphine Luquet, Cirad-Agropolis

François Tardieu, Cirad-INRA

Renée Lafitte, IRRI

Jean-Marcel Ribaut, CIMMYT

M Dingkuhn, Cirad-Agropolis

M de Raïssac, Cirad-Agropolis

JC Combres, Cirad-Agropolis

C Welker, Cirad-INRA

Scott Chapman, CSIRO. Australia

Graeme Hammer, University of Queensland

B. Bouman, IRRI

M Bänzinger, CIMMYT

M Reynolds, CIMMYT

R Trethowan, CIMMYT

Eva Weltzien, ICRISAT

Frederico Duraes, EMBRAPA

Mark Cooper, Pioneer

MID-YEAR REPORT

Following some defection in the participants initially included in the project (see § “Deviations from the work plan”), new contacts have been made with Embrapa in order to settle a strong interaction between two GCP SP1 commissioned projects : “Whole Plant Modelling (WPM)” and “Drought Phenotyping Network (DPN)”. The rationale is that both projects will mutually benefit each from the other : WPM by testing materials and models on a unique reference experimental network dedicated to drought tolerance ; DPN by exploring at a small scale a phenotyping approach supported by modeling. The initiative received very positive reaction from Embrapa.

On these new basis, the launching workshop for WPM project has been organized by Embrapa Rice and Bean in Goiania, from May 23th to 25th. 18 participants from different Embrapa centers, from Inra, Cirad and Csiro met during three days to share their own experiences on phenotyping and modeling. Discussions on days 2 and 3 led to the paper “Discussion and final recommendations of the workshop” (S Chapman, M Dingkuhn, M de Raïssac, June 2005) that fixes the modified two year work plan for components 1 and 2 of the project, assuming that Embrapa was not directly involved in component 3 -which activities are to start in 2006-.

- Environment characterization will be carried out using crop model SarraH, from Cirad. Simple meteorological models could be used to complete study.
- A post-doc has been recruited by Embrapa (Alexandre Bryan Heinemann). He is currently at the Rice and Bean Center and is collecting all meteorological and soil data required to run SarraH model for TPE definition (Target Population of Environment) in the studied regions (Minas Gerais, Goias, Tocantins, Piaui). From September 2005 to September 2006, he will be received at the Cirad/ECOTROP laboratory in Montpellier to carry out this TPE definition for upland rice, sorghum and maize.

- Camilo Andrade, agronomist from Embrapa Maize and Sorghum will be trained at Montpellier on SarraH model in November 2005.
- Edson Bastos from Embrapa Mid North will be trained on SarraH model in November 2006, with special focus on building up specific modules for cowpea and legumes use.
- Complementary experiments will be conducted on the Drought Phenotyping Network, for maize and upland rice in seven sites on wet and dry seasons, on a reduced sub sample of the GCP reference collection, in order to improve phenotyping ability of available models, for subsequent applications to other materials from GCP collection.
- Experiments will be conducted by Embrapa, with support from Marcel de Raïssac and Delphine Luquet, for model input variable measurements.
- Alexandre Bryan Heinemann and Camilo Andrade will receive training on Apsim at Csiro, Australia, in beginning of 2006.
- on experiment designs are in progress between all participants from Csiro, UQ, Inra and Cirad. Call for post-doc will be launched in September 2005.

Tangible outputs delivered:

- Post doc recruited
- Definitive workplan established (see document above-mentioned)
- Launching workshop achieved
- Scientific community constituted
- Data collected for model running.

Initially, the project proposal was based on collaborative work with :

- Inra, Cirad, Csiro and UQ for modeling part
- Embrapa on maize, sorghum and upland rice in Brazil
- Irri on upland rice in Philippines
- Cimmyt on maize in Africa.

Unfortunately, some professional evolutions at Irri and Cimmyt prevented their scientists to participate. That led the project to reinforce planned collaboration with Embrapa, with gain in consistency for the experimental part. At the scientific level, the project is not modified, testing and improving methodologies for model supported phenotyping on cereals, with the objective to enlarge crops, partners and geographical regions at the term of the project. There are no financial consequences of this evolution and the project did not suffer any delay in progress.

In order to complete two experimental years before the tenue of the opened final workshop, the project has been extended up to 31 May 2007.

8. Population Structure, Phenotypic Information, and Association Studies in Long-Generation Crops

Principal Investigators:

M Carmen de Vicente, IPGRI
 Martin Fregene, CIAT
 Luc Baudoin, CIRAD
 Kodjo Tomekpe, CARBAP, Cameroon
 Merideth Bonierbale, CIP
 Jean-Louis Noyer, CIRAD

Co-Principal Investigators:

Toby Hodgkin, IPGRI

Jean Christophe Glaszmann, CIRAD

Marc Ghislain, CIP

Reinhard Simon, CIP

Vincent Lebot, VARTC, Vanuatu

Collaborating Scientist:

Nicolas Roux, IPGRI-INIBAP

MID-YEAR REPORT

As an attempt to integrate association analyses (in a broad sense = comparison between molecular and phenotypic diversity) in the course of ongoing characterization and breeding activities, similar exercises of LD assessment have been undertaken for additional cases where accurate phenotypic data are available for materials amenable to LD mapping. These are the cassava breeding program at CIAT, the potato breeding program at CIP, the banana breeding program at CARBAP, Cameroon, the coconut breeding program at VARTC, Vanuatu, as well as the yam (*Dioscorea alata*) germplasm evaluation program at VARTC, Vanuatu.

In the former three species, the possibilities for choosing the best materials have been analysed. Numerous morphoagronomic data have been assembled and analysed in various germplasm compartments. The preference will be for materials derived from breeding activities, that represent a large well characterized phenotypic diversity, preferably have known pedigrees and represent related materials that have a limited number of meioses since their main foundation events. Various options still exist; the assessment of the level of LD will be of primary importance.

The case is more clearly defined in coconut. The materials consist of 200 trees representing four generations of breeding materials. A total of 219 trees have been sampled in VARTC and DNA has been extracted in CIRAD, Montpellier, France. They will be analysed by a scientist from Sri Lanka during a training session in Montpellier in September-November. A total of 31 loci will be surveyed, including 13 international reference markers and 9 couples of linked markers. In addition, a coconut breeder from Vanuatu will visit CIRAD in October for a three week training session on DNA extraction, principles and applications of molecular breeding and data management.

For yam, the idea is that the insular history may have involved bottlenecks that have established strong LD in the well characterized populations of Vanuatu; a mapping exercise is undertaken to identify linked markers among the already used AFLPs, for quick assessment of LD. Phenotypic data are scored for 331 *D. alata* accessions from the Vanuatu collection since 2000 (SPYN project 1999-2003). From these 331 accessions 117 were characterised through an AFLP analysis where 124 loci were retained from 5 primer pairs' patterns. Phenotypic and molecular data analyses were achieved in early 2005 and the Vanuatu collection is now re-organised essentially by the suppression of duplicates. Two hundred and twenty-one remaining accessions have been planted again at the end of July 2005 to continue the phenotypic evaluation and will be complemented with new accessions obtained from running collects in Vanuatu islands. Twenty-six descriptors and used to describe the Vanuatu collection, 9 for the tubers, 14 for the aerials traits and 3 for the resistance to pest and diseases. Fifty-one SSR were screened on a small interspecific sample and 16 were retained for the capacity to exhibit useful patterns for all cultivated species belonging to the *Dioscorea* genus. DNA sequences are available on the EMBL Genbank

since July 2005 (Tostain et al, submitted). The 35 remaining SSR are currently screened for their ability to reveal useful patterns inside of the *D. alata* species. DNAs of the 221 accessions still present in the phenotypic evaluation process are conserved in Cirad laboratories.

Attempts to germinate seeds and grow seedlings derived from controlled hybridizations following protocols developed at ICAR are planned from August.

Tangible outputs delivered:
Numerous phenotypic data have been organized

Deviations from the work plan:
Choosing the populations within the breeding programs is taking longer than expected

SP2 COMMISSIONED GRANTS

9. Systematic Evaluation of Rice Mutant Collections for Conditional Phenotypes with Emphasis on Stress Tolerance

Principal Investigator:

Andy Pereira, WUR

Co-Principal Investigators:

Hirohiko Hirochika, NIAS, Japan

Hei Leung, IRRI

Emmanuel Guiderdoni, Agropolis

Mathias Lorieux, IRD/CIAT

Manabu Ishitani, CIAT

Tiegang Lu, CAAS

Qifa Zhang, HAU, China

Deming Jin, HAU, China

Collaborating Scientists:

Members of the IRFGC

Gyn An Pohang, University of Science and Technology, Korea

Srinivasan Ramachandran, Temasek Lifesciences Laboratory, Singapore

Narayana Upadhyaya, CSIRO, Australia

Venkatesan Sundaresan, UC Davis

MID-YEAR REPORT

Compilation of a stress associated gene (SAG) database for comparative genomics: A number of approaches were taken to categorize 'stress associated genes'. A number of rice microarray publications describing lists of rice genes induced by various abiotic stresses and overexpression of stress resistance genes, were used to compile a publicly available SAG list of about 400 genes. Another list was generated of 116 disease induced/related genes using knowledge and discussions at CIAT/IRD/Agropolis. This was supplemented by using 22K rice microarray experiments to identify more genes induced by BTH-induced defense. A third approach was taken to identify rice orthologs of 48 Arabidopsis genes identified as stress associated using the criteria of being induced in a stress resistant genotype obtained by overexpression of stress resistant genes. In addition a list was made of 16 rice orthologs of Arabidopsis genes known to be involved in RNAi associated mechanisms (including stress).

Identify knockout mutants in these stress associated rice genes: The OryGenes DB (<http://orygenesdb.cirad.fr>) has about 56,000 publicly available insertion sequences tagged by FST sequence on the rice genome sequence as annotated by TIGR, as shown below (please see author for figure). This DB was used to search for SAG knockout inserts and provided about 22% (61/298; 20/58) insertion coverage. These insertion lines are now being categorized for phenotyping at CIAT.

Generate gain-of-function genotypes: To generate stress resistant genotypes using overexpression, constructs were initiated for about 20 rice SAG genes (identified as orthologs of Arabidopsis genes) and about 5 are being transformed. Contributory projects generated drought stress resistant rice genotypes (overexpression of the Arabidopsis *SHINE* gene) that were tested in the greenhouse and are now being tested for physiological parameters, and will be tested under field/GH conditions at China.

Microarray analysis to identify genes involved in plant activator-induced defence: The 'plant activator' Benzothiadiazole (BTH) protect plants from diseases by 'priming', enhancing resistance in rice against *Rhizoctonia solani* and *Magnaporthe grisea* (Fig. 1). A 22k rice oligo-DNA array was used to identify genes induced by BTH: with 260 and 104 genes showed differential expression of >1.5-fold and >2-fold. The induced genes included defense related genes, four WRKY-family transcription factor genes (Fig. 2) and one NAC-family gene. Mutants of two of the five transcription factor genes were identified using available FST databases and will be screened against *Rhizoctonia solani* and *Magnaporthe grisea*.

Forward screen drought stress: At CAAS and HAU in China, the drought stress screen is being done using a collection of T-DNA tagged lines. At NIAS an *in vitro* based screen revealed 3 salt-sensitive mutants by screening 3,000 lines of *Tos17*-induced mutants. In a mutant analyzed that was not tagged by *Tos17*, a mapbased cloning strategy was initiated. As shown in Fig. 3, the mutant, named *rss2*, showed severe dwarf phenotype when 150 mM NaCl was added in the medium. By bulked segregant analysis, *rss2* was mapped at the middle of the chr. 11. By using 1,500 F2 plants, *rss2* was mapped within 800 kb region. This region will be narrowed down by mapping using new markers.

Tangible outputs delivered:

- An extensive SAG database is compiled using a variety of criteria from available resources. This comprises of the following:
- 400 abiotic stress associated genes identified from publications
- 116 disease stress associated genes obtained from publications
- 260 (>1.5 fold) and 104 (>2 fold) BTH induced disease stress associated genes
- 48 putative rice orthologs of Arabidopsis genes revealed in a drought stress regulon
- 16 putative rice orthologs of Arabidopsis stress related RNAi mechanism genes

A novel strategy was developed to identify genes responsible for pathogen resistance, using BTH induction. The availability of this SAG list opened up an efficient way to identify these candidate genes, and this research is being followed up by validation of the SAG candidates. Screening for phenotypes will be carried out at the most convenient location for minimizing transfer of transgenic plants.

10. Collection, Distribution, Phenotyping, and Genotyping Directed towards Utilisation of Existing Wheat Genetic Stocks to Enhance

Tolerance/Resistance of Wheat Cultivars to Abiotic and Biotic Stresses with Emphasis on Drought

Principal Investigators:

Maarten van Ginkel, CIMMYT

Hans-Joachim Braun, CIMMYT

Collaborating Scientists:

Peter Langridge, University of Adelaide

Xueyong Zhang, CAAS

Marion Röder, Gatersleben, Germany

Tetsuo Sasakuma, Kihara Institute for Biological Research, Japan

Hitashi Tsujimoto, Tottori University, Japan

Masahiro Kishii, CIMMYT

John Snape, John Innes Centre

Jorge Dubcovsky, University of California Davis

Bikram Gill, Kansas State University

Bernd Friebe, Kansas State University

Perry Gustafson, USDA-ARS

Adam Lukaszewski, University California Riverside

Mark Sorrells, Cornell University

MID-YEAR REPORT

A main output of the project was the workshop organized to establish the network and mechanisms involving institutions worldwide that will provide wheat genetic stocks for use in gene discovery. Excerpts from the discussion and outputs of this workshop are presented here as the project update report.

Generation Challenge Program (SP2) Wheat Genetic Stocks Utilization Workshop 5-7 April 2005, El Batan, Mexico

Participants: Andreas Börner, Hans-Joachim Braun, Kim Campbell, Nick Collins, Jonathan Crouch, John Dixon, Bernd Friebe, Perry Gustafson, A. Mujeeb Kazi, Masahiro Kishii, Robert Koebner, Adam Lukaszewski, Jennifer Nelson, Tom Payne, Tatiana Pshenichnikova, Tetsuo Sasakuma, Mark Sorrells, Shawn Sullivan, Richard Trethowan, Hisashi Tsujimoto, Jan Valkoun, Manilal William, Xueyong Zhang

Project Overview¹: This project proposes the establishment of a consortium existing of a central site with an activities network of key institutions worldwide that will provide the genetic stocks for use in gene discovery through comparative analysis. The consortium aims to:

- Conserve existing relevant wheat genetic stocks;
- Make these available for agronomic and molecular characterization that will lay the foundation for association genetics across diverse wheat germplasm;
- Multiply and on request distribute seed of these stocks;
- Initiate phenotyping for drought tolerance traits.

¹ The full project proposal can be found at:

http://www.generationcp.org/vw/Download/Commissioned_Research_2005/10_SP2_Wheat_genetic_stock.pdf

Wheat Genetic Stocks amenable to shorter-term, with good immediate impact prospects

- Pavon centric rye translocations
- Primary genepool derivatives
- Synthetics, and derivatives
- Saratovskaya 29 populations
- Mapping population parents (inventory needed)

Wheat Genetic Stocks more amenable to medium to longer-term (7+ years) utility

- Alien sources, secondary and tertiary genepools
- Addition lines < Amphiploids (e.g. triticale)

Wheat Genetic Stock selection criteria

1. Genetically stable
2. Adapted background
3. Available without restrictions
4. Seed availability
5. Possible phenotypic, genetic characterization
6. Trait(s) of interest → (will determine) phenotyping procedures → (will determine) amounts of seed required
7. Short-term guaranteed outputs (SP3, 3-5 years) + longer-term intermediary products (SP2, 7-10 years)

What is drought tolerance? (Project phenotypic characterization.)

- *Per se*
- Avoidance
- Root health (nematodes, rots)
- Micronutrient deficiency / toxicity (Bo, Zn, Al, Cu)
- Root biomass / architecture
- Photothermal response
- Canopy temperature depression
- Salinity

A preliminary list of wheat genetic stock candidates available for distribution and characterization by the workshop participants was drafted (appended), with over 3,300 lines identified. Project participants will be requested to provide an itemized list of available, candidate wheat genetic stocks shortly after the conclusion of this workshop.

Initial seed stock reserves of 30-50g per accession are needed to begin further multiplication and phenotypic characterization. It was agreed that stocks currently available, given this 30-50g requirement, will be sent to CIMMYT immediately. Perry Gustafson will coordinate collection of stocks in the USA allowing a consolidated shipment to CIMMYT.

Seed stocks that may require multiplication to achieve sufficient seed supplies (30-50g) will be done by each contributing partner, with a target of May 2006 to seed the required stock seed supply to CIMMYT.

Subject to revision, initial seed multiplication cost estimates were (US\$):

John Innes Centre	\$1000
ICARDA	free
Cornell University	free
University of Missouri	free

Gartersleiben	free (but some funds may be required by Spanish partners)
Australia	\$500
China	free
Japan	\$1000 (amphiploid multiplication & verification)
Kansas State University	\$2000
Univ. California-Riverside	\$2000
CIMMYT	free

Genetic stocks with sufficient seed available (30-50g) can be sent to CIMMYT-Mexico (or via Perry Gustafson, see below):

- USA germplasm to be sent to Perry (USDA) for redistribution to Mexico
- All seed to be sent to Mexico first. Winter wheat germplasm will then be transshipped to the TURKEY/CIMMYT/ICARDA and ICARDA-Syria programs.
- Generation CP subcontracts will be distributed to cooperators to formalize participation in the project

“How to send seed to CIMMYT-Mexico” can be found at:

http://www.cimmyt.org/english/wps/obtain_seed/shl-sendseedtomexico.htm

Wheat Genetic Stock collections that maybe vulnerable or “at risk”

- Vavilov Institute (VIR)
- University of Stellenbosch, South Africa
- Bulgaria (addressed via the EWAC)

Additional possible partners for the project

- George Fedok, Canada
- Andre Comeau, Canada
- Ludihana, India
- ICAR, Delhi, India
- Punjab Agricultural University, India
- Nordic Gene Bank
- Argentina, Castilar (Buenas Ares) “Sr. Prina”
- Turkish Genebank, Menemen, Izmir
- Moshi Feldman, Israel
- PBI, Cobbitty, Australia (Ken Shepard, Bob McIntosh)
- Bob Metzger, Oregon State Univ., USA → \$2000 seed multiplication
- University of Nebraska, USA (Rosalind Morris, Bob Metzger)
- Steven Shau, Leonard Joppa. Langdon series (needs to be contacted)

Avenues available to communicate about the project

- Website, Announcement of meeting, List of germplasm,
- GrainGenes announcement
- AWN
- Poster of the 7th IWC, ASA, EWAC,
- Brochure

A Steering Committee consisting of all institutes present at this meeting was established (list appended), with Hans Braun volunteering to Chair the SC. Meetings of the SC will

be conducted electronically, via email. The SC will monitor implementation of Phase 1 activities, and draft a Phase 2 project proposal.

1. Other business:

Bernd Friebe noted that the Wheat Genetic Resources Center, Kansas State University holds a number of potentially vulnerable stocks, primarily due to seed viability. Workshop participants endorsed the value of these stocks, and the following participants volunteered to assist with seed multiplication and chromosomal verification of these stocks:

Material	No. entries	Multiplication	Verification
Amphiploids	111	KSU	KSU
Addition / Telosomic Addition	105	A. Mujeeb-Kazi, P. Gustafson	P. Gustafson, A. Mujeeb-Kazi, H. Tsujimoto
Substitution	143	CIMMYT, ICARDA	H. Tsujimoto, X. Zhang
Translocation	74	CIMMYT, ICARDA	H. Tsujimoto, X. Zhang

Preliminary list of wheat genetic stock candidates available for distribution and characterization by the project

Contributing Institution	Description	Number in the set	Back-ground
CIMMYT/Kishii	Synthetic D genome	250	
CIMMYT/Kishii	Synthetic DW parents	50	
CIMMYT/Kishii	Synthetic A genome	150	
CIMMYT/Kishii	Drought DH (synth derivatives)	500	
CIMMYT	Tcl		
Cornell/Sorrells	Mapping pop. Parents	9	
Cornell/Sorrells	Mapping pop. Parents	6	
Cornell/Sorrells	Mapping pop. Parents	10	
Waite/	Mapping pop. Parents	14	
Waite/	Wheat/Th. Intermedium 7A recomb lines		
Waite/	Wheat/Rye 1R recomb	20	
Waite/	Wheat/Rye 6R		
Waite/	Imperial Rye/wheat transl		
JIC/Koebner	Mapping pop. Parents	6	
JIC/Koebner	H. chilense/CS transl	1	CS Hobbit
JIC/Koebner	H. chilense/CS 5D sub	2	Moulin
JIC/Koebner	CS/H. vulgare subs	20	CS
JIC/Koebner	CS/rye subs	7	CS
JIC/Koebner	CS/rye trans	3	CS
JIC/Koebner	Holdfast/rye sub	11	Holdfast
JIC/Koebner	CS/S. montanum subs	8	CS
JIC/Koebner	CS/Th. Bessarabicum subs	8	CS
JIC/Koebner	CS/Th. Bessarabicum transl	1	CS

JIC/Koebner	Vilmorin/Th. Intermed. Sub	3	Vilmorin
JIC/Koebner	CS/T. urartu subs	3	CS
JIC/Koebner	Wheat/rye transl	10	various
JIC/Koebner	Ae. Bicornis subs	5	Holdfast
JIC/Koebner	Ae. Comosa sub 2A	5	various
JIC/Koebner	Ae. Comosa sub 6A	1	CS
JIC/Koebner	Ae. Comosa 4D transl	1	CS
JIC/Koebner	Ae. Comosa 2D transl	2	Hobbit
JIC/Koebner	Ae. Sharonensis	1	Brigand
JIC/Koebner	Ae. Sharonensis	3	CS
JIC/Koebner	Ae. Sharonensis	3	CS
JIC/Koebner	Ae. Umbell.	3	CS
JIC/Koebner	Ae. Univ.	3	CS
JIC/Koebner	Ae. Univ.	7	CS
JIC/Koebner	Ae. Variabilis	7	CS
JIC/Koebner	H. chilensis subs.	2	CS
JIC/Koebner	H. chilensis subs.	3	CS
JIC/Koebner	H. chilensis subs.	3	CS
JIC/Koebner	H. chilensis subs.	3	CS
KOMUGI/Sasakuma	Synthetic progenies	20	
KOMUGI/Sasakuma	RILs	4	
KOMUGI/Sasakuma	Mutant, NIL	10	
KOMUGI/Sasakuma	Cytoplasmic sub	10	
KOMUGI/Sasakuma	Many DH		
KOMUGI/Sasakuma	Many landraces		
KOMUGI/Sasakuma	6x Tcl	18	
KOMUGI/Sasakuma	Amphiploids	20	
EWAC	Saratovskaya subst	23	
EWAC	CS/Tauschii subst	9	
EWAC	CS/Tauschii introgr	85	
EWAC	Saratovskaya 5A subst	50	
EWAC	Diamant 5A subst	10	
EWAC	Tcl (Eastern Europe)	100	
EWAC	Tritordeum	250	
EWAC	Novosibirskaya isogenic	70	
WGRC	WGRC germplasm releases	50	
WGRC	Wheat alien subs	30	
WGRC	Alien transl	40	
WGRC	Synthetic progenies	300	
ICARDA	DW population parents	10	
ICARDA	Synthetic progenies	2	
ICARDA	DW introgr	30	
CAAS	Thinopyrum introgr	10	
CAAS	Synthetic progenies	10	
USDA/Gustafson	Tcl	200	
			Mapping
USDA/Gustafson	Wheat	10	pops
USDA/Gustafson	Wheat	400	Sears coll.
USDA/Gustafson	Inbred ryes	48	Sears coll.
USDA/Gustafson	Wheat/rye subs	40	Sears coll.
USDA/Gustafson	Sears subs	100	Sears coll.

UCR/Lukaszewski	Wheat-Rye subs and trans	50	Pavon
UCR/Lukaszewski	Assorted subs, transl (7S, 7E)	10	Pavon
UCR/Lukaszewski	Recomb of 7E in Pavon/Opata	30	Pavon
UCR/Lukaszewski	Assorted alien transl	15	assorted
UCR/Lukaszewski	1b subs in Pavon	15	Pavon
UCR/Lukaszewski	5+10 introgr fixed 1RS	20	DW
UCR/Lukaszewski	D subs, transl in Rhino	30	tcl
UCR/Lukaszewski	D subs, transl in Presto	50	tcl
Joppa/NDSU	Langdon series	14	
TOTAL		3337	

Additional information (not presented) was collected on: Spring or winter growth habit; Verification of Freedom to Distribute; Estimated current seed supply; Availability of phenotypic characterization data; Availability of genetic characterization data.

11. Legume Mutant Resource Development

Principal Investigator:

Matthew Blair, CIAT

MID-YEAR REPORT

This project has initiated the development of mutant resources for common bean as a model legume given that fewer mutant stocks have been created for grain legumes available compared to those for *Arabidopsis* (McCallum et al., 2000; Colbert et al., 2001; Greene et al., 2004), forage legumes (*Lotus* principally) (Perry et al., 2003) and the cereals (Till et al. 2004; Henikoff et al., 2004). We have made progress in the following five points:

- 1) Seed multiplication: We multiplied seed for the target TILLING genotype BAT93 over the course of two seasons in Colombia to increase the amount of seed to a total of over 20 kg. This multiplication was based on type specimens which were purified and fingerprinted with 150 microsatellite (SSR) markers to ensure homogeneity of the mutagenized genotypes and quality control during TILLING. The large-scale seed multiplication was done in order to ensure a seed supply for the mutagenesis work (next point).
- 2) Chemical mutagenesis: 9 ethyl methane-sulfonate (EMS) was used to create a M1 population of 300 fertile individuals in Univ. of Geneva which was shipped to CIAT for multiplication. In addition Univ. of Geneva has initiated further mutagenesis with an additional 4 kg. of seed that we prepared at CIAT as described above and are using an optimum EMS concentration that was determined by Pankhurst et al. (2003).
- 3) Generation advance: in an enclosed screenhouse, we have advanced from the M1 to the M2 generation and produced a total of 800 M2 progeny from 100 M1 mutant families (8 individuals per family).
- 4) Phenotypic screening: the M1:2 plants have been screened for phenotypic differences compared to the non-mutated control genotype, BAT93. Phenotypic mutants (dwarfing, leaf fasciation, leaf variegation, spindly growth, etc.) have been documented and photographed.

- 5) DNA extraction: a miniprep DNA extraction technique has been developed to extract DNA from 400 of the mutant M1:2 plants. The DNA quantity and quality has been checked in preparation for plans for pooling of individuals from different mutant families or pooling of the individual plants analyzed in the phenotyping described above.

A well-developed mutant stock, particularly in genotypes of common bean (*Phaseolus vulgaris*), a simple diploid species with a small genome (650 Mb), will serve the broad community involved in tropical legume improvement aiding gene-discovery both in common bean, the most widely consumed grain legume for human consumption and a major protein and mineral source in East Africa and Latin America, as well as in two tropical legume relatives: cowpea (*Vigna unguiculata*), a food crop important in West African agriculture and soybean (*Glycine max*) a major industrial and feedstock crop around the world. Mutant stocks in common bean will allow researchers to conduct both forward (systematic phenotypic screening) and reverse genetics (TILLING or Targeting Induced Local Lesions In Genomes) experiments aimed at understanding the genes involved in abiotic and biotic stress tolerance as well as those genes involved in biological nitrogen fixation in the tropical legumes. Mutations will be sought in common bean genes that have been isolated at CIAT and shown to be associated with drought tolerance. The phenotypic effect of these mutations will be analyzed as a proof of concept for the value of the mutant stocks generated by this project.

Tangible outputs delivered:

- pure seed has been produced at CIAT for BAT93 and shipped to Univ. of Geneva for mutagenesis
- mutagenesis protocol has been worked out by collaborators at Univ. of Geneva and seed of M1 families sent to CIAT

The project has begun with a good start and we are mostly on schedule. We have produced an initial number of mutants for troubleshooting all the steps in the DNA extraction and pooling process but will need to increase the number of EMS mutants to start TILLING (targeted induced local lesions in genomes) protocols. We have not yet decided whether to implement mutagenesis of a second target genotype, BAT477 although this has been a focus of the drought research we are conducting in parallel with the mutagenesis work. We plan to search for allelic series of mutations and the plants that contain them for the drought responsive genes identified by M. Ishitani probably targeting the three homologues of the DREB gene identified by his group. One potential bottleneck is the extraction of CEL I endonuclease or the troubleshooting of the Li-Cor evaluation of mutants (McCallum et al., 2000).

12. A Saturated Potato Mutant Population for Functional Genomics among Solanaceae and Tuber Crops

Principal Investigator:

Meredith Bonierbale, CIP

Marc Ghislain, CIP

Collaborating Scientists:

Glenn Bryan, Scottish Crop Research Institute

Robbie Waugh, Scottish Crop Research Institute

Dani Zamir, The Hebrew University of Jerusalem

MID-YEAR REPORT

Out of the 8 ver accessions available at CIP (see table1), we germinated seeds (presumably from selfing) and isolated DNA from 50 plants. We used 7 AFLP combinations (Eco/Mse - chosen based on polymorphism in previous mapping work) and found that plants from 2 accessions (SHGRF 4019 and TRHRG 161) were indeed resulting from strict selfing and with the least heterozygous (TRHRG 161) with 7.5% of polymorphic AFLP markers. Assuming this is still far too high for our purpose, we have chosen 5 plants with minimum polymorphic loci to generate new selfed seeds. These are scheduled to be available early next year.

Seeds of the ‘CPC54’ *S. verrucosum* accession have been sent to CIP.

Table 1: Accessions of *S. verrucosum* available to the project for developing potato mutant collection (‘CPC’ accession is available from SCRI and ‘763 ___-series from CIP).

CIP number	PI number	Collector number	Species	FAO	In vitro/seeds
CPC54	-	BLS 5628	ver	not known	seeds
763029	275256	HAW 1528	ver	y	seeds
-	116163	-	ver	not known	In vitro & seeds
763779	558482	SHGRF 4019	ver	y	seeds
763855	498010	TRHRG 23	ver	y	seeds
763857	498061	TRHRG 161	ver	y	seeds
763858	545747	TRHRG 197	ver	y	seeds
763863	310966	UGN 1289	ver	y	seeds
-	-	TAR 91F	ver	y	seeds

Tangible outputs delivered:

Least heterozygous of CIP *S. verrucosum* accession identified.

Deviations from the work plan:

Delay in fund transfer from GCP to CIP (only effective on the fourth month the year) as well as funds from CIP to SCRI and University of Hebrew (not yet effective) made us losing the first half of the first year of the work plan.

Secondly, the transfer of homozygous seeds of the accession CPC54 from SCRI to CIP was extremely slow due to ineffective phytosanitary process between Scotland and Peru.

The consequences are that M0 seed bulking will not happen in 2005 but early in 2006.

13. Crop Gene Expression Profiles and Stress-gene Arrays

Principal Investigator:

Tiegang Lu, CAAS

Co-Principal Investigators:

Guozhen Liu, Beijing Genomics Institute (BGI), China

Shoshi Kikuchi, NIAS

Manabu Ishitani, CIAT

MID-YEAR REPORT

Three sets of experiments were at different stages of implementation with the common theme of identifying stress response genes using two rice gene chip platforms (Agilent 22K gene chip and BGI 60K gene chip). Both rice and wheat RNAs were used to test the feasibility as well as limitation of heterologous probing. The experiments are expected to provide a rich dataset generated from a common array platform. From these experiments, a set of experimentally-determined stress-related genes will be selected to conduct a pilot study of stress gene arrays.

Drought is a yield-limiting factor in rice and wheat production under rainfed environments. Wheat is one of the more drought-tolerant cereals, providing a useful platform to understand tolerance mechanisms in rice. We undertook a comparative gene expression study of drought responsiveness in rice and wheat by making use of the genome sequence information and tools available from rice. Two rice varieties (IR64 and Apo) and two wheat varieties (Seri and Weebil) with contrasting levels of drought tolerance were subject to a similar drought stress regime. Water stress was imposed on pot-grown plants at 33 days after sowing and leaf samples were collected from both control and stressed plants when available soil moisture fell below 30%. Expression profiling was done using the BGI-Ricechip-60K. We detected 29494 signals in rice-rice hybridization and 19053 in wheat-rice hybridization (>60% of that in homologous hybridization). Among the significantly expressed genes, 77 were up-regulated and 153 were down-regulated in both IR64 and Apo. Of the up-regulated genes, 156 were specific to Apo (tolerant) and 29 genes were specific to IR64 (susceptible). Of the under-expressed genes, 269 were specific to Apo and 140 were specific to IR64. In wheat, 52 genes were found to be drought responsive, of which 12 genes showed a common expression pattern in both rice and wheat. Five genes were up-regulated in wheat but down-regulated in rice during drought stress. This study demonstrates the feasibility of using a common gene chip to identify a relatively small set of genes involved in drought response in two related cereals. Experiments using 22K rice Oligoarray (Agilent Technologies) are in progress to validate these results. The confirmed gene sets may provide potential markers for associating gene function to drought tolerance phenotypes in rice and wheat.

This experiment was built upon the initial success of Experiment 1 where the 60K BGI rice chip was found to be useful for expression analysis of wheat. In this experiment, CAAS selected wheat genotypes with known tolerance to drought and salinity. As an initial test of the chip system, the genotypes were subjected to artificial stress conditions (dehydration under PEG-6000 for 3 time points within 24 hr and salt-stress under 1% sodium chloride at 1 and 6 hr). RNA from the relevant genotypes under different stress treatments were isolated by CAAS and provided to BGI for hybridization. As of June, the hybridization experiments were done but data analysis was not yet completed.

A visit was made by CIAT (M. Ishitani) to NAIS in March 2005 to discuss experimental plan. The team decided to complement the other experiments by focusing on analysis of the gene expression profile of stay-green rice genotype, Fedearroz 50 under drought-stress condition. Two different developmental stages were tested under water-limited conditions to see whether the stay-green phenotype will be beneficial for drought tolerance. The test was conducted using cylinder filled by soil with different soil water capacities (100, 75, 50 and 25%) in the greenhouse. Physiological analysis was also conducted such as photosynthetic activity. CIAT is responsible for providing RNA samples for the microarray analysis with rice 22 K oligoarray (Agilent Technologies) at NIAS. Work is in progress at the time of writing.

Tangible outputs delivered:

Muthurajan R, Praba ML, Lafitte RH, Satoh K, Kikuchi S, Gan Q, Liu G, Leung H. 2005. Comparative analysis of gene expression in rice and wheat during drought stress at vegetative stage. Abstract submitted to the Fifth Rice Genetics/Third Rice Functional Genomics Symposium, to be held November 19-23, 2005, Manila, Philippines.

No significant deviation from original workplan. The work is built upon the collaboration initiated in GCP Year 1 commissioned work. The new partnership initiated with Beijing Genomics Institute has broadened the group's ability to assess different gene chip platforms. Members of the group interact primarily through email so far. All parties desire more interaction among the group. We expect this will happen when the group has generated sufficient data from respective experiments to enable comparative analysis. A face-to-face group meeting would be desirable in the later half of 2005.

14a. Stress Response-enriched EST Resources for Targeted Species— Pearl Millet

Principal Investigator:

Tom Hash, ICRISAT

Co-Principal Investigators:

MK Reddy, International Center for Genetic Engineering and Biotechnology

Arjula Reddy, Central University of Hyderabad, India

MID-YEAR REPORT

Pearl millet hybrids HHB 67 (843A x H 77/833-2) and 843A x ICMR 01029, which are near-isogenic for a portion of linkage group 2 containing a major QTL for terminal drought tolerance and have been repeatedly demonstrated to exhibit differential response to terminal drought stress, were sown in a line-source irrigation experiment in the summer drought nursery at ICRISAT-Patancheru in late January in anticipation of approval of this project. The field experiment consisted of four replications of 4-m x 20 row plots (60 cm between rows) distributed along a bordered line-source sprinkler irrigation system. The crop was fully irrigated up to the boot leaf stage of growth, and from then on received weekly irrigations from the line-source sprinkler system so that rows nearest the line-source sprinkler remained unstressed while those most distant from the line-source sprinkler experienced moderately severe terminal drought stress resulting in a 40% reduction in grain yield.

Unfortunately, it was not possible for collaborating partners at CuoH and ICGEB to purchase the kits and other lab consumables required for RNA sample collection before the standing crop reached maturity in mid April, so a separate greenhouse sowing of these hybrids was undertaken explicitly for this purpose.

In mid April a dry-down experiment was sown under greenhouse conditions at ICRISAT-Patancheru. This experiment consisted of 20 large pots (30 cm diameter) filled with a soil mixture of (soil:sand:manure, 3:1:1) of each of the two pearl millet hybrids (thinned to one plant per pot shortly after seedling emergence), with 10 pots of each genotype randomly assigned to fully-irrigated or drought stressed treatments. Greenhouse conditions included natural solar radiation and day lengths with air temperatures maintained at 28/22°C for day and night. Plants were grown under well-watered conditions up to boot leaf stage following seedling emergence. The afternoon before the drought treatment was initiated; all pots were fully watered (saturation). After draining overnight, each pot was enclosed in a white plastic bag that was drawn tight around the plant stem to prevent direct soil-evaporation, and a small tube was inserted in the plastic bag for pot re-watering. Pots were weighed after bagging and

this initial weight recorded for each pot. Thereafter the pots were weighed every morning. Daily transpiration was calculated as the difference in pot weight on successive days. To avoid too rapid imposition of stress and to homogenize the development of drought stress across replicated potted plants, the decrease in soil moisture was limited to a net loss of 100 gm per day per pot, controlled by partial re-watering of the stress-treated pots. Well-watered control pots of both genotypes were maintained at their initial weight by adding the daily water loss back to the pot.

Stress-treated plants began to show visible symptoms of stress on 04 June and tissue sampling for RNA isolation was performed on 06 June (one week after imposition of the stress treatment), by which time the stressed plants of HHB 67 had reduced their transpiration to approximately 39% of that of their non-stressed counterparts.

RNA was isolated from panicles and flag leaves of stressed and non-stressed plants of the two hybrids that were at as nearly identical phenological stages of development as possible. Stressed plants of the two genotypes had comparable levels of transpiration relative to their non-stressed counterparts. Differential EST library development is now underway.

RNA samples collected from differentially stressed and non-stressed plants of near-isogenic hybrids differing primarily for a major QTL on pearl millet linkage group 2 that is associated with terminal drought tolerance.

Due to late release of funds for this project, first to ICRISAT and subsequently from ICRISAT to the collaborating partners at CuOH and ICGEB, the partners were unable to purchase lab consumables required for RNA extraction in time to collect samples for the field-grown plants, so a greenhouse dry-down experiment of plants grown in large pots was conducted to produce the differentially stressed near-isogenic plant pairs required for RNA sample collection. This will result in several months delay in development and sequencing of the differential EST libraries, and hence in their subsequent annotation and mapping, but should not otherwise affect the project outcome.

The genotypes used were hybrids of pollinator lines near-isogenic for a putative drought tolerance QTL from donor parent PRLT 2/89-33 in the genetic background of elite hybrid parental line H 77/833-2. These materials were used, rather than the seed parent materials (based on donor parent 863B and recurrent parent 841B) originally proposed, because we have a stronger body of evidence to show that marker-assisted selection has been successful in transferring improved levels of terminal drought tolerance in case of the pollinator lines. This change in genetic backgrounds is not expected to have any consequences as the target QTL(s) on pearl millet linkage group 2 are the same for both sets of genetic materials.

The same plants used for RNA sample collection were also assessed for their ABA and proline levels, and it was observed that 843A x ICMR 01029 has higher ABA and proline levels than HHB 67. This was observed under non-stress conditions, but the genotypic differences were greatly enhanced under stress conditions. The more drought-tolerant genotype 843A x ICMR 01029 appeared to detect the onset of stress earlier (67% available water) than its more drought-sensitive near-isogenic counterpart HHB 67 (48% available water), so that leaf gas exchange slowed first in the drought-tolerant hybrid, but continued for a longer period so that photosynthate required for grain filling was maintained.

A recent paper (Bertin et al. 2005 TAG 110: 1467-1473) demonstrates that the EST resources from this project can be used to develop single-strand conformation polymorphism single

nucleotide polymorphism markers, using annotated rice genomic sequences to initially predict the intron-exon borders in pearl millet ESTs and then design primers that amplify across the introns. The resulting markers have considerable potential for use as conserved orthologous sequence markers across the grasses. We will therefore attempt to identify and map such SNP markers as well as indel markers and SSR markers from the pearl millet EST sequences generated by this project.

14b. Stress Response-enriched EST Resources for Targeted Species— Cowpea

Principal Investigators and Collaborating Scientists:

Sarah Hearne , IITA

Morag Ferguson, IITA

Chris Town, The Institute For Genomic Research, USA

MID-YEAR REPORT

Plants of four cowpea lines; Vu7778 (drought susceptible), Tvu11986 (type I drought tolerance), Dan Ila (type II drought tolerance: stay green) and 12008D (fodder type) have been grown in the in the presence and absence of drought stress. Root, stem and leaf tissues have been harvested at one time point from both stressed and irrigated replicates; in addition the same materials have been harvested from re-watered stressed individuals. Tissue has been stored and is ready for RNA extraction

A fodder type of cowpea has been included as we have joined forces with ILRI collaborators who have inputted their own funds to improve the size of the potential data set.

15. Musa Genome Frame-map Construction and Connection with the Rice Sequence

Principal Investigator:

Takuji Sasaki, NIAS

Co-Principal Investigators:

Nicolas Roux, INIBAP-IPGRI

Isabelle Hippolyte, Agropolis-Cirad

Manoel Souza, EMBRAPA

Collaborating Scientists:

Pat Heslop-Harrison, University of Leicester

Jaroslav Dolezel, IEB

MID-YEAR REPORT

- Two genomic BAC libraries of *Musa acuminata* cv. Calcutta 4 have been deposited in the Musa Genome Resource Centre:
 1. C4BAM comprising 17280 clones and representing 3x genome coverage
 2. MA4 comprising 55296 clones and representing 9x genome coverageThese two libraries are ready for distribution (BAC clones, colony filters)
- A genomic BAC library of *Musa balbisiana* cv. Pisang Klutuk Wulung has been deposited in the Musa Genome Resource Centre. The library comprises 36864 clones and represents 9x genome coverage. The library is ready for distribution (BAC clones, colony filters)

- An EST cloneset (208 x 384-well plates) has been obtained from Syngenta and deposited in the Musa Genome Resource Centre: The clones can be distributed to project participants.
- The actual strategy to prepare the pool is still being discussed by the project participants. The final conclusion will be made by September 2005. After this date, DNA pools will be prepared.
- A genomic library of *Musa acuminata* cv. Tuu Gia cloned in a BIBAC vector has been deposited recently in the Musa Genome Resource Centre. The library comprises 30700 clones and represents 5.1x genome coverage. The library is being prepared for the distribution of BAC clones and colony filters.
- Due to administrative issues, a genomic BAC library of cv. Grande Naine has not been received until now from CIRAD. However, it is expected that the library will be received by the end of 2005.

Three F1 populations have been created from *Musa acuminata* diploid accession :

1. Pahang doubled haploid (♀) X Pisang Madu (♂): 190 individuals will be ready for DNA extractions in March 2006
2. Borneo (♀) x Pisang Lilin (♂): 280 individuals will be ready for DNA extractions in December 2005
3. Malaccensis (♀) x Gabah Gabah (♂): 200 individuals will be ready for DNA extractions in December 2005.

The recovery of individuals has been made by *in vitro* embryo rescue. The regenerated plantlets are still *in vitro* at the moment.

To choose the best population for mapping, the 6 parents have been analyzed for: (see appendix)

- Determination of the level of heterozygosity of the parents;
- Determination of genetic distance between both parents
- Evaluation of structural heterozygosity

614 repetitive DNA sequences were isolated from *Musa acuminata* cv. Calcutta4 after Cot selection. The clones were sequenced at TIGR and their sequences are now being analyzed for homology to DNA sequences in public databases and their abundance in the genome of *M. acuminata* and 35 species representing genetic diversity of the *Musa* genus.

The choice of the 17 BACs to be sequenced is in progress.

High density membrane from the BACs libraries of *Musa acuminata* Calcutta 4 and *Musa balbisiana* PKW have been prepared.

During our project coordination meeting held at INIBAP the 6 and 7 of June 2005, NIAS communicated with CIRAD and EMBRAPA and they agreed that they will send their library after they remove most abundant clones in September. NIAS will rearrange them and sequence 5000 clones from both ends.

BAC clones selected by CIRAD and Univ. Leicester will be end-sequenced at IAEA, then sent by IEB to NIAS in September for sequencing.

The ADH1 (alcohol dehydrogenase) region will be targeted for comparing BACs from *M. Acuminata* and *M. balbisiana* genome. The identification of 14 other genes related to biotic and abiotic stresses is also in progress.

The new organization of the Global *Musa* Genomics Consortium web site has been defined in order to give a better visibility to the data produced and analysed. A first draft on the new graphic chart has been done. The Genome Browser, Gbrowse, has been installed and configured and existing BACs has been integrated.

Tangible outputs delivered:

- Around 32,000 EST clonesets were obtained from Syngenta and deposited at the Musa Genome Resource Cente (MGRC)
- A *Musa acuminata* cv. Tuu Gia BIBAC library was deposited at the MGRC
- Creation of three mapping populations containing at least 150 individuals
- Diversity analysis of the 6 parents generating the three mapping populations
- 614 repetitive DNA sequences were isolated from *Musa acuminata* cv. Calcutta4 after Cot selection
- BACs libraries high density membranes from *M. acuminata* Calcutta 4 and *M. balbisiana* Pisang klutuk wulung

I. Genetic analysis of the 6 parents used to create the 3 mapping populations:

a. Level of heterozygosity of the parents

The level of heterozygosity of the parents was established based on the analyze of 21 SSR, previous RFLP data for all accessions, and previous SSR data for Borneo

	Nb of alleles	% heterozygosity SSR markers	% heterozygosity 42 RFLP markers
Pahang DH	21/21 μ sat	0 %	-
Pisang Madu	34/21 μ sat	62 %	67 %
Borneo	30/16 μ sat +56/33	76%	24 %
Pisang Lilin	37/21 μ sat	76 %	56 %
Malaccensis	28/20 μ sat	40 %	33 %
Gabah Gabah	40/21 μ sat	90 %	61 %

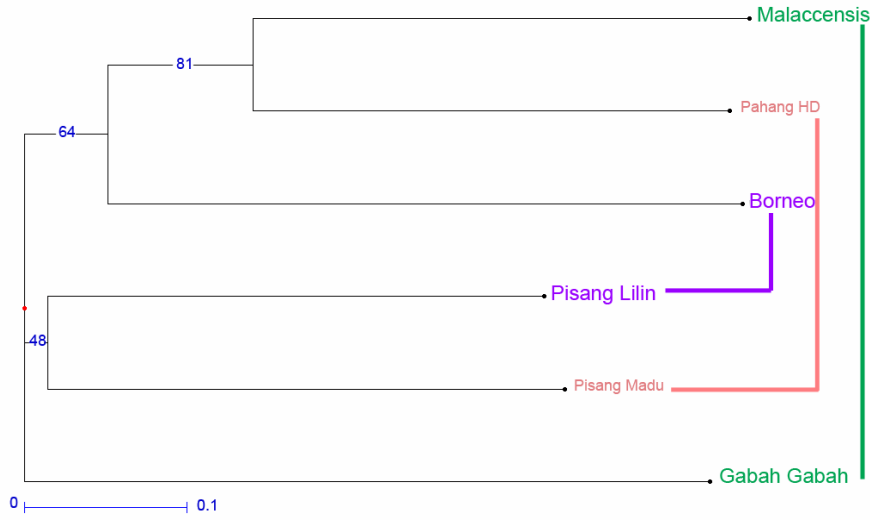
b. Genetic distances between parents

Table of genetic distances calculated based on 20 SSR loci, simple matching distance

	1	2	3	4	5
2	0,775				
3	0,78125	0,78125			
4	0,75	0,625	0,78125		
5	0,6	0,775	0,78125	0,75	
6	0,85	0,75	0,84375	0,75	0,9

PahangHD	1
Pisang Madu	2
Borneo	3
Pisang Lilin	4
Malaccensis	5
Gabahgabah	6

Genetic tree: Table of genetic distances calculated based on 20 SSR loci, simple matching distance, 100 bootstrap, unweighted neighbor joining tree (Darwin software)



c. Structural heterozygosity of the male parents:

The structural heterozygosity was estimated based on the analysis of pairing at meiosis

	Metaphases							Anaphases				Conclusions
	N b c e l l s	1 IV 9 II	1 IV 8 II 2 I	1 III 8 II 3 I	1 III 9 II 1 I	10 II 2 I	11 II	Nb c e l l s	11/ 11	12/ 10	13/ 9	
P. Lilin	7	-	-	-	5	1	1	1	1	-	-	1 translocation
P. Madu	17	3	1	2	7	-	-	7	3	3	-	1 translocation
P. jari buya (=gab ah	6	2	-	4	-	-	-	2	2	-	-	1 translocation

gaba h)											
--------------------	--	--	--	--	--	--	--	--	--	--	--

Legend: IV: tetravalent; III: trivalent; II: bivalent; I: monovalent.

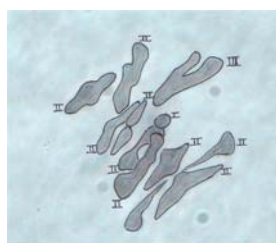
For Pisang Lilin one trivalent (III) was observed in the five cells analyzed suggesting that P. Lilin is heterozygous for one translocation. These results corroborate the conclusions of Dodds and Simmonds (1948). This translocation was called the “Northern Malayan translocation” by Shepherd in 1999. The features of the multivalents suggest that this translocation is located on the distal portion of the chromosome.

For Pisang Madu, we systematically observed one tetravalent (IV) or one trivalent (III) in the 17 observed cells. These results suggest that Pisang Madu is heterozygous for one translocation. The occurrence of tetravalents with a “frying pan” shape and one trivalent suggests that this translocation is in interstitial position, proximally to the centromer. The disequilibrium in the repartition of the chromosomes at anaphase will lead to plants with at least 22 and 23 chromosomes in the progeny.

The Gabah Gabah clone has not yet been analyzed. Nevertheless, this clone is a mutant of Pisang Jari Buaya which was analyzed in 1992 by CIRAD (pictures above). The observation on 6 cells suggested that this clone is heterozygous for one translocation probably in distal position. These results have now to be consolidated by observations on the clone itself upon a larger number of cells.



1



2

Illustration:

- 1- Picture of a Metaphase of Pisang Madu (on one plane) and 2- its interpretative diagram (amplification x 2000) : 1 trivalent; 9 bivalents; 1 monovalent.

Based on the results obtained, we shall by the end of summer decide which population will be used for the map construction. The 2 other populations will be save as frozen leaves.

II. Figures SP2-15 July 2005

- a. Participants of the 1st GCP Musa-Oryza coordination meeting, 6-7 July 2006, INIBAP, Montpellier
- b. Musa Genome Resource Center
- c. Mapping populations
- d. Characterization of a CoT 0,05 library
- e. Homology to genebank sequences
- f. Musa genome to rice chromosome

16. Validation of Conserved Orthologous Markers

Principal Investigators and Collaborating Scientists:

Jizeng Jia, CAAS
Lifeng Gao, CAAS
Qingming Sun, CAAS
Nicolas Roux, INIBAP
Pat Heslop-Harrison, University of Leicester
Mathieu Rouard, INIBAP
Merideth Bonierbale, CIP
Roland Schafleitner, CIP
Reinhard Simon: CIP
Michael Baum, ICARDA
Sripada M. Udupa, ICARDA

MID-YEAR REPORT

100 COS II primers for candidate genes for disease resistance/defence, drought tolerance or other functions were selected with Cornell U/SGN and synthesized. Primers were screened for polymorphism in subsets of 3 mapping populations, BCT, PD, PCC-1, of which 24 were polymorphic in at least one cross. Mapping is in progress to be followed by sequencing of products to validate amplification of target DNA. Six primers amplified monomorphic fragments, and mapping efforts will continue with SSCP, CAPS or SNPs (See table). Samples of bean (M. Blair, CIAT: DOR364 and G19833) and sweetpotato (CIP) were included in the screening and 30% gave clear PCR products. Of 15 primers COS II, 8 were scored in the BCT and PD populations, of which 6 mapped as predicted from tomato (SGN), to chr IV, VI, VII. One whose position is not available from SGN mapped in potato chr X. 12 products of COS I and 7 of COS II are being sequenced. Genotypes of 4 wild and one cultivated potato species of known resistance phenotype were assembled toward application of COS to understanding diversity in germplasm. *Bioinformatics*: COS identification strategies are being reviewed in collaboration with SP4, using the custom BLAST from Paracel and genome annotation tool on the HPC. EMBRAPA was contacted to help establish the GENOMA EST annotation pipeline (including primer design) on CIP's HPC in collaboration with CIAT and installation is scheduled for August. Further links with SP4 are established via collaboration on the common functional gene catalog. Available annotation tools for metabolic pathways and integration with sequence information were identified (BioCyc software) and are scheduled for evaluation and in 2nd half of 2005. Documentation on COS annotation strategies was initiated for publication on a web-site. In Agr & Agri-Food Canada, SNPs have been identified in potato DNA sequences corresponding to 31 tomato-At COS based on comparisons of sequences available in NBCI. Six have detected polymorphism among potato genotypes.

Fifty abiotic stress (drought, cold and ABA) induced gene sequences from microarray experiments of Arabidopsis were used to identify putative orthologues and develop COS markers for legumes. These genes include senescence/dehydration-associated protein-related (ERD7), aldehyde dehydrogenase (NAD⁺), pathogen-inducible alpha-dioxygenase, protein phosphatase 2C-like protein, probable succinate dehydrogenase flavoprotein subunit precursor, glyoxalase II, 3-ketoacyl-CoA thiolase-like protein, probable receptor-like protein kinase [imported], nodulin-like protein, embryo-specific protein 1 (ATS1), etc. Sequences were aligned with other dicots preferably EST sequences of *Medicago*, soybean and lotus. Primer design at conserved regions are in progress. Amplification products of 6 COS markers

with six legumes of ICARDA' interest (faba bean, lentil, chickpea and grasspea) were eluted from agarose gels, and prepared for sequencing and/ cloning.

During the first six months of SP2-7, CAAS focused on the development of disease resistance EST-SSR markers across monocots, including wheat, rice, maize and barley. From the public databases of NCBI and TIGR, a total of 48 SSR-containing ESTs were discovered and 61 EST-SSR primers were designed based on wheat ESTs. Of which, 21 ESTs contain LRR or NBS-LRR domain, 21 contain PK or LRR-PK domain, the other 6 ESTs were related to disease resistance. 46 EST-SSR primers have been synthesized and experiments showed that 24 primers get expected and clear PCR products, 12 showed polymorphism in wheat varieties and have been used for genetic diversity analysis, and 6 primers used for genetic mapping in three mapping populations. Of 15 primers used to validate the transferability, 12 produced strong bands across the four monocots tested, and the other 3 primers did not get amplicons in maize.

Of 8 COS markers (Fulton, Tanksley et al) used, 2 isolated the expected genes from Musa. COS1263 gave strong amplification from all Musa accessions studied, with a length of 600 bp. COS1006 amplified a single fragment of 220bp in all triploid AAA accessions, with weak products in other triploids. Another gave retroelement fragments, and others poor products.

A) One primer was redesigned from Musa ESTs, and isolated the expected polygalacturonidase gene. Interestingly, this was absent from the 30,000 sequence Musa EST database. SNPs were found in the successfully amplified sequences. B) Genes for dihydrofolate reductase and chalcone synthase were targeted, using sequences where present in the Musa EST database and homology to model species. C) Newly designed primers for abiotic stress-related genes include drought responsive elements and candidate genes discovered through differential display using stressed and control plants (Dhairyasheel Desai). Two genes used are the drought-related gene DRE binding factor 1 and DREPP2 protein. This work has set the scene for a larger scale and more systematic comparative study of genes between rice, Musa and other model species. Primers working from Musa (and other spp) are being tested with DNA provided by other GCP members (rice, T. Sasaki; cassava, M. Fregene).

The sequencing of orthologous genes is very important to understand the variation present between different species. However, it is important to communicate how much preliminary work is required for primer design toward this objective, and to what extent it can be expected that conserved gene primers can be designed using straightforward methods across all species. As reported at the end of 2004 in this collaborative activity, while conservation of genes between species can be detected by bioinformatic approaches at the DNA (e.g. >60% homology over > 100 bp as a typical threshold) or protein level (>30% similarity over 20 amino acids), the identification of conserved regions for primer design, and the amplification of the targeted regions is proving to be more difficult than expected. Our previous studies have taken advantage of information about the individual genes being targeted (making primer design more straightforward); for example, the design of primers around leucine (with six possible codons, and important in many resistance genes) is hardly possible, as conserved sites seem to be inappropriate separations for genomic PCR. Conservation was often low between primers, and some publications rooted in model organism work are perhaps over-optimistic regarding gene conservation.

In conclusion, only 2 of 8 COS primers (Cornell U) amplified orthologous products for Musa as published, with refinement of primer design, ideally based on EST comparisons, being

required to obtain more useable primers for targeted genes. Targeting genes of interest, using both ESTs from Musa and/or genes from other species is generally effective in finding the gene conserved in banana.

Tangible outputs delivered:

- 100 COS II primers (SGN) synthesized
- 45 COS II (SGN) amplified, surveyed for polymorphism and/or mapped in potato
- 46 EST-SSR primers synthesized
- 24 EST-SSR primers amplified, surveyed for informativeness and/or mapped in wheat
- 12 EST-SSR primers amplified across 4 monocots, and 3 more produced amplicons in wheat, rice and barley.
- Abiotic stress genes identified *in silico* and aligned for COS marker development
- Primer designed and amplified on a panel of four legumes
- Sequence analysis of PCR products initiated

Deviations from the work plan:

None.

I. Table of COS I and COS II amplified in potato (CIP)
(Please see author for complete list)

II. Figures SP2-7 July 2005

- a. Screening of COS II in Potato survey set CIP
- b. COS II markers mapped to Potato chr 4 CIP
- c. COS II markers mapped to Potato chr 6 CIP
- d. Amplification of COS II marker in 4 Dicots CIP(CIAT)
- e. Experimental approach ICARDA (a)
- f. Experimental approach ICARDA (b)
- g. Amplification of COS II marker in 3 Dicots ICARDA
- h. Screening RGA-SSR primers primarily across Monocots CAAS
- i. Diversity analysis using RGA-SSR in Wheat CAAS
- j. Mapping RGA-SSRs onto Wheat chromosomes CAAS
- k. Summary INIBAP/ U Leicester
- l. Primer sequences prepared for Musa INIBAP/ U Leicester
- m. Amplification profile of DREBP in Musa germplasm INIBAP/ U Leicester

17. Comparative QTL Mapping for Drought Tolerance

Principal Investigator:

Mathias Lorieux, IRD/CIAT

MID-YEAR REPORT

Progress was made in the areas of genotyping and phenotyping for drought tolerance in the populations of interest for fine mapping of QTLs:

1) Genotyping: We used a large number of newly-developed genomic and cDNA based microsatellite markers to anchor and extend RAPD and AFLP maps developed for two Mesoamerican gene pool populations based on the crosses of BAT881 x G21212 and DOR364 x BAT477. As locus-specific markers the microsatellites have been useful for determining synteny and correlations between the genetic maps as well as for comparative mapping of the drought tolerance traits. Recombinant inbred line populations are useful for

these genetic studies because they are being grown over a wide range of environments in a consistent and statistically reliable fashion (large plots, many replicates). We also used cross-legume comparative markers to link the common bean genetic map with the genetic maps for *Medicago truncatula* and other legumes based on markers developed at Univ. of California - Davis. Our initial work focused on linkage group Pvb03c from common bean that has a low saturation of microsatellite markers in maps from narrow crosses within the Mesoamerican gene pool.

2) Phenotyping: greenhouse studies were continued for the RIL parents (BAT 881, G 21212, SEA 5, MD 23-24) and a check (BAT 477), where plants were grown in large plastic cylinders (100 cm long and 15 cm diameter) covered with PVC tubes with three levels of water supply: 100% field capacity (well-watered), 60% field capacity (moderate drought stress) and 30% field capacity (severe drought stress) as main plots and genotypes as sub-plots. Treatments of water stress were imposed after two weeks of initial growth of plants and maintained by weighing each cylinder every week. Data was taken on shoot biomass distribution, root biomass and root length distribution in different soil depths, leaf and stem nutrient composition, ash content and TNC (total nonstructural carbohydrates).

Tangible outputs delivered:

Phenotyping protocol

Deviations from the work plan:

No deviation from plan.

35. Sequencing Multiple and Diverse Rice Varieties: Connecting whole-genome variation with phenotype

Principal Investigators:

K. McNally, IRRI

D. Mackill, IRRI

H. Leung, IRRI

R. Bruskiewich, IRRI

Collaborating Scientists:

K. Frazer, Perlegen Sciences, Inc

D. Cox, Perlegen Sciences, Inc

C. Xu, Perlegen Sciences, Inc

E. Peacock, Perlegen Sciences, Inc

MID-YEAR REPORT

This project provides partial support (28% of current funding) for undertaking genome-wide SNP discovery by re-sequencing multiple, diverse rice varieties through DNA-DNA hybridization on high density oligonucleotide arrays and is a partnership between IRRI and Perlegen Sciences, Inc. and spans subprograms 1, 2 and 4. Current funding to support this effort is \$0.7 million with the balance from IRRI.

In Perlegen's approach to re-sequencing, all non-repetitive regions of the genome greater than about 60 bases in length are tiled using 25-mer oligonucleotide with a sliding window offset by one base for both strands and where the middle base is 4-fold degenerate. Hence, 8 oligonucleotide are present on the array to interrogate every position contained in the non-repetitive regions. Once arrays are fabricated, target DNA is fragmented, labeled, and hybridized to the array. Its application on SNP analysis of human was reported in Science on

18 February 2005 (“Whole-genome patterns of common variation in three human populations” by Hinds *et al*). Currently, Perlegen is also applying this technology for SNP discovery in mouse and Arabidopsis.

The design of the project involves 2 phases. First phase (supported by current funding) will generate data on up to 15 diverse rice varieties for about 125 Mbp of the genome. Since the size of the rice genome is 390 Mbp and assuming that ~40% is repetitive (transposons, centromeric and telomeric repeats, microsatellites, etc.), approximately 60% of the genome will be covered in this phase. Three different approaches for the preparation of target DNA are being investigated -- direct labeling of high quality CsCl₂-banded genomic DNA, whole genome amplification, and long-range PCR.

Funding for a second phase to complete the genomic coverage and extend the set of rice varieties from 20-30 is being sought through partnerships under the umbrella of the International Rice Functional Genomics Consortium. After the initial costs of array design, the cost for re-sequencing one variety is about \$100,000. All data from this project will be in the public domain and will enable genome scanning in rice as well as the design of SNP assays for specific genes of interest. For example, assuming a haplotype block size of 100 kbp, a genome scan of rice could be accomplished by interrogating around 3,900 features, a very doable task using current microarray platforms.

- IRRI has assembled the collection of varieties/germplasm for seed increase and purification. A main criterion for choosing varieties of germplasm is their comparative diversity while demonstrable utility for plant breeding is an added benefit. A dendrogram (based on SSR fingerprinting) was constructed for ~22 varieties to document their genetic distances.
- The primary sequence data for the design of oligonucleotide will be the release 4 of the high-quality BAC-by-BAC *japonica* sequence of Nipponbare (International Rice Genome Sequencing Project). Regions unique to indica will be tiled using the whole genome shotgun *indica* sequence of 93-11 (Beijing Genomics Institute). Masking for repetitive regions is being done in collaboration with Tom Bureau (McGill University).
- For testing the different hybridization techniques, 20 mg of purified genomic DNA from IR 64 and Shan-Huang-Zhan 2 (*indica* types) and Azucena and Li-Jiang-Xin-Tuan-Hei-Gu (*japonica* types) were isolated at IRRI and shipped to Perlegen Sciences on September 9, 2005.
- Letter of Research Agreement was finalized between Perlegen Sciences and IRRI on August 26, 2005 allowing shipment of genetic materials to Perlegen.

Deviations from the work plan:

No deviation.

SP3 COMMISSIONED GRANTS

18. Development of Low Cost Gene Based Trait Assay Technologies in Cereals

Principal Investigator:

Casiana M. Vera Cruz, IRRI

Co-Principal Investigator:

Manilal Williams, CIMMYT

MID-YEAR REPORT

A. Recipient germplasm materials from NARES collaborators

1. On rice improvement for bacterial blight (BB) resistance at IRRI, NARES collaborators from China, the Philippines, Indonesia, India and Africa were surveyed for the germplasm materials serving as recipients in their national program for improving resistance to BB. Each country submitted 2-6 recipients for introgression of bacterial blight resistance genes, and DNA of these recipients are being sent when the seeds are not available yet for shipment. However, to be useful in MAS, the DNA originated from the same batch of seeds used for crossing work, if crosses were made, or from breeder's seeds.
2. In China, the DNA of four susceptible recipients representing restorer lines for hybrid rice breeding were sent to IRRI for sequencing the Xa21 and Xa7 S alleles. In the Philippines, Xa21 and Xa7, in addition to Xa4 (not used in this project), have been introgressed into advanced lines of hybrid rice maintainers IR58025B and IR68888B, thus the DNA from breeders' seeds of IR58025B were used. DNA extracted from seeds of Indian and Indonesian susceptible recipients have been received at IRRI (with the corresponding documents). Three to four individual plants were used to extract DNA, and two of these DNA samples are being amplified, purified and sent for sequencing in commercial laboratories.
3. On maize improvement for quality protein maize (QPM) at CIMMYT, eight QPM donor sources from diverse agroecological zones (mainly tropical lowland and subtropical) and nine non-QPM recipient sources from tropical highland, tropical lowland and sub-tropical regions have been selected for sequence comparison of the alleles at the opa2 locus. The non-QPM sources have been selected based on adaptability and superior performance in respective regions. Currently, there are no available sources of QPM for tropical highlands.
4. The QPM sources selected for the project are as follows: tropical lowland, white – CML 159, CML 144, CLQ 4203, CLQ–RCW Q01, CLQ–RCW Q50; tropical lowland, yellow – CML 161, CML 165; and sub-tropical, white – CML 176. These QPM sources are important QPM sources covering areas of lowland tropical and sub tropical maize environments in Asia, Africa and Latin America. They also contain excellent kernel modifying capacity and have been extensively used in line conversion activities.
5. The recipient sources selected from diverse agro-ecological zones are: tropical white – CML RCW 22, CML 343, CML 254, RCW01; tropical yellow – CML 348, CML 451; highland white – CML 244, CML 349; and Ethiopian white – F7237, A-7018. These normal maize materials have been selected based on their use and importance in Asia, Latin America and East African highland regions.
6. The seed material of QPM and non-QPM source material have been collected and DNA extractions from young seedlings have been completed.
7. The markers Xphi057 and Xumc1066 have been used in initial studies assessing sequence differences among different QPM donor and recipient material. Three different allele sizes exist among the eight QPM donor sources for Xphi057 and two different allele sizes exist for Xumc1066. Allelic diversity is higher among the normal, non-QPM material for the two markers.

B. Progress on the development of the dot blot-based detection assay

8. We have initiated the development of dot-blot allele-based assay to determine if the hybridization and DIG-detection system (<http://www.roche-applied-science.com/>) is possible with the current system of markers and template DNA carrying R, S and heterozygous alleles. We used the original Xa21 primers to amplify the target

template DNA and the probe from IRBB21 and IR24 while the allele sequences of the S recipient rice cultivars from the NARES partners were not available at the time of evaluation.

9. Using the original Xa21 primers, both R and S alleles hybridized to the DNA templates regardless of genotype (homozygous R, S or heterozygous). While this showed that the dot-blot assay works for this system following manufacturer's recommendation, there was no specificity of detection by R and S probes to the target alleles, suggesting that the sequence of the R-allele probe was highly homologous to the S-allele probe. Thus, when all sequences of S alleles from the recipient cultivars in NARES breeding program become available, highly specific primers will be designed for the probe based on sequences of recipient alleles.
 10. Various parameters were also evaluated to optimize conditions for the assay. They were concentrations of DNA blotted onto nylon membrane, post-hybridization stringency wash conditions, and different incubation times of detection substrate. These parameters are being reevaluated as the specific probes are becoming available.
- C. Initial sequence comparison of Xa21 and xa5 R and S alleles
11. We made sequence comparison focusing first on Xa21 and xa5 R and S alleles in the GenBank and the available target sequence in the donor and recipient materials for the project. We aligned Xa21 sequences deposited in the GenBank for the seven gene family members composed of the Xa21 functional allele, Xa21C, D, A1, A2, and F which originated from *O. longistaminata* (Song et al., 1995). Xa21E was reported to originate from the susceptible IR24 background of NIL IRBB21 (Wang et al., 1996). The Xa21 Forward primer of the original polymorphic primers for gel-based assay was found on the 6 gene sequences, but not on the E sequence of Xa21 (Forward or Reverse) primers.
 12. The phylogeny of these gene family members was analyzed using Clustal X. Xa21D and E grouped in one cluster while Xa21A2, C, and A1 formed another group. Interestingly, the Xa21 R allele, which we amplified and sequenced from IRBB21 resistance donor, formed a robust cluster (bootstrap value of 99.8) with the functional Xa21 deposited in the GenBank. Also, the Xa21 S alleles from IR24, the control Vandana, and the recipient maintainer IR58025B from PhilRice formed one tight cluster that is different from the R allele cluster suggesting that there are diverged regions or SNPs between the Xa21 R and S alleles, which may serve as useful probes for developing allele-based assay technologies.
 13. The cloned sequences of xa5 R and S alleles deposited in the GenBank were used in sequence alignment to design/identify specific primers for the probe. We used the region where the 3 base pair difference is located. To amplify the target sequence from the breeding lines (DNA template), we will use the sequencing primers published in Iyer and McCouch (2004) instead of the polymorphic primers derived from RG556.
- D. Probe/Marker design based on sequence comparison of Xa21 R/S alleles and QPM alleles
14. Sequences of the susceptible Xa21 allele from IRBB21, IR24, IR58025B, Vandana, and Moroberekan were aligned and compared with Xa21 R allele from IRBB21. These sequencing results will be used as the bases for the synthesis of highly specific oligonucleotide probes (Cahill and Schmidt, 2004; <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.508>) for end-labelling with a terminal transferase.
 15. Two groups of oligonucleotide probes were designed based on: (a) diverged regions and indels, and (b) SNPs between R and consensus S allele sequences. To increase the specificity of hybridization for dot-blot assay, probes/markers were designed as short as 18-20 nucleotides, thus, avoiding long stretches of identical sequences

between R and S alleles. However, compared to conventional DNA probe, oligonucleotide probes that are allelic and have single mismatch are unstable at high hybridization stringency

(<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.458>), therefore, we are evaluating the use of low temperature condition during hybridization between probe and template, and optimize also other conditions for hybridization and detection for dot-blot assay.

16. Sequence analysis is currently in progress in order to evaluate any diverged regions and/or indels, and for identifying SNP differences among different donor and non-QPM recipient sources.
- E. BLAT (BLAST-Like Alignment Tool) analysis for Xa21 and xa5 probes
 17. BLAT is a high-speed mRNA/DNA and translated protein alignment algorithm (Kent, 2002; <http://www.genome.org/cgi>). Using BLAT, all R probes for Xa21 were not located in the Nipponbare genome which may indicate that these R sequences are unique to Xa21 from *O. longistaminata*. Of the S probes, only 5'-GTGCTGGAAATAGTAACCGG-3' is located in Chr 11 of Nipponbare. The other S probes were not located in the Nipponbare genome suggesting that they are unique sequences in the Xa21 S alleles in the recipient cultivars.
 18. Three xa5-R and three xa5-S markers were designed for xa5. Of these, the R probe 5'-CCCGGAGCTCGCCATTCAAG-3' (position 20-39) and S probe 5'-AGTTCTTGTCCAGTTTGATA-3' (position 38-57) were located in Chr 5 of rice. BLAST results of the xa5 gene in the Nipponbare genome shows it is similar to LOC_Os01g73890 (Chr 1 transcription factor gamma iia) and LOC_Os05g01710 (Chr5 transcription factor gamma). Alignment of the Chr5 LOC_Os05g01710, xa5 R allele and xa5 S allele showed that the first SNP region in Nipponbare is similar to the R allele (CG), and the second SNP region is similar to the S allele (TC). This is probably the reason why both R and S probes are located in LOC_Os05g01710 using BLAT. The probes containing two diverged nucleotides between xa5 R and S alleles were not found, indicating that these are likely the useful probes/markers.
 19. For preliminary evaluation, three selected probes (two indels and one SNP) designed for Xa21 and one for xa5 R and S alleles have been sent for synthesis. Evaluation for probe efficiency is in progress.

Based on current results, we hypothesize that to obtain specific primers and probes for MAS application, sequences need to be designed from the specific combination of donor R allele and recipient S allele used as parental cross in the breeding program. Degenerate primers or probes may reduce the specificity of detecting the target R or S allele. Hybridization conditions and detection need to be optimized between R/S alleles and the target templates.

Tangible outputs delivered:

1. Susceptible recipient germplasm for developing bacterial blight resistant lines and improving QPM in national breeding program gathered from participating NARES.
2. Collection of DNA of a few recipient rice cultivars from NARES collaborators received for sequence analysis of S alleles.
3. Oligonucleotide probes for Xa21 and xa5 R and S alleles (IR24 only for xa5) developed based on (a) diverged regions/indels, and (b) SNPs for evaluation by dot-blot, microplate-based, and modified TAM assays.

Plans for the next 6 months:

1. Collect seeds of the recipient germplasm when all documents for seed exchange have been accomplished between CG centers and country of NARES partner. As needed,

- PCR products amplified from these seeds of R donor alleles and recipient S alleles in NARES breeding program and control cultivars shall be sent for sequencing as further basis for designing specific primers and probes.
2. Complete the design of primers and probes for R and S alleles for 3 Xa genes and QPM alleles.
 3. Optimize the dot-blot assay using the new oligoprobes designed from Xa and QPM R and S alleles of donor and recipient rice and maize cultivars, respectively.
 4. With the availability of new probes, develop and optimize the microplate-based assay technology.
 5. Validate the modified TAM approach.
 6. Utilizing the Genomics Facility of the National University of Mexico (UNAM), conduct evaluations of glass slide-based genotyping technique with specific probes
 7. New staff and shuttle researcher from CNRRI
 - a. A consultant with experience in high throughput genotyping techniques is expected to join the project at CIMMYT in September.
 - b. Our primary collaborator in developing the technology for rice, Dr. Jianli Wu and his staff Ms. Jie Chen will join in late September to work on development of various technologies for the regional hub laboratories, especially the FRET-based detection assay.
 8. The microplate reader (~\$34,000.00) for absorbance and fluorescence reading was purchased in August 2005 (from IRRI capital budget). It is expected to arrive in mid-September. Similar equipment purchases also have been made at CIMMYT.

Literature Cited

Cahill DJ and Schmidt DH. 2004. Use of Marker Assisted Selection in a Product Development Breeding Program. In "New directions for a diverse planet". Proceedings of the 4th International Crop Science Congress, 26 Sep – 1 Oct 2004, Brisbane, Australia. Published on CDROM. Web site www.cropscience.org.au

Strachan T and Read AP. 1999. Nucleic acid hybridization assays. In Human Molecular Genetics 2. BIOS Scientific Publishers, Ltd.
<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.458>

19. Evaluation and Deployment of Transgenic Drought-Tolerant Varieties

Principal Investigator:

John Bennett, IRRI

Collaborating Scientists:

Renee Lafitte, IRRI

Matthew Reynolds, CIMMYT

Vincent Vadez, ICRISAT

Enrique Chujoy, CIP

Kazuko Yamaguchi Shinozaki, JIRCAS

Kazuo Watanabe, University of Tsukuba

MID-YEAR REPORT

- A 3-day Start-Up Meeting was held at IRRI in March and was attended by representatives of CIMMYT, ICRISAT, IRRI and JIRCAS.

- Dr. Fernando Ezeta of CIP visited JIRCAS and the University of Tsukuba in May to discuss collaboration on the evaluation of DREB+ potato.
- Following a DREB meeting organized by JIRCAS at IRRI in April, CIAT was invited to join the task for work on DREB+ rice. With permission from the GCP, a Letter of Agreement was exchanged between CIAT and IRRI to formalize this new arrangement.

Using *OsLip9::AtDREB1B* and *OsLip9::OsDREB1B* constructs and *Agrobacterium*-based protocols, CIAT obtained over 400 independent transformants in three important rice cultivars. Production of homozygous plants by selfing and characterization of insertion events by molecular analysis will now be followed by preliminary phenotypic evaluation.

Wheat cultivar Bobwhite was successfully transformed with the *Atrd29A::AtDREB1A* construct in 2000. Pot studies began in 2002 and field trials were initiated in 2004. Ten transgenic events were studied under fully irrigated conditions and pre-anthesis drought stress relieved after anthesis. The 2005 field trial compared continual water stress with re-watering treatment. Several sets of physiological data were collected for the analysis: heads number, biomass, total yield and canopy temperature.

Fourteen independent transformants of groundnut carrying *rd29::DREB1A* events were compared with non-transformed parent JL24, based on two parameters: (i) the soil moisture value where transpiration begins to decline (ii) the number of days plants take to deplete transpirable water after watering was withheld. Typical transpiration response values were normalized against controls and against transpiration data before stress occurred to account for plant-to-plant variations. When transpiration response was plotted as a function of soil moisture available, transpiration declined at lower values of soil moisture (drier soil) in the transgenic plants compared with non-transgenic controls. Certain transgenic lines have higher transpiration efficiency.

Following the Start-up Meeting in March, IRRI evaluated the “dry-down” technique recommended by ICRISAT for applying drought stress to pot-grown plants. Reproducible data were obtained when at least 50% of the daily transpired water was added back at the end of each day. This technique is now being used to modify the rate of stress development. In anticipation of transforming indica and japonica rice varieties with *OsLip9::OsDREB1A* constructs, RT-PCR was used to measure drought-responsiveness of endogenous *OsLip9* and *OsDREB1A* genes in reproductive-stage tissues (flag leaf blades and sheaths, panicles, anthers and peduncles).

Tetraploid potato cv. Desiree was transformed using an *Arabidopsis thaliana* stress inducible promoter *rd29A* and the *DREB1A* gene, which confers multiple tolerances to abiotic stresses (e.g., dehydration, and elevated soil salinity). Transformed Desiree lines showed a direct correlation between *DREB1A* expression levels and tolerance to salinity. By making filial progeny from the transgenic lines which were estimated as furnishing single copy *DREB1A* gene, crossed with a non-transgenic cultivar, the estimation of the genotype at the transgenic locus of these tetrasomic tetraploids were made by chi-square test, and it provided suggestion that these were simplex. Results showed that low copy heterozygous loci containing the *DREB1A* gene could sustain the expression, and consequently provide significant tolerance to salinity in tetrasomic tetraploid potatoes. (Breeding Science, in press)

Tangible outputs delivered:

Minutes of Start-up Meeting held at IRRI in March 2005, including first draft of review manuscript on present status of research on DREB+ transgenic crops for peer-reviewed publication.

Deviations from the work plan:

Staff changes at IRRI limited the ability of the institute to follow the work plan. IRRI management decided to transfer the research funds to a new partner institute (CIAT) for research on DREB+ transgenic rice.

20. Optimising Marker-assisted Breeding Systems for Drought Tolerance in Cereals through Linkage of Physiological and Genetic Models

Principal Investigator:

Scott Chapman, CSIRO

Collaborating Scientists:

Mark Dieters, University of Queensland

Graeme Hammer, Agricultural Production Systems Research Unit (APSRU)

Jiankang Wang, CIMMYT

Maarten van Ginkel, CIMMYT

Richard Trethowan, CIMMYT

Eva Weltzien, ICRISAT

Tom Hash, ICRISAT

Gary Atlin, IRRI

Marianne Banziger, CIMMYT

Mark Cooper, Pioneer

MID-YEAR REPORT

This project only began with the appointment of Jiankang Wang in July 2005, and with his visit to CSIRO/The University of Queensland from July 26 to Sep 11 2005. In his time in Australia, we worked on further developing the research plan, and made a good start on achieving some of the programming and analysis tasks. The project outputs are listed at the end of the report and are referred to below. Note that this project has been designed to present examples of many of the available options in utilizing simulation to improve MAS. It is not a comprehensive study of any single approach, but we are attempting to develop some general guidelines (see further below) as we progress.

The aim of this task is to develop some simulation examples that could be extended to other crops and different genetic models. The initial work is based on a study of an existing case study (wheat breeding at CSIRO) to combine known genes (using 'perfect' or near-perfect markers) into single genotypes for use as parents or further field screening. Some generic work was done during this process that will be applicable to many other research questions in this area.

1. Population genetics model on efficient use of marker-based selection in plant breeding has been built.

When using markers in selection to create a genotype that contains a combination of favourable alleles or markers (target genotype), there are many choices in the design of crossing and selection strategies. Population genetic theory was used to develop some general

guidelines to combine parental alleles in an existing wheat breeding program for unlinked genes and perfect markers. The following questions relevant to marker based development of a target genotype have been studied.

- How many genes can be selected simultaneously in one cross?
- What is an efficient method to pyramid genes from F2?
- Type of cross to use: single cross, backcross or top cross?
- In what order should markers be selected?

2. A case study in wheat has been completed.

When the genetic model is extended with linkage being present and the association between gene and marker is not perfect, the population genetics theory becomes intractable. A breeding module (QuCIM) that was developed in a previous project using the QUGene software (owned by Uniquist at the The University of Queensland) was used to determine the selected proportion in each selection stage and the optimum crossing and selection strategies.

In total, 9 marker-linked genes were considered in the case study: alleles at the *Rht-B1*, *Rht-D1*, and *Rht8* loci affect plant height, *Sr2* is an adult plant stem rust resistance gene, *Cre1* is a cereal cyst nematode resistance gene, *VPM* is an *Aegilops ventricosa* chromosome translocation carrying genes for leaf (*Lr37*), stem (*Sr38*) and stripe (*Yr17*) rust resistance, the *Glu-B1* and *Glu-A3* loci code for grain storage proteins, and *TIN* loci affects the tiller number. Molecular markers are 'perfect' for these genes except *Rht8* and *Sr2* where diagnostic markers are a small chromosomal distance from the respective gene, i.e. 0.6cM (recombination frequency $r=0.006$) for *Rht8* and 1.1cM ($r=0.011$) for *Sr2*. The *Rht-B1*, *Rht-D1*, *Rht8*, *Sr2*, *Glu-B1*, *Glu-A3* and *TIN* molecular markers are codominant, and *Cre1* and *VPM* markers are dominant in genic expression. *Glu-A3* and *TIN* are linked on the short arm of chromosome 1A with a distance of 3.8cM ($r=0.0366$). There is no linkage among the other seven genes.

Three parents were used, i.e., Sunstate, Silverstar/TIN, and HM14BS. The best strategies using marker based selection in genotype building for the single cross between Sunstate and Silverstar/TIN, and the top cross among the three parents have been studied and will be presented briefly at this year's annual meeting.

3. The QUCim breeding module is being modified to realize the linkage with physiological models.

We are currently modifying the simulation module QuCim/QuLine so that the simulation model and physiological model can communicate with each other. A new functionality called Plug-In will be added to QuCim/QuLine. A physiological model on leaf elongation rate in maize is under development as an example physiological model. In conjunction with GCP Comissioned project #7 ('-Whole Plant Physiology modeling of drought tolerance in cereals'), this analysis will be later extended in output 1 to include the leaf elongation physiology embedded inside the APSIM crop simulation module

From the population genetics model and simulation experiment, some general and specific rules regarding the efficient use of marker based selection have been proposed.

1. Based on the minimum population size of having at least one individual to be selected, we propose that the selected proportion should be not less than 0.005 for each selection stage to keep the population at a reasonable size for most crop species. With this restriction the

number of markers being selected should not exceed three if the target genotype is selected in F2 or seven if in doubled haploids (DHs) or recombination inbred lines (RILs). Up to 12 or 13 markers can be selected simultaneously if enrichment selection (i.e. selection for full set of alleles in either homozygous or heterozygous combinations) is applied in F2 prior to DH or RIL selection, although additional enrichment (say in F3 or F4) was of no value. In practice, if multi-stage selection is applied, employing a similar an evenly-distributed selected proportion for all stages will result in the smallest minimum population size, and therefore decreases the cost associated with selection.

2. If $3^{n_i} / 2^n > 1$ (n_i is the number of target alleles in parent P_i , where $i=1$ or 2 , and n is total number of genes in segregation in the two parents), a backcross with P_i as the recurrent parent will increase the proportion target genotypes; otherwise, no backcross is needed. When a top cross is made, the parent with the largest number of favorable alleles should be used last.

3. Where markers exist at different frequencies in the population or have different costs, the order of their selection should be such that the markers are selected sequentially based on the value of $\frac{c}{1-f}$ (where c is the cost per marker dataset, and f is the selected proportion for the marker). This will minimize both the number of marker screens and the cost of conversion.

4. Selection of the target genotype in F2 requires a substantially greater minimum population size, unless the frequency of the target allele in the F2 exceeds 0.6. When the frequency of the target allele exceeds 0.27, selection of the target genotype in DHs/RILs requires the smallest minimum population size. Otherwise, selecting the target genotype in DHs/RILs after F2 enrichment is the best strategy. When more than 3 genes are segregating, the frequency of enriched genotypes in F2 is less than 0.27. Therefore, in most cases, F2 enrichment selection followed by homozygous selection in DHs/RILs should be the best strategy, to minimize the population size for screening. Enrichment selection at two selection stages (in F2 and F3), never becomes the best strategy. Using F2 enrichment increased the frequency of selected alleles, allowing large reductions in minimum population size for recovery of target genotypes (commonly around 90%) and/or selection at a greater number of loci. So the gain from another enrichment selection in F3 after the enrichment in F2 is minor.

A draft research paper has been prepared on the work described above, i.e. the wheat marker-assisted selection case study.

Due to the short period for start-up, appointments and complex contract agreements, the project did not begin until July 2005. The main consequence of this is that the delivery time of outputs needs to be discussed with the GCP, as it simply is not possible for all outputs to be delivered in 18 months cf 24 months. I would like to discuss this at the annual meeting, as the options include dropping some of the outputs or rolling them into alternative projects.

Outputs/tasks 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.
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

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.
4. **For maize, demonstrate APSIM/QU-GENE link for existing QTL work on leaf growth.** In conjunction with commissioned SP1 project (CIRAD/de Raissac) a short report will be written on genetic analysis of the leaf elongation model.
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.
6. **Develop process and template to make licence agreements** between software owners (UQ/APSRU) and GCP participants (see below) to utilize existing and new tools.
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.

21. Planning for Effective Product Development, Delivery, and Use

Principal Investigator:

Victoria Henson-Appollonio, IPGRI

Collaborating Scientists:

Silvia Salazar, Consultant (Lawyer), Costa Rica

Maria Ines Mendosa, Consultant (Lawyer), Columbia

Shawn Sullivan, Consultant (Lawyer), USA

Zenete Franca, IFPRI-ISNAR

Rosemary Wolson, University of Capetown

Jocelyn Webster, AfricaBio

MID-YEAR REPORT

The principal investigators that will be involved in this project (“Unlocking the genetic diversity in peanut’s wild relatives with genomic and genetic tools”/EMBRAPA; “Drought Tolerant Rice Cultivars for North China and South/Southeast Asia by Highly Efficient Pyramiding of QTL’s from Diverse Origins”/CAAS, and “Revitalizing Marginal Lands: Discovery of Genes for Tolerance of Saline and Phosphorous Deficient Soils to Enhance and Sustain Productivity”/IRRI, have been contacted and asked to provide details of their projects so that we can begin to put development, delivery and use plans together in the next few months. In addition, we have been asked by the SP3 Leader, Dr. Jonathan Crouch to include the other SP3-commissioned projects in our study: “Development of low tech gene-based trait assay technologies in rice and wheat”/IRRI and CIMMYT; “Development of low tech gene-based trait assay technologies in rice and wheat”/IRRI; and “Simulation of marker-assisted selection strategies for optimizing molecular breeding systems for drought tolerance in cereals”/CSIRO. We are in the process of contacting these additional researchers.

In the past six months we have participated in a series of SP5-led events that coincide with the activities of this project as well. Victoria participated in an electronic forum that was looking at issues associated with product delivery and uptake by users. In addition Victoria participated in a workshop, organized by the SP5 Leader, Dr. Carmen de Vicente for the purpose of developing a GCP Strategy and a work plan dealing with enabling product

delivery. We developed draft guidelines for “Product Delivery Plans” that will be included in the proposal requirement, for the next GCP call-for-proposals, to be initiated later this year.

Tangible outputs delivered:

Draft Guidelines for Product Delivery Plans (for inclusion with new proposals).

We are very fortunate in that this project coincides very nicely with work that is ongoing under SP5. The budget that has been committed to this project for a workshop next year will facilitate the work in SP5 as well. We have not had our first meeting of institutional teams and this is a delay. However, the SP5 workshop has helped us to more clearly define the project and as a result we expect our institutional meeting to be more effective.

SP4 COMMISSIONED GRANTS

22. Development of GenerationCP Domain (Data) Models

Principal Investigator:

Richard Bruskiewich, IRRI

Collaborating Scientists:

Reinhard Simon, CIP

Manuel Ruiz, CIRAD

Tom Hazekamp, IPGRI

Masaru Takeya, NIAS

MID-YEAR REPORT

- Preliminary task discussions held with a small number of SP4 scientists at the Plant & Animal Genome meeting in San Diego. Some additional work undertaken on the original use case database prototype. Some discussions over email with other task leaders (e.g. Guy Davenport of the template task).
- The task was heavily kick-started at the Wageningen GCP meeting focusing on domain modeling (see http://cropwiki.irri.org/gcp/index.php/Wageningen_workshop_report). The five designated domain model editorial teams in the task (IRRI + four collaborators listed above) prepared a preliminary assessment of their respective modeling themes. This was refined during the meeting to establish Phase I modeling guidelines. A decision was also taken at the meeting to establish CropWiki documentation of the task (which is being elaborated on a continuing basis). This CropWiki site is also supplanting the functionality of the original Use Case database as well, with respect to use case development.
- During March/April, phase I modeling was undertaken by all editorial teams and largely published on the CropWiki (with some additional coordination over email).
- During the May IRRI hosted GCP platform workshop, further task discussions in all modeling areas was undertaken. In addition, to overcome the issue of UML tool interoperability and integration of the domain models with (Java) platform development, a technical decision was taken to consolidate all the domain models into EclipseUML and to develop the models as <<interface>> UML stereotyped specifications rather than <<class>> specifications.
- A CVS repository for these EclipseUML models was established and is being populated (the DEMETER-DSI subproject of the GCP middleware project in CropForge).

- This task is being effectively coordinated with the Data Template task efforts (with Guy Davenport at CIMMYT). MOBY support for domain models is partially prototyped into the Java GCP platform (at IRRI).
- A follow-up task coordination meeting is being convened on August 15th and 16th in Vancouver, BC, Canada, to assess task status, share task implementation experiences, and to ensure the proper integration of domain models with the development activities of related tasks/projects (i.e. data template, web services, repository and GCP platform).
- Reports on progress of specific domain model teams:
- Generic Core Models (Richard Bruskiewich): a generic meta-data domain model for global identifiers, entity and feature descriptions is designed for application to all editorial theme areas.
- Germplasm/Phenotype/Genotype (IRRI team+Jennifer Lee): ICIS GMS model has been extended and captured in UML to meet GCP needs; a Phenotype Ontology inspired by rice mutant database and ICIS phenotype trait ideas is designed (anchored on the generic domain models) and Genotype models are rapidly evolving in collaboration with Jennifer Lee (Germinate).
- Passport (Tom Hazekamp):
- The passport descriptors used by contributors to the 11 GCP Composite Collections were collected. This resulted in a set of more than 200 distinct passport descriptors. This set was used as the base for the development of the GCP passport domain model.
- Drafts of the GCP passport domain model were published on the GCP CropWiki for discussion and consolidation. A Phase I GCP passport domain model was delivered as input for the GCP-SP4 Platform Development Workshop held at IRRI from 10-20 May.
- External stakeholders were involved to obtain further inputs into the development of the GCP passport domain model:
- Drafts of the GCP passport domain model were shared with EURISCO partners.
- From 23-25 May a technical consultation on passport domain models and schema's was held at IPGRI with representatives from Institut fuer Pflanzengenetik und Kulturpflanzenforschung (Germany), Nordic Genebank (Sweden) and the Botanic Garden and Botanical Museum Berlin-Dahlem (Germany). The Botanic Garden is directly involved in the development of the ABCD schema within the Taxonomic Database Working Group (TDWG).
- Functional Genomics (Masaru Takeya):
- The GCP functional genomics domain model is being designed for the gene-based integration of experimental data, and the comparative biology/genomics. The structure of model is contrived to exchange data smoothly for public models such as MIAME MAGE-OM, MIAME/plant and HUPO PSI-OM.
- Current drafts of the domain model cover gene expression, proteomics, and mutants. The gene expression model consists of microarray experiment, SAGE and MPSS.

Tangible outputs delivered:

- CropWiki documentation site was established and is being maintained for domain model discussions at http://cropwiki.irri.org/gcp/index.php/Domain_Modeling
- Domain models files posted to CropForge CVS:
- Original project models at <http://cropforge.irri.org/scm/cvsweb.php/?cvsroot=gcpmodels>
- Latest models are integrated as the “DEMETER” subcomponent of GCP platform development middleware: <http://cropforge.irri.org/scm/cvsweb.php/Demeter-DMS/?cvsroot=gcpmiddleware>
- Ontology database is commissioned at <http://ontology.generationcp.org/>

- Active URL site for GenerationCP domain model publication of XML model specifications (and use as a URN) is established at <http://www.generationcp.org/model/> (but not yet populated with XML files).

Task objectives with a detailed step by step checklist of activities, task schedule and milestones were elaborated in the original task work plan. Actual implementation of the task has evolved somewhat differently.

Current Deviations from Task Objectives:

1. Use case compilation: although use case documentation guidelines are published (http://cropwiki.irri.org/gcp/index.php/Use_Case_Documentation_Guidelines), a use case database with interface commissioned (see http://cropwiki.irri.org:8080/GCP_UseCaseDb/) and although the CropWiki has been demonstrated to be a useful tool for use case documentation (see http://cropwiki.irri.org/gcp/index.php/Use_Case_Inventory), the task is lagging somewhat in systematic use case documentation.
2. Extend year 1 models: seems to be effectively underway but consolidation of efforts remains relatively inefficient at the present time (hopefully this will be resolved at the August GCP task meeting).
3. Commission a Community Editorial Process: although five (and a half²) editorial teams are well established and working on their models, again, the community process is poorly integrated and it is unknown how broadly each team is consulting the wider scientific community (and integrating their efforts with other GCP tasks).
4. Collaboration with Data Template Task: direct integration with data template task efforts is underway but not complete (however, the task 25 is closely coordinating his efforts with the task 22 task leader).
5. Commission software tools for integration of domain models with web services: prototyping of software is underway but accelerated progress awaits arrival of Martin Senger (MOBY expert) to GCP (hosted at IRRI) on August 1st.
6. Commission software tools for integration of domain models into GCP platform: EclipseUML consolidation of domain models into the platform is proceeding in a promising manner but is incomplete. A core (generic) meta-data model for the domain models and platform is now proposed but details of this core model remain to be successfully communicated to other editorial and GCP platform project teams for application to their respective activities. Again, the August Vancouver meeting should meet this objective.

Current Deviations from Task Outcomes:

1. Use case development: although the IRRI workshop in May clarified some use cases for prioritization (see <http://cropwiki.irri.org/gcp/index.php/PlatformDevelopment>), as indicated above, use case documentation needs to be significantly improved (on the CropWiki).

2. Domain modeling design and application methodology

At the February Wageningen meeting, a proposal to construct the domain models in Protégé/OWL RDF was shelved in favor of continued modeling using conventional UML modeling approaches. The UML approach has not been without some difficulties: it was discovered in March/April that interoperability of different UML drawing tools is very poor. These incompatibilities has not been completely resolved as several editorial teams still use

² Jennifer Lee is working on genotype models in collaboration with Thomas Metz at IRRI, even though she is not a distinct editorial team.

distinct tools, although a general decision was taken at the IRRI May workshop to consolidate all the models into one UML tool, EclipseUML (a task which remains ongoing). Domain model publication (in a GCP repository) requires further elaboration. Domain model “workflow” integration with data templates, MOBY web services and GCP platform development tasks/projects is progressing more slowly than originally anticipated (but progressing).

3. Model Repository, Populated with Year 2 models: repository components established (see “Tangible Outputs” above) but still poorly populated and integrated (status to be reviewed at the August meeting).
4. Application of Domain Models to GCP products: progressing on all fronts but tangible end user products are still forthcoming.

Current Deviations from Task Methodology:

Protégé abandoned in favor of UML for domain model development. Inventory and integration of external modeling standards is incomplete and requires review. Although the domain models are being published (in CropForge, CropWiki and elsewhere), further coordination and better consolidation of all documentation sites is required. Application of the domain models to GCP software components (templates, web services, platform software) is progressing but more slowly than originally envisioned. The status of the task methodology will be reviewed at the August Vancouver meeting.

Current Deviations from Task Schedule & Milestones:

Meetings at PAG, Wageningen and IRRI have all been held as proposed, with steady task progress; however, publication and integration of task editorial team outputs within a common modeling workflow remains inadequate. Accelerated progress over the coming two months is optimistically anticipated, however, as efforts in EclipseUML model integration and associated Java software tool engineering come to fruition and are effectively communicated to team members at the August Vancouver meeting. It will be critical at the proposed meeting to finalize a set of “spanning use cases” that are feasible for tangible delivery by the GCP annual meeting in Rome in late September.

Two GCP meetings – in Wageningen and at IRRI – were essential in ensuring steady progress in this task; task progress in defining GCP domain models is progressing steadily; however, given the technical challenges encountered so far in the work, better integration of the models and associated activities between editorial teams is required to ensure that task objectives are fully met. A third project meeting is now being convened on August 15/16, in Vancouver, BC, Canada, for this purpose.

Critical to success will be the efficacy in which task members can share and consolidate their ongoing efforts, work more effectively to improve documentation of use cases on the domain models, and apply their models more efficiently to tangible products that use the domain models – templates, web services and platform software - for deployment in a timely manner to end users within the coming months (i.e. at the GCP annual meeting?).

23. Implementation of Web Services Technology in the GCP Consortium

Principal Investigator:

Samy Gaiji, IPGRI/SGRP

Participating Institutions:

All Generation CP Consortium members
Other CGIAR Centres (CIFOR, WorldFish, World Agroforestry Centre, ILRI,
IFPRI), SGRP1, MOBY2, GBIF3, FAO4, EURISCO5, USDA-GRIN6, ICT-KM7, CSI8

MID-YEAR REPORT

The beginning of the year has been used to survey the Generation CP Consortium members in terms of software, hardware and system requirements. An important role has been taken by the evaluation of web services, in particular of the BioMOBY standard. There are a number of web service standards available, but MOBY, both because of its wide use in bioinformatics and because of its flexibility and modularity. We participated in a series of meetings to lay out the strategy for making available the data that is available in the Consortium. It was clear from the beginning that the development of the middleware would be a long term strategy and that there was the need to set some short term goals. The first one is to unlock the great amount of data that the Consortium members hold, the problem being that this data is in several formats and does not necessarily follow defined standards. This goal should be handled by the modeling task for standards, by repository task for dissemination and by the web services task for availability. A meeting was held in IPGRI bringing together experts on the DiGIR, BioCASE and BioMOBY standards with members of the Consortium. The main outcomes were a better understanding of the technologies and, most importantly, an agreement between the experts to integrate these technologies in the future to offer “the best of both worlds”. This is an important achievement that will give its fruits in the long run.

Regarding the deployment of web services, it was decided that a tool that would aid in deploying web services could be a short term solution to provide access to the Consortium data. The idea is to allow creating web services without having to program them from scratch, allowing members of the Consortium without specific technical knowledge to deploy these services. After a series of technical consultations it was decided to create a centrally deployed tool, rather than a personal one, integrating it into the Generation CP Repository. This approach would allow easy access to the Programme standards and provide a single access point for data consumers and providers. Currently we are developing a framework that will be the base for both the tool and for the Repository, this makes both tasks dependent upon each other. By the end of July a first prototype of the tool should be available to collect the work of the modeling task and to allow database managers to map their resources to these models. This first phase will identify which domain models are available through which data providers. A second prototype will add the ability to map BioMOBY objects to domain models. The final first version should add the ability to define BioMOBY web services and generate the necessary scripts to deploy them. Once the tool is tested, the actual deployment phase can begin by providing training to the data providers on how to set up the environment to host web services and map their database schemas to domain models, to domain modelers on how to load data models and to data “consumers” on how to create BioMOBY objects and use them to define web services.

Tangible outputs delivered:

- January 31st – February 4th: Technical consultation with Mathieu Rouard on BioMOBY implementation and technical issues.
- February 14th – February 18th: Participation in the Generation challenge programme meeting in Wagenigen, Netherlands.
- February 21st – February 25th: Participation in the ICIS Workshop 2005 in Roedmond, Netherlands.

- May 23rd – May 25th: Technical Consultation on Domain Models at IPGRI, Italy.
- June 13th – June 17th: Workshop on “Web Services, its technical fundamentals and future implementations” in IPGRI, Italy.
- A database to record information related to data providers, data sources, database schemas have been created. This database will become part of the central repository.

Tangible outputs expected:

- End of July 2005: A first prototype of the Model Mapper tool allowing database schema mapping to domain models should be available and on-line for testing.
- September 2005: A second prototype of the tool adding support for mapping BioMOBY objects to domain models and database schemas should be available for testing.
- End of October 2005: The first version of the Model Mapper Toolkit should be operative, this version should allow generating web services.
- November 2005: Deployment of web services at Generation CP Consortium member sites.

Originally the mapping toolkit was intended to be an installable software package to be deployed at each of the Consortium member’s sites. In consideration of the fact that the main reason for adopting web services is to allow data sharing and standards adoption, we thought that a tool deployed in a central site would be more advantageous and the Generation CP Repository was selected as the best candidate to host such a tool. The idea is to have at hand all the domain models, allow data providers to map their databases upon the domain models, allow domain model developers to create BioMOBY objects mapped to the domain models and finally allow web service clients to create web services that could be implemented by downloading scripts from the Repository.

This strategy solves two main issues:

- A tool that is deployed locally has difficult access to the Programme standards, also it does not aid in bringing together the scientists since it doesn’t put them in contact.
- Web services are useful to those who consume data, however the toolkit is an aid to publish existing data as web services: this means that both the service consumers and the service providers must have access to the tool.

By deploying the tool centrally those who need data can create web services that can then be “proposed” to the data providers, these can then take care of implementing and refining these web services according to the individual provider specific constraints. This way each party retains responsibility on their resources and is therefore able to better manage them.

This change of plans has direct consequences on the “Creation and maintenance of Generation CP Repository” activity specifically for the part where central repository has to provide support to the functionality related to the web services. This is also due to the fact that this tool will be a part of the Central Repository web site, The framework that is being developed for the tool will be the framework upon which the Repository will be developed: this means that delays in one may affect the other, but it also means that advances in one will benefit the other, specifically for functionality related to the web services.

The choice of deploying the mapping tool centrally might also change the deployment strategy in terms of training, data providers will have to be able to deploy the environment that will host the web service scripts, whereas the others will need training on a tool that is available on the internet, so for these there is no need to provide a personalized training.

24. Application of MOBY for GenerationCP Consortium

Principal Investigator:

Richard Bruskiewich, IRRI

Collaborating Scientists:

Natalia Martins, EMBRAPA

Mathieu Rouard, INIBAP

Shoshi Kikuchi, NIAS

Masaru Takeya, NIAS

Koji Doi, NIAS

MID-YEAR REPORT

- Preliminary task discussions held with a small number of SP4 scientists at the Plant & Animal Genome meeting in San Diego with some further discussions at the Wageningen GCP meeting.
- Some CropWiki documentation of the task is available (see http://cropwiki.irri.org/gcp/index.php/MOBY_Applications).
- Project PI (R. Bruskiewich) and one other task representative (Masaru Takeya, NIAS) attended a MOBY workshop held in Vancouver, BC at the beginning of May.
- The May IRRI hosted GCP platform workshop included participation of Martin Senger (EBI representative involved in MOBY development) and Markus Markus Döring, of the BioCASE/GBIF community, for a MOBY/ GBIF protocol comparison (see http://cropwiki.irri.org/gcp/index.php/Informal_GBIF_MOBY_Comparison) which was later elaborated at the Rome ICT/KM+GCP sponsored interoperability meeting. Other discussions relating to GCP platform development (e.g. ortholog gene catalog and gene expression repository) have highlighted MOBY web services in their near term workplans.
- IRRI programmers are continuing to work on integration of web services into the GCP platform (using the Java ICIS platform as a reference implementation). Significant prototyping of MOBY code in the GCP platform is completed.
- A new IRRI-hosted (Filipino) software engineer starts work on July 4th on the task of MOBY and related interoperability (e.g. GBIF) protocol integration. This individual will be joined by Martin Senger on August 1st (see below under “Deviations”) to accelerate GCP platform integration of MOBY web services.
- Collaborator Mathieu Rouard (INIBAP) reports the following for this task report: “...According to the objectives of this task, we decided to implement web services to specific GCP scientific analysis use cases. An internship, David Baux has been recruited to go further in the developments which have been initiated last year.

The use cases selected are still focused on the assessment of the allelic diversity (creation of a core collection) on Musa.

A web service interacting with the statistic software R has been done and produces a distance matrix from a complex disjunctive matrix.

Besides, A Moby service allowing selecting a germplasm list, from characterization and biotic stresses, has been created in order to provide an allele frequency distribution table. This service aims to answer the following use case available in the GCP Germplasm white

paper: Identify the genes or alleles that are involved in and responsible for the resistance of a germplasm to a specific pathogen.

Another Moby service is being implemented, and will perform an AFC (Correspondence factor analysis) with a distance matrix and will return the results in a diagram form (image file).

The use of these services is possible through a web client implemented in house and available on the inibap server but we are also exploring the use of Taverna. A new Moby specific plug-in, making easier the creation of workflows, is available and will be integrated in the next release of Taverna. We are expecting another plug-in, being developed at Evry (France), which will permit to generate web-based workflows...”

Tangible outputs delivered:

- Some basic CropWiki documentation of the task is available (see http://cropwiki.irri.org/gcp/index.php/MOBY_Applications).
- Some domain model teams have defined use cases ripe for implementation as MOBY web services (e.g. from Reinhard Simon, CIP: see list of features related to Location and Environment model on CropWiki at http://cropwiki.irri.org/gcp/index.php/Location_model#Features_Supported_by_this_Do_main).
- Mathieu Rouard, INIBAP reports 3 new Moby services developed and registered in the Moby Central
- in lieu of a subcontract to Dr. Mark Wilkinson of the University of British Columbia, Canada, task is funding Mr. Martin Senger to work as a consultant at IRRI from August 1st for a ~ 7 month period.
- From Reinhard Simon, CIP)
- Due to late release of funds (end of May in case of CIP) all activities at CIP were more or less affected. In this case, implementation of BioMoby services is pending and scheduled for later this year and early next year.
- Most task collaborators did not choose to attend the May MOBY workshop in Vancouver, thus missing out on a good training opportunity.

Documentation of target MOBY web services lag and require better integration (i.e. on the CropWiki) to GCP domain modeling activities.

25. Creation and Maintenance of Templates for GCP Data Storage in Repositories

Principal Investigator:

Guy Davenport, CIMMYT

Participating Institutions:

Agropolis, CIMMYT, IITA, IPGRI, IRRI, and SCRI

MID-YEAR REPORT

In January an attempt was made to collate all templates developed under SP4 tasks from 2004. We able to locate the Fingerprinting template for SSR data developed for SP1, however we not able to find out about the progress of other templates from 2004. A list of data types being produced by SP1, SP2 and SP3 was developed on CropWiki

(http://cropwiki.irri.org/gcp/index.php/GCP_Data_Type_Inventory). Through discussions with GCP funded scientists we developed a priority of list of data types for which templates should be developed in the first stage. These consisted of three data types, on which development was carried out by the groups as follows:

- Passport data – IPGRI - Thomas Hazekamp
- SSR Fingerprinting data – CIRAD - Brigitte Courtois
- Mapping and map data (including QTL data) – CIRAD - Manuel Ruiz

Priority was given to the development of templates for Fingerprinting and Passport data, since these were the most urgent. An updated template for SSR fingerprinting data was developed by Brigitte Courtois with input from other team members and published on the bioinformatics portal of the GCP website (www.generationcp.org/bioinformatics.php) at the end of June 2005. Development of a GCP passport data template was started by Thomas Hazekamp based on the GCP passport domain model being developed within Task 22 – Development of GCP Domain models. However, this template is not quite complete. Therefore a passport template based on the EURISCO passport descriptors was developed and published on the GCP bioinformatics portal at the end of June 2005. The GCP passport data template will be completed and published by the end of July 2005. Work on the Mapping and map data template was delayed and it is expected that it will be completed and published by the end of August 2005.

To provide validation of these templates and conversion of the data they hold into an XML exchange format, we started development of web-based tools. In order to be compatible with the GCP Platform these tools are being developed in Java and the XML format will conform to the domain model being developed within Task 22 – Development of GCP Domain models. A prototype of these web based tools was developed and demonstrated at the GCP workshop February 2005 in Wageningen. The proposed plan was to develop these tools further to provide additional functionality and support for the new templates. However, after Guy Davenport and Andrew Farmer attended the GCP funded Eclipse RCP/uDig workshop July 2005 in Lima it was decided a better approach would be to develop an Eclipse RCP client for the templates. This client could be developed relatively easily using many existing Eclipse plug-in modules, to provide a readily extensible platform independent client for the validation, visualization, editing, submission and conversion of GCP datasets. A plug-in for the Germinate database will also be developed using the existing Germinate Loader project currently available on cropforge. A prototype of this RCP client will be demonstrated at the GCP ARM in September.

An initial call was made to all GCP Task leaders and a second call to bioinformatics staff at GCP centres for information about how they store data. The response was not unanimous or particularly well defined, except in a few cases. The reason for the poor response was perhaps due to how the questions were phrased. Once all the responses have been completely reviewed a new call will be made in August, with a repeated request and an additional request for feedback on the published templates. However, phenotyping data has already been established as a priority for Phase 2 of development. In addition Andy Flavell's group will be developing models for the storage of SNP and DArT fingerprinting data.

An updated template for SSR fingerprinting data and a EURISCO passport data template was developed and published on the bioinformatics portal of the GCP website (www.generationcp.org/bioinformatics.php). In addition to the templates the download contains detailed information about how to fill the templates and example data.

Deviations from the work plan:

- Sarah Hearne replaced Francis Moonan for IITA.
- Tom Hazekamp replaced Raj Sood for IPGRI.
- Andy Flavell (University of Dundee) replaced David Marshall for SCRI
- Andrew Farmer from NCGR was consulted in the development of the templates for mapping and QTL data.
- The first development phase was delayed by three months and will now be completed at the end of August with the publishing of the template for Mapping and map data. The SSR fingerprinting data template was one month late and the passport data template will be two months late.
- An Eclipse RCP client for the templates will now be developed to provide a readily extensible platform independent client for the validation, visualization, editing, submission and conversion of GCP datasets.

26. Creation and Maintenance of Generation CP Repository

Principal Investigator:

Raj Sood, IPGRI/SGRP

Participating Institutions:

All Generation CP Consortium members

Agropolis, CIMMYT, IRRI, IITA, IPGRI/SGRP, ICRISAT, CIP

MID-YEAR REPORT

Feedback on requirements and design was obtained from Guy Davenport (CIMMYT). These refer specifically on requirements for the implementation of the GCP repository for data templates and suggested to use a platform based on Hibernate and a Postgres DBMS. The technical options for implementation are still under consideration.

The requirements and the design of the central repository heavily depend on the template task. The data templates are very crucial, as they will be used to load and retrieve data from the central repository. The use of templates for data upload/retrieval and making that data available to GCP partners are the key requirements of the central repository. In order to meet the above goals the following technological solutions have been identified and discussed with the template task leader.

CIMMYT is tasked to develop a web based tool which will be based on Java/RPC/Eclipse technology. This tool will parse data to and from templates and will establish links with other platforms.

The data in the central repository will be stored using the relational database and the application will have functionality to import and export data in XML format. The web based tool and the central repository will speak through XML. This will allow the java-based tool to interact with the central repository to retrieve and upload data in XML automatically using the templates.

The front end of the central repository will allow partners to view, and load new data sets in the central repository. And with the separate log in and password, GCP partners can view and retrieve all the available data sets from the central repository.

In February the collection of passport data on the 11 year 1 GCP composite crop collections started. The coordinator for each collection was contacted by email, telephone and sometimes through mediation of staff in IPGRI regional offices to obtain passport data on the GCP composite collections. To date passport data for 9 crops were obtained representing a total of 14,953 accessions.

In June a training workshop on Interoperability was organized by IPGRI which trained technical staff from all CGIAR GCP partners and others in web service solutions. Training sessions were given on BioMoby, BioCASE and GBIF solutions. This workshop generated a lot of enthusiasm with the trained CGIAR staff. To capitalize on this data collection activities for the Central Repository will focus on developing appropriate schema's for use with the BioCASE wrappers. The development of a schema based on the Template for GCP Data Submission for Passport data as developed for GCP Task 25 is underway. To build up a collection of core GCP data, the Central Repository will rely on these web service implementations in addition to the manual harvesting of data from GCP data providers.

Several Wiki software packages were evaluated as a platform for the Wiki-type environment. This environment will provide a collaborative environment for the publication and further development of data templates, data models, XML schema's stored at the Central Repository. MediaWiki software was selected as the preferred solution. A wiki application has been installed and is being customized. In coordination with the web services implementation task leader, the instructions related to installing and working with the web service will be created in this collaborative environment. The same "Wiki" will also provide necessary information related to the data upload and retrieval mechanisms used for the central repository.

Implementation GCP Repository for XML schema's, data models and data standards
In coordination with the web services implementation task leader, a database has been created to record the information related to data providers and data sources, and database schemas.

Tangible outputs delivered:

- Passport data for 14,953 accessions of the GCP composite collections were collected
- Wiki software installed and being customized for Central Repository
- A database to record information related to data providers, data sources, database schemas have been created.

Many of the design, implementation and content management activities for the Central Repository depend on inputs or deliverables from other GCP Tasks. These have their own development cycles and some of these are only just beginning to emerge at the end of the second quarter of 2005 (e.g. GCP templates for data submission). As a result design and implementation activities for the Central Repository have been moved further back in the year. To effectively cope with the shifting of activities additional personnel will be dedicated to Central Registry Tasks during the second half of 2005.

27. Integration of the High Performance Computing (HPC)-facilities in the Generation CP toolbox

Principal Investigator:

Anthony Collins, CIP

Participating Institutions:

CIP, IRRI, ICRISAT, IITA, CIRAD, ICARDA

MID-YEAR REPORT

The principal activities to achieve the objective are:

1. Creation and publication of appropriate documentation.
Structure running on the CIP HPC has been documented and tested within the CIP for Genetic Biodiversity Analysis.
2. Priority Tool Establishment
Support for customized use of BLAST has been provided for the SP2 COS activity by CIP. IRRI is working on gene microarray analysis use cases.
3. Development of Policy Management
Liaison with the CG IT Managers has led to agreement to standardize CGIAR Active Directory information by July 31, 2005, to facilitate the global Bioinformatics COP of HPC users by September 2005. This will simplify the system administration of and access to all the HPCs, while maintaining appropriate security.
4. Site activity Use Analysis, Surveying, and Expansion Roadmap Planning
Pending ... July - December 2005
5. Support.
CIP has contracted a full-time LINUX administrator who is in training to provide high-level "HELP DESK" support for all HPC sites
6. Promotion
This new LINUX Administrator is also assuming responsibility for consolidating the HPC web site, editing and compilation of documentation, and stimulating the COP relations
7. Capacity building
Following IRRI in 2004, CIP has connected its site to the global Internet2, and is testing a pilot Access Grid video conferencing facility with IRRI. This is anticipated as a step contributing to SP5 goals in 2005/2006.

Tangible outputs delivered:

- IRRI has launched the CropWiki for HPC, and a CropForge S/W repository to be used in common with the HPC programs is currently in implementation.
- http://cropwiki.irri.org/gcp/index.php/High_Performance_Computing
- <http://cropforge.irri.org> (note GCP HPC Structure)
- <http://hpc.cip.cgiar.org> (nextactivity is to consolidate CIP / ICRISAT / IRRI inputs)

No deviations from plan, but due to late arrival of funds, in fact we are only recently ramping up activities, hence the sketchy report. Expect good progress to report by the September annual meeting.

28. Improvement of Quality of Existing GCP Databases

Principal Investigator:

Graham McLaren, IRRI

Collaborating Scientists:

Guy Davenport, CIMMYT

Edwin Rojas, CIP

Fernando Rojas, CIAT

Chandra Subas, ICRISAT

Akinnola Akintunde, ICARDA

Visvanathan Mahalakshmi, IITA

Elizabeth Arnaud, INIBAP
Manuel Ruiz, CIRAD

MID-YEAR REPORT

Guidelines were established during the SP4 meeting in Netherlands in February for conducting base-line quality surveys of data in GCP repositories. These reports are to be compiled as a resource for developing a QA strategy for GCP data in August.

Also during the February meeting themes and structure were agreed for for platform development tasks. A call for sub-tasks was defined and published. During March and April, nine sub-tasks were defined. A developers' workshop was held at IRRI in May and progress on sub-task is indicated:

28.1 General Platform Architecture – G. Davenport

A developers' workshop was held at IRRI during May. The model implementation sub-tasks defined the overall platform architecture and specified technologies for middleware implementation.

http://cropwiki.irri.org/gcp/index.php/General_Platform_Architecture_WR_IRRI_May_2005

28.2 Implementation of middleware for germplasm, genotype and phenotype domain models – R. Bruskiwich

A Web-based querybuilder has been designed and prototyped for the CGP platform which will utilize the passport, germplasm and phenotype middleware which is being implemented.

http://cropwiki.irri.org/gcp/index.php/Query_Builder

28.3 Implementation of middleware for passport domain model – M. Rouard

First generation UML documentation of passport data model completed. Implementation commenced. http://cropwiki.irri.org/gcp/index.php/Passport_Domain_Module

28.4 Implementation of middleware for location and environment domain models – I. Mukema

The location model has been defined as part and relevant Java APIs generated.

http://cropwiki.irri.org/gcp/index.php/Location_model. Java libraries and tools reviewed for GIS processing and uDig, based on Eclipse RCP and Geo Tools has been selected for implementation of the location model.

28.5 Implementation of middleware and interface for mapping domain model – M. Ruiz

The mapping domain model was completed and published as an Eclipse UML document and a module prototype was developed in Java using the strategy developed at the IRRI workshop to answer some basic use cases. The prototype allows marker, map and gene searches using a generic mapping API, the mapping model in the middleware and object-relational mapping to CMAP and GBrowse repositories via Hibernate.

28.6 Mondrian data warehousing for GCP data – E. Rojas

Presentations and training was given on the Mondrian technology at the IRRI workshop

http://cropwiki.irri.org/gcp/index.php/Mondrian_Data_Warehouse_WR_IRRI_May_2005

and a prototype datamart for passport data was designed

http://cropwiki.irri.org/gcp/index.php/Mondrian_Data_Warehouse and implemented

<http://web.riu.cip.cgiar.org:8080/passport/>

28.7 LIMS web applications – sample tracking – B. Jayashree

The AGL-LIMS is under testing and is presently being used for two on-going GCP SP1 activities: genotyping of the composite germplasm set of chickpea and the molecular characterization of groundnut. CIMMYT has been provided with access to AGL-LIMS for testing.

28.8 LIMS: further development of GMLIMS and IGGeMS – Akintunde, A
Continued development of GMLIMS and IGGeMS following demonstration and feedback in February.

28.9 Integration of the CMTV application into the GCP platform – G. Davenport
This task was integrated with task 28.1

28.10 Adaptation of functional genomics tools for GCP platform – R. Bruskiwich
A workplan meeting was held at IRRI in May where partners agreed on work going forward
http://cropwiki.irri.org/gcp/index.php/Standalone_Tool_Adaptation_for_Functional_Genomics_WR_IRRI_May_2005

A data model has been adopted for sequence data and a prototype entry form to support the Ortholog Gene Catalogue has been developed
http://cropwiki.irri.org/gcp/index.php/Ortholog_Catalog_May-Sept_2005_Workplan

Tangible outputs delivered:

- Comprehensive platform architecture agreed and published on CropWiki
- Eclipse UML modeling tool agreed and deployed to generate Java interfaces
- Individual task plans, designs and progress described on CropWiki
- Eclipse RCP tool adopted and training delivered for wrapping JAVA based platform applications
- Prototype GCP Mondrian data warehouse developed
- Prototype Querybuilder interface to Phenotype data model developed
- LIMS systems enhanced at ICRISAT and ICARDA
- Prototype sequence interface and data model implemented

Deviations from the work plan:

Most participants indicated slower than expected start-up due to slow arrival of funds and/or staffing and recruitment delays.

Integration of sub-task elements has been hampered by the necessity to wait for agreed preliminary domain models and technology.

29a. Creation of Institutional Bioinformatics Capacity (CIAT)

Principal Investigator:

Joe Tohme, CIAT

Co-Principal Investigator:

Mathias Lorieux, CIAT

MID-YEAR REPORT

The project aims to establish a system that can compile and analyze through high throughput tools the data generated in studies of genetic diversity assessment, germplasm characterization, gene expression and QTL analysis, breeding and marker assisted selection. To achieve this we will develop:

- Pipelines in different fields to perform automated analysis.
- A laboratory management system (LIMSYS) connected to pipeline analysis tools and specific analysis plug-ins.
- Transference of knowledge to assistant researchers and students in the different projects.
- Implementation of the developed tools by users in CIAT for specific projects.

Expected Outputs

- Development of high-throughput analysis pipelines for genetic diversity assessment, germplasm characterization, gene expression and QTL analysis, breeding and marker assisted selection (In Development)
- Improvement in the efficiency of time and quality for mass data analysis.
- Generation of new tools and capacities to link data of diverse projects (MAS, gene expression, comparative genomics). (In Development)
- Generation of qualified researchers able to create high-throughput analysis pipelines (master students) and to perform in-depth bioinformatics analysis(users). (The pipeline for Cassava EST's is done)
- High-quality competitive publications generated using the bioinformatics resulting tools. (In Development)
- Use of pipelines and bioinformatics resources to generate capacity enhancement among research institutes and groups through extensive collaboration.

Deviations from the work plan: None.

29b. Creation of Institutional Bioinformatics Capacity (CIMMYT)

Principal Investigator:

Guy Davenport, CIMMYT

MID-YEAR REPORT

29c. Creation of Institutional Bioinformatics Capacity (CIP)

Principal Investigator:

Reinhard Simon, CIP

MID-YEAR REPORT

Main outputs related to CIP (see Task 29 proposal):

- (A) Use case identification and prioritization for SP4 and HPC
- (B) CIPPEX prototype
- (C) Eclipse RCP evaluated and advanced tutorials provided
- (D) IntiMap upgraded
- (E) Additional functionality for molecular marker/data analysis based on ERCP and DIVA-GIS code basis
- (F) Bioinformatics workflows / COS discovery / primer design
- (G) Further integration of R web services/BioMoby web services
- (H) Trained staff and trainees

1. In support of the COS activities at CIP available information will be organized around metabolic pathways. The MetaCycle database and software was identified as a promising infrastructure. Personnel was identified to collect and organize information. Work will be started in July 2005 but detained due to unexpected technical difficulties with MetaCycle software.
2. In support of deployment of varieties processing of satellite images will be included in DIVA-GIS software along with integration of HPC.
3. Commodity tools to handle BLAST searches for use in gene expression data analysis are being applied and will be organized in repeatable workflows.
4. STABLE software ported to R
5. GENOMA from EMBRAPA will be installed on HPC.

CIPPEX functionality is further refined as it is being used and reviewed by breeders and molecular biologists. Database and experimental management were further improved: The database structure for the molecular part is a complete new design- however, inspired by concepts from GERMINATE and ICIS. Presently, the system stores also weather data from CIP's weather station obtained semi-automatically.

Eclipse RCP evaluated and advanced tutorials provided:

Mainly done; see report on Platform development. Written tutorials will be posted on CropForge. An e-conference bases tutorial is planned for 2nd half of 2005.

IntiMap is a wrapper program around MapPop to provide a Graphical User Interface and draw simple maps of end-results (PAG XI-C13 & PAG XI-P894).

To be done (before September ARM).

Additional functionality for molecular marker/data analysis based on ERCP and DIVA-GIS code basis to be done (partially before ARM in September).

During the workshop at IRRI I realized that the EMBRAPA tool Genoma already very much realizes this functionality. Discussing with EMBRAPA staff we realized our mutual interest to test a version on the HPC. This will probably done in coordination with CIAT staff.

CIPSTAT: partly based on advances from web-services funding by IPGRI in 2004; now continued under 29 and 23. Responding to user requests in the context of CIPPEX development. Several new routines were added. BioMoby web services still need to be implemented.

1 trainee in bioinformatics was trained on COS discovery and HPC.

1 trainee in statistics and genetics (partial re-implementation of Structure in R).

Tangible outputs delivered:

- 5 use cases identified (4 involving use of HPC)
- 2 trainees
- CIPPEX further enhanced (XX new features)
- New R- web services implemented
- Eclipse RCP tutorial provided (see also 28.1)

Due to late release of funds (end of May in case of CIP) all activities at CIP were more or less affected. In this case, some activities were not yet initiated while others are only partially implemented.

In order to speed up development across GCP related activities and make up for the lost time I initiated purchase of new hardware under this activity both for improved data collection in the molecular lab and speedier development of Java programs on new hardware.

29d. Creation of Bio-informatics Capacity for Central and West Asia and North Africa

Principal Investigator:
Murari Singh, ICARDA

MID-YEAR REPORT

With an objective to develop Bioinformatics capacity at ICARDA for promoting the development and use of state-of-the-art bioinformatic tools for all crop improvement and biodiversity conservation programs and to support its Integrated Gene Management megaproject research activities as well as to develop capacity building in the NARSs of CWANA under the GCP, the following activities were undertaken:

During Jan- June 2005, ICARDA recruited two programmers for database development. Four staff members (M. Baum, A. Akintunde, Kamel Chabane, K. El-Shamaa, H. Abed and M. Singh) participated in a number of professional meetings and workshops at San Diego, Brisbane and Wageningen, and gained knowledge in the areas of genomic databases, search tools and SSR-EST and Inter-operatability. A Perl program was developed for data management for use in Popgene software. For identifying expressed-genes from micro-array experiments, we have derived expressions for threshold of extremes based on control RNA expression data.

We facilitated and assisted a Ph. D. student from Ethiopian NARS in completing data management, biometrical analysis of his four experiments on Tef. During his three months of stay at ICARDA, the student was assisted with computing tools, assistance on data management, use of in-house developed Genstat modules to analyse data for evaluating inbred lines, and various analyses based on genotype x environment interaction, genotypic correlations, selection index, path analysis, MapQTL to identify QTLs on data from his one experiment.

In a microarray study, the 22K Affymetrix GeneChip Barley 1 array was used to monitor changes in the transcription levels under drought stress of two barley cultivars, drought tolerant and sensitive, respectively., The microarray data were analyzed with the software of

Affymetrix GCOS1.2 and HarvEST barley 1.34. The preliminary results showed that 77 genes were significantly differentially expressed in both varieties through comparison analysis of gene expressions under stress and non-stress condition. These 77 genes were expressed in both varieties when plants were exposed to water stress. Therefore, they are likely to be genes responsive to drought stress and not important in drought tolerance. When gene expression was further analyzed, 372 genes were identified to be significantly differentially expressed between two varieties under drought stress. Those genes with known function were classified into 15 different functional categories in biological process. Some of these genes are known genes related to drought tolerance, while the others are unknown function or novel genes which may be involved in drought tolerance.

Tangible outputs delivered:

- Perl-program for data management for Popgene software.
- One scientist from Ethiopia-NARS trained personnel in biometric analysis of Tef experiment and QTLs for Tef traits
- Expressions of confidence limits for threshold of extremes based on control RNA expression data from an Affymetrix micro-array experiment.

Deviations from the work plan: None.

29e. Creation of Institutional Bioinformatics Capacity at ICRISAT

Principal Investigator:

Subhash Chandra, ICRISAT

MID-YEAR REPORT

- Thirteen staff and students trained in the use of AGL-LIMS (Applied Genomics Laboratory –LIMS)
- 7 postgraduate students trained in sequence analysis and database management systems
- 2 doctoral students trained in sequence analysis and annotation, primer design
- Regular support provided in data management, data format conversions, mining of public databases, effective querying of databases, experimental design and data analysis of genomic studies.
- Following systems/databases are being maintained for the benefit of genomics research community: Inventory Management System, EST-SSR databases, Genomic SSR databases and an Environmental Stress Transcripts database. The Inventory Management System automates inventory issue and maintenance, is accessible to all laboratory users from the intranet, and permits the store manager to keep track of stock. The system is regularly updated and maintained. The SSR databases contain updated information on SSRs mined from public collections of ESTs from three cereal crops and 4 legumes. Annotation information is regularly updated, the database is accessible on the intranet and the feedback form allows users to enter information on the usefulness of the marker in the laboratory/crop used. The database on Environmental Stress Transcripts is being developed. It contains information from environment stress libraries across 11 crops and is a useful resource to researchers in gene mining and comparative studies.
- Bioinformatics staff recruited:
One Scientist, One Scientific Officer, One Programmer, two Junior Programmers

Deviations from the work plan: None

29f. Creation of Institutional Bioinformatics Capacity

Principal Investigators:

Guy Davenport, CIMMYT
Graham McLaren, IRRI
Joe Tohme, CIAT
Murari Singh, ICARDA
Simon Reinhard, CIP
Samy Gaiji, IPGRI
Subhash Chandra, ICRISAT
Visvanathan Mahalakshmi, IITA (BECA)
S.Hearne, IITA (BECA)

MID-YEAR REPORT

The final contract for GCP Commissioned Research was sent to IITA on August 23rd 2005, however previous to this, discussions were held among IITA scientists, BECA and ILRI on how best to strengthen bioinformatics capacity. Two areas of priority emerged from these discussions:

1. The need for a standardized LIMS system in operation in both IITA-Ibadan and IITA-Nairobi. Under SP1 significant data has been generated at BECA through genotyping (3000 cassava accessions for 24 primers, and 2800 cowpea for 17 primers). DNA extraction was done in Ibadan with genotyping at BECA. This highlighted the need for a standardized LIMS system across locations. ICRISAT, having genotyping operations both at BECA and headquarters in India are apparently facing a similar need. The possibility of installing and training staff at BECA and IITA-Ibadan on the LIMS system developed by ICRISAT was discussed with Dave Hoisington. Dave agreed that this would be possible and could be funded from IITA and/or ICRISAT funds under this module. This however still needs to be discussed further with ILRI and BECA. Dr. Sarah Hearne (IITA-Nairobi) who has done most of the genotyping at BECA will be instrument in getting this system operational. Part of her employment costs are thus charged to this project.

To clarify PI and collaborator issues, since most of the high-throughput genotyping is done at IITA-Nairobi using the BECA platform, there appears to be a greater need or demand for a LIMS system at this location, therefore it has been agreed that Morag Ferguson will take over from Visvanathan Mahalakshmi as PI for this project. Mahalakshmi will remain as the main collaborator at the Ibadan location.

Tangible outputs delivered: None.

Deviations from the work plan: None

29g. Creation of Institutional Bioinformatics Capacity (IPGRI)

Principal Investigator:

Samy Gaiji, IPGRI/SGRP

MID-YEAR REPORT

29h. Creation of Institutional Bioinformatics Capacity

Principal Investigator:

Morag Ferguson, IITA

MID-YEAR REPORT

- At IRRI Task 29 has supported the positions of two NRS bioinformatics staff as well as bioinformatics training for our staff and resource material. Bioinformatics training has 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.
- 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.
- Graham McLaren attended the course on Computational and Statistical Aspects of Microarray Analysis in Bressanone, Italy from June 19th-24th 2005.
- Emily Deomano is attending 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.
- Subscriptions to three journals were also taken out under this project: Journal of Bioinformatics, Briefings in Bioinformatics and Applied Bioinformatics.

Deviations from the work plan: None

30. Development of Decision Support Systems for Sampling Germplasm

Principal Investigator:

Xavier Perrier, Agropolis-Cirad

Participating Institutions:

CIRAD-Agropolis, IPGRI, WUR

MID-YEAR REPORT

1.1 Cluster analysis with dependent data

dissimilarity measure with a vector of weights on markers depending on the map position

- R/S-plus dynamic link library: 'MAPweights.dll'

1.2 Joint analysis of molecular and phenotypic data

1.2.1 Weighted analysis

Define a dissimilarity on sets of variables of different nature (molecular markers, DNA sequence, morphological characters...) and different type (ordinal, nominal, binary...) in weighting each set .

Refinement on weight definition in discussion

- R/S-plus functions Daisy and Wgtdaisy
- Procedure to add several weighted dissimilarity matrices in Darwin software (in development)

1.2.2 Phenotypic under marker constraints

Clustering on phenotypic characters but under constraint of genetic structure inferred from molecular markers to control effects of homoplasy and convergence.

- Procedure NJtree under topological constraints in Darwin software.

2. 'Structural' LD free sub sampling

2.1 'Structural' LD concepts

- Comparison between linkage disequilibrium and structure disequilibrium
- Developments and tests of structure disequilibrium measures
- Problem of multiallelic loci with rare alleles
- Problem of diploids when phases are unknown

2.2 Sampling strategies

- Maximum length tree strategy: stepwise algorithm to prune a diversity tree of more redundant accessions
- Minimum structure disequilibrium strategy: stepwise algorithm to remove accessions of greatest contribution to disequilibrium
- Validation by iterate random sub sampling
- Procedures for the two strategies in Darwin software (only for homozygous). First versions available but require refinements on disequilibrium measure and some algorithmic improvements for large data sets. Extension to heterozygotes with unknown phase to develop.

Deviations from the work plan: None

31. Development of Ortholog-Function Display Tools

Principal Investigator:

Richard Bruskiewich, IRRI

Collaborating Scientists:

Kimmen Sjölander, University of California at Berkeley

Brigitte Courtois, CIRAD

Manuel Ruiz, CIRAD

Christophe Perin, CIRAD

Mathieu Conte, CIRAD

Masaru Takeya, NIAS

B. Jayashree, ICRISAT

Natalia Martins, EMBRAPA

MID-YEAR REPORT

- Preliminary task discussions held with Manuel Ruiz and Mathieu Conte of CIRAD at the Plant & Animal Genome meeting in San Diego. Mathieu traveled to the UC Berkeley lab of task collaborator Kimmen Sjölander for discussions on the ortholog gene analysis pipeline.
- A recruiting drive was initiated for the task-funded IRRI-hosted post doctoral scientist but not yet successful (see in “Deviations from the Workplan” below).
- Further informal discussions were undertaken between Manuel and NIAS representatives during the Wageningen GCP meeting in February.
- All collaborators on the task (save Dr. Sjölander) were present at the May IRRI hosted GCP platform workshop where further planning and design progress was achieved and documented on the CropWiki within the context of the GCP task 28 platform project for functional genomics standalone tool adaptation (see http://cropwiki.irri.org/gcp/index.php/Standalone_Tool_Adaptation_for_Functional_Genomics_WR_IRRI_May_2005#Comparative_Stress_Gene_Catalog). Various components of the task were delegated to task members (see http://cropwiki.irri.org/gcp/index.php/Ortholog_Catalog_May-Sept_2005_Workplan).
- Guided by the data template task leader (G. Davenport), a local University of the Philippines M.Sc. candidate computer scientist student is working at IRRI on data template development for the compilation of sequences for the comparative gene catalog. IRRI bioinformatics staff are deploying a GMOD Chado database for the stress catalog. Available public domain software for ortholog gene family tree and multiple sequence alignment display (ATV, JalView) are being reviewed for adaptation to the GCP platform interface for the catalog.
- CropWiki documentation site was established for comparative gene catalog discussions at http://cropwiki.irri.org/gcp/index.php/Standalone_Tool_Adaptation_for_Functional_Genomics_WR_IRRI_May_2005#Comparative_Stress_Gene_Catalog
- Postdoctoral Scientist
Recruitment of the project postdoctoral scientist (at IRRI) has been significantly delayed. A very small set of postdoctoral applications were received at IRRI for the GCP position and short-listed. One short-listed candidate withdrew from the competition. A second short-listed candidate was interviewed over the telephone but found deficient (by the three IRRI scientists doing the interview). Interviewing of a third shortlisted candidate awaits the return of members of the IRRI selection committee back to IRRI (from home leave).
- First prototype database for comparative gene catalog

Not yet commissioned. Domain modeling and data template design for required gene data is ongoing but incomplete.

32. Development of Crop gene Expression Database and Data Mining Tools

Principal Investigator:

Shoshi Kikuchi, NIAS

Collaborating Scientists:

Richard Bruskiewich, IRRI

Hei Leung, IRRI

MID-YEAR REPORT

Our goals of this project theme are

1. Establishment of the user-friendly gene expression database of the crops.
2. Data among crops are connected by the linkage of orthologous genes.
3. All the data are accessible through WEB service. As described in our proposal.

For these goals, one Post-doc researcher (Dr. Koji Doi) is hired from the beginning and one assistant for him will be hired. And for the bioinformatics and computer work, we have made contract with Hitachi Software Engineering Co., Ltd. With this company we have been working together for the establishment of RMOS, KOME, and RED databases.

For the establishment of new “Gene expression database”, we have taken the way of the enlargement of currently existing “Rice Expression Database (RED)” database rather than starting from de novo one. RED database is the microarray database based on the data of 9k-cDNA (EST) array, which was established in the Japanese Rice microarray project (1999-2002).

Microarray database has two major components, one is the information of the probed genes and the other is the information of target RNA samples. Currently, by the collection of rice full-length cDNA clones, nearly 30,000 rice expressed genes have been collected and incorporating the gene annotation data presented by TIGR, 62,000 unified-TUs (transcription units) on the rice genome (TIGR Pseudomolecules rel.3) are fixed. Within these TUs, 9k cDNA-based probes, 22k rice oligoarray probes (currently used among many researchers) and coming 44k rice new oligoarray probes will be covered. For the information of target RNAs, we will follow the MIAMI /Plant (Minimum Information About a Microarray Experiment) standard. As for data deposition to the database, all the information on the RNA samples will be described following this standard.

After search of available gene expression data in several databases such as NCBI-GEO, Array Express in EBI, many kinds of gene expression data from Arabidopsis are opened to the public but other the data from other plants even from rice are very rare. These are the original plan to the project and the content was presented in the Wageningen meeting, Feb 2005.

For the data mining process of the gene expression data, useful annotation data of probed genes, such as the mapping information to the genomic DNA, cis-element information in the upstream promoter region, gene expression data in other condition are necessary. It is meaningful to find hidden correlation among these characteristics. If some cis-elements

frequently appear in the upstream region of the same genes, and those genes show similar expression pattern, it would be efficient to clarify function and role of genes as well as those of cis-elements. We have employed two methods for this purpose; correspondence analysis and association analysis. Correspondence analysis is useful to visualize correlation, while association analysis achieves exhaustive examination of correlation between characteristics, with checking their confidence level. We have tried to examine relationship between cis-element existence and gene expression pattern for genes. The former data were from KOME database, and the later are categorized into 500 clusters by K-means clustering method from in-house data. Preliminary analysis showed some unexpected correlations between expression patterns, for which more detailed examinations are planning. Arabidopsis and rice are the model cases for these kinds of data. Then we have tried the clustering analyses of the gene expression profiles and the existence of the cis-elements in the clustered genes set. These data will be shown in the new Gene Expression Database.

In the middle of May 2005, the Platform-engineering workshop was held in IRRI. From our team, two post-docs (Doi and Satoh, who is related to the SP-2 project), one system engineer (Kohji Suzuki) from Hitachi Software Engineerings Co., Ltd and me attended. At the meeting we have presented the progress of the work related to the project and discussed with the schedule until the annual meeting in Rome (Sept-Oct, 2005). At that discussion, our internal task and external task were discussed. As for our internal work, we will establish the pilot version of the new Gene Expression Database in the end of August, 2005 and from the beginning of the August, along with the visit of Dr. M. Senger to IRRI, we will prepare for the WEB based remote communication system.

After returning from IRRI, we have started the establishment of new Gene Expression Database. As a first step of the progress, 62K universal TU on rice genome was fixed. Supported by the Japanese rice genome budget, we have performed the validation experiments of a 44k x 2 custom oligoarray with five kinds of RNA samples and the probes for the final array (44k) are nearly fixed. These data are strongly related to the gene expression database and the annotation of rice genes.

Tangible outputs delivered:

Japanese Society of Plant Physiology Meeting (March, 2005 in Niigata)

Koji Doi, Kouji Satoh, Shigemi Iizumi, Setsuko Kimura, Hisako Ooka, Kimihisa Tasaki, Hitomi- Akiyama Yamada, Jung-Sook Lee, Shoshi Kikuchi “Cluster analysis of the gene expression data by the oligoarray system in callus formation or regeneration process of rice.” *Plant Cell Physiol.* 46, Suppl. s141.

Deviations from the work plan:

- Not so many gene expression data are currently available in the public database as described in the report. So we have focused in the first project year (2005) to the in-house data.
- Because of the reason 1) and the orthologue linkage among many crops which have been studied by the CIRAD team in SP-4 project is mainly focused to the specific gene sets, such as GRAS family genes, which is known after discussion in the workshop in IRRI (May 2005). Bridging the data among crops will be focused in the second year of the project.

33. Development of an Integrated Decision Support System for Marker-assisted Plant Breeding

Principal Investigator:

Subhash Chandra, ICRISAT

Collaborating Scientists:

J-M Ribaut, CIMMYT

CG McLaren, IRRI

AE Melchinger, University of Hohenheim, Germany

FA van Eeuwijk, Wageningen University

H Mohanty, University of Hyderabad, India

JH Crouch, CIMMYT

MID-YEAR REPORT

The goal of this two-year project (2005-06) 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 journey from phenotyping-and-genotyping of genetic entities to the 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 a logical sequence of nine activities. These are: (1) Analyze potentially useful free software for quality and reliability, (2) Select software for inclusion in iMAS, (3) Develop iMAS system, (4) Develop online decision guidelines, (5) Test iMAS system, (6) Refine and release iMAS system, (7) Develop iMAS user manual, (8) Training in use of iMAS, and (9) Consultation and support for iMAS.

An email survey was conducted to elicit information on the freely available software used by and/or known to molecular breeders for (a) Experimental design for phenotyping, (b) Biometric analysis of phenotyping data, (c) Linkage map construction, (d) QTL analyses, (e) Association mapping, (f) Sample size determination for for-/back-ground selection, and (g) Estimation and display of recombinant genome content of selected individuals. The results of this survey were circulated to all respondents and all others concerned. Based on the survey results, a set of 12 softwares was identified (at Wageningen in Feb) across the eight tasks (a)-(g) for assessment of their quality and reliability by project partners (ICRISAT, IRRI, CIMMYT) using well-defined assessment criteria. The detailed reports of these assessments have been posted on the GCP website.

Project partners met at CIMMYT in early June to select the software for inclusion in iMAS based on assessment reports. It was agreed to include the following software in iMAS: IRRISTAT for tasks (a), (b) and (e); GMendel (and possibly MapManager QTX and MapDisto) for task (c); PlabQTL and Win QTL Cartographer for Task (d); Tassel (and possibly Structure) for task (e); PopMin for task (f); and GGT for task (g). This meeting at CIMMYT also reviewed and finalized the plan of work for

The development of the iMAS system is in progress according to the plan agreed among the project partners in the early June meeting at CIMMYT.

The text of online decision guidelines is under preparation as per the following plan: IRRI for IRRISTAT; ICRISAT for GMendel, PlabQTL, WinQTLCart, Structure and Tassel; and CIMMYT for PopMin and GGT.

An email survey was conducted to elicit information on the freely available software used by and/or known to molecular breeders for (a) Experimental design for phenotyping, (b) Biometric analysis of phenotyping data, (c) Linkage map construction, (d) QTL analyses, (e) Association mapping, (f) Sample size determination for for-/back-ground selection, and (g) Estimation and display of recombinant genome content of selected individuals. A summary report of the results of this survey has been circulated to all concerned.

Based on the survey results, a set of 12 softwares was identified (at Wageningen in Feb) across the eight tasks (a)-(g) for assessment of their quality and reliability by project partners (ICRISAT, IRRI, CIMMYT) using well-defined assessment criteria. The detailed reports of these assessments have been posted on the GCP website.

The software development for the iMAS system has been finalized in the early June meeting at CIMMYT.

Deviations from the work plan: None

34. GenerationCP Use Case and Software Engineering Collaboration and Management

Principal Investigator:

Thomas Metz, IRRI

MID-YEAR REPORT

Two web-based collaborative development systems, one for the development of software, and one for the development of textual content were installed and tested locally at IRRI in January 2005. The systems were presented, discussed, and accepted during a SP4 workshop (February 2005, Wageningen).

The first system is a Wiki, based on the free, open source MediaWiki software that was developed for the Wikipedia project (<http://en.wikipedia.org>).

The second system is a free, open source platform for software development (<http://www.gforge.org>), which is based on an earlier version of the now proprietary SourceForge (<http://sourceforge.net>) software.

Since the initial deployment of the systems there has been a steady growth of users and content. Both systems have created substantial interest and are already being used outside the direct scope of the GenerationCP.

A request has been made (21/04/2005) to the CAS-IP office to develop disclaimer and terms-of-use pages for the collaborative sites, and propose an appropriate license under which the GCPWiki content should be released. This important part of the deployment of collaborative workspaces is still outstanding.

In response to an interest expressed by CIP, IRRI and CIP are in discussion to share the responsibility, management, and support for the CropForge collaborative software development platform. The plan is to move CropForge to a better, dedicated server, have a neutral domain name (<http://cropforge.org>), and get commercial support for maintaining, upgrading, and trouble-shooting. This move is seen as a good starting point to provide

hosting of software development projects to a wider agricultural research community, also beyond the immediate scope of the GenerationCP.

From a technical infrastructure perspective, most of the objectives of this project have already been achieved. The initial attention to deploying the technical infrastructure is now switching towards widening and deepening the use of the collaborative systems. User training and support, as well as wider adoption and systematic use, will require continued efforts and resources, also beyond the current project period.

A web-based system for collaborative software development (<http://cropforge.irri.org>) is in use since February 2005. Currently (04/07/2005), there are 32 hosted projects and 73 registered users. Many, but not all of the projects currently hosted on CropForge are GCP related. Some of the software projects are in the beginning stages, but others were ongoing projects that moved their source code management and other support and discussion to the CropForge server. Each project has its own administrator who is responsible for the project content, project resources configuration and project team management.

A web-based system for collaborative development of textual content (<http://cropwiki.irri.org/gcp>) is in use since February 2005. The site is only accessible to registered users. Currently (04/07/2005), there are 1188 total pages, discounting for user pages, "talk" pages, pages about GCPWiki, minimal "stub" pages, redirects, and others that probably don't qualify as content pages, there are about 255 pages that can be considered proper content pages. There have been a total of 16647 page views, and 2451 page edits since the wiki was setup. That comes to 2.06 average edits per page, and 6.79 views per edit. There are 88 registered users. There are 8 users from different institutes (IRRI, IPGRI, CIMMYT, GCP office) with administrator roles. The content of the site is mostly limited to SP4 according to the proposal, but there it could be used by other sub-programs as well.

Project 34 has contributed to and supported two SP4 informatics workshops. During the first workshop (February, Wageningen), the collaborative systems were introduced and used by a few early adopters. During the second workshop (June, IRRI), the collaborative systems were already heavily used by the participants. Formal training was given during the June workshop and individual support has been ongoing since the February workshop.

Deviations from the work plan: None

SP5 COMMISSIONED GRANTS

CB1. Training Programme on Genetic Diversity Analysis of Germplasm

Participating Institutions:

CIRAD

IRD

MID-YEAR REPORT

CB2. Development of Training Materials for a Course in Genomics and Comparative Genomics, and Design of Course Curriculum

Principal Investigator:

Theresa M. Fulton, Cornell University

MID-YEAR REPORT

The goal of this project is to develop training materials and a curriculum for a course in genomics and comparative genomics.

- to be useful either as a self-tutorial, or as the basis for a course of approximately 2 weeks duration
- will include definitions of terms, illustrations of concepts, photographs, real-life examples, appropriate applications, lists of key references, and other items as appropriate

Outputs:

Training materials and curriculum, Powerpoint presentation format

Gains:

These materials will be useful as both as concept material for a training course, and as a stand-alone self-tutorial for scientists interested in self-education. They will also be a good complement to already existing training materials, such as the de Vicente and Fulton molecular marker training module CD.

Duration:

2005

Milestones:

2005

- August 31, topics decided, materials organized and developed where necessary
- September 31, First draft disseminated for comments, edits
- December 1, Final edits, given to GCP

Location of Research Activities:

Institute for Genomic Diversity, Cornell University, Ithaca, New York, USA

Users (Beneficiaries):

- Primary intended recipients: Generation Challenge Program, global scientists especially in developing countries
- Beneficiaries: Scientists, educators, interested laypeople
- Target groups: Scientists in developing countries

CB4. Development of Training Materials for a Course in Bioinformatics and Design of Course Curriculum

Principal Investigator:

Richard Bruskiewich, IRRI

Collaborating Scientists:

Tin Wee Tan, National University of Singapore/APBioNet

Theresa Fulton, Cornell University

Marja Thijssen, Wageningen University

Natalia Martins, EMBRAPA

Francis Moonan, IITA

MID-YEAR REPORT

A CropWiki page was set up to support development of the course (see http://cropwiki.irri.org/gcp/index.php/Subprogramme_5).

It was originally anticipated that an GCP funded IRRI bioinformatics staff scientist, Dr. Gloria Despacio-Reyes, would primarily be involved in executing the work for this project. However, Dr. Despacio-Reyes left IRRI prematurely some months ago but no replacement has yet been recruited.

The task leader (R. Bruskiewich) has unfortunately been too fully engaged in other GCP (SP4) commissioned research to find time to work on this project alone. Perhaps the project should be reassigned to another qualified GCP scientist/Institution?

CB5. Development of Reference Molecular Marker Kits to Analyze Diversity of Germplasm for the Year 1 GCP Crops - Rice

Principal Investigator:

Kenneth McNally, IRRI

Collaborating Scientists:

Reflinur, S.P., ICABGRRD, Indonesia

N. Ruairaidh Sackville Hamilton, IRRI

Claire Billot, CIRAD

Brigitte Courtois, CIRAD

Laure Benoit, CIRAD

Cesar Martinez, CIAT

Matthias Lorieux, CIAT

Claudio Brondani, EMBRAPA

Long-zhi Han, CAAS

Marie-Noelle Ndjiondjop, WARDA

Susan McCouch, CORNELL

MID-YEAR REPORT

Based on a request from SP5 sent on April 1, 2005, nominations were solicited from two of the NARES institutes collaborating with IRRI that have a capacity for in-house use of fluorescent labeled SSRs. These two NARES institutes were the Korean Rural Development Administration (RDA) Genebank and the Indonesian Center for Agricultural Biotechnology and Genetic Resources and Research Development (ICABGRRD).

Dr. Yong-Jin Park (RDA) nominated Mr. Ki-jin Park, M.Sc., and Dr. Sutrisno (ICABGRRD) nominated Mr. Reflinur, S.P., M.Sc. Materials supporting each of the nominees were sent to Carmen de Vicente, SP5 Leader, for screening. In mid-May, Mr. Reflinur (ICABGRRD) was selected to carry out this task at IRRI.

A letter of agreement was sent to IRRI on June 10 for the approval of the task assignment and transfer of funds. Processing of this agreement was just recently completed. As yet, the funds are not available to initiate the task.

Mr. Reflinur has also committed to 3 months of genotyping work under the GCP 2005 competitive project with Dr. Rebecca Nelson. This issue was unknown to me at the time the nomination was made. However, an advantage is that travel funds can be share across the two projects giving additional funds to complete the task.

Furthermore, since the genotyping for rice under 2004 Commissioned project #02c (Genotyping of composite germplasm set, Tier 1) has been delayed due to various problems and the work is on-going, a delay in the start of this activity is preferred. Hence, Mr. Reflinur would first accomplish the commitment to the other project followed by work on the molecular marker kit. This timing is adequate for the genotyping and analysis of the rice core collection to be completed.

A NARES scientist, Mr. Reflinur, S.P., from the Indonesian Center for Agricultural Biotechnology and Genetic Resources and Research Development (ICABGRRD) was selected for training under this activity.

Due to constraints on finishing the genotyping activity for rice and the commitment of Mr. Reflinur for genotyping on another GCP activity, we propose that the molecular marker kit development be delayed for 3 months (tentative starting date of November 1).

CB8. Functional Genomics to Improve African Crops

Principal Investigator:

Roeland CHJ van Ham, Wageningen University

MID-YEAR REPORT

This project has been postponed until 2006.

CB9. Use of Molecular Markers for Mining Useful Allelic Diversity – A Summary of SP1 Genotyping for Germplasm Scientists

Principal Investigator:

Carmen de Vicente, IPGRI

MID-YEAR REPORT

This workshop was recently completed, with report currently pending. Please see the GCP website for a final report.

CB10. GR Policies: A Workshop Session in China Devoted to Genetic Resource Policies

Principal Investigator:

Niels Louwaars, Wageningen UR

MID-YEAR REPORT

Paper to be published in Fall 2005.

CB11. CIMMYT Plant Genetic Diversity and Molecular Marker Assisted Breeding: A Training Course

Principal Investigator:

Marilyn Warburton, CIMMYT

MID-YEAR REPORT

- The uptake of new tools depends on the ability of national partner scientists to use the new technologies, which may include the need for capacity building in some cases. Therefore, the training courses offered in the regions for Subprogram 3, entitled “Plant Genetic Diversity and Molecular Marker Assisted Breeding”, is geared towards National Program scientists with the desire and possibilities to utilize markers (via diversity analyses and Marker Assisted Selection) in their breeding programs.
- The objective of this proposed project is to organize a training course at INIA, La Platina, Santiago, Chile on “Plant Genetic Diversity and Molecular Marker Assisted Breeding.” The course is scheduled to be held October 3 – 21, 2005.

In addition to the training of 24 National Program scientists in techniques relating to Marker Assisted Selection and Backcrossing and Diversity analyses, the workshop will create the following outputs:

1. Training materials (lecture notes, powerpoint presentations, laboratory manuals, data analysis manuals, and analysis software manuals) will be given to each student in hard copy and/or CD and also placed on the GCP webpage for downloading to interested people.
2. Public/private partnerships established between the GCP and two foundations (the Kirkhouse trust and the Syngenta Foundation) to provide additional funding for participants to the workshop.
3. Capacity building of the laboratory facilities at INIA, where computers and possibly small laboratory equipment will be purchased for the purposes of the workshop and for their use afterwards.

Duration:

2005 – 2005

Milestones

2005

- Planning of meeting: location and dates selected (INIA La Platina, Santiago Chile Oct. 30 21, 2005).
- Course announced and applications invited (May, 2005).
- Resource people selected and confirmed (see list of collaborators, below; May 2005)
- Course schedule completed and approved (April 2005) (see attached)
- Student Selection Committee formed (including Patricio Hinrichsen, Inia, Carmen de Vicente, GCP, Petr Kosina (CIMMYT), and Marilyn Warburton (CIMMYT). (April 2005)
- Deadline for submission of applications (students): August 1, 2005.
- Deadline for submission of outlines (teachers): August 15, 2005.

Users (Beneficiaries):

- Primary intended recipients: 24 National Program scientists will be trained during the workshop: the breakdown of the students (and their funding source) is as follows:
 - 12 students fully funded by the Generation Challenge Program
 - 2 students fully funded by the Kirkhouse Trust
 - 1 student fully funded by the Syngenta Foundation
 - 3 students funded by their own institutions
 - 6 students attending from INIA and local universities in Santiago. Additional resource people in the laboratory and the computer room will be provided by INIA so that these students do not take the time of the primary resource people, who will attend the paying students.
- Beneficiaries: the National Programs where the 24 students work will benefit from their new skills and knowledge, and these students will be encouraged during the workshop to give seminars and demonstrations upon their return home in order to continue to spread the benefit of the workshop. In addition, new skills of the students in the area of grant writing and new contacts with the IFS will open new funding opportunities for these National Programs in the future. Finally, the resource people invited to give lectures will have the opportunity to attend other people's lectures and discuss together (with other resource people, students, INIA staff, and GCP staff) new ideas and possible new collaborations.
- Target groups: National Program scientists with the ability (present or in the near future) to apply what they have learned in their research.

CB11. Cornell University Plant Genetic Diversity and Molecular Marker Assisted Breeding: A Training Course

Principal Investigator:

Theresa M. Fulton, Cornell University

MID-YEAR REPORT

The objectives of this workshop were to provide both conceptual and hands-on training in the use of plant genetic diversity and molecular marker assisted breeding, with emphasis on practical applied usage and improving the links between plant breeding, germplasm management and utilization, and molecular biology methods, with a particular focus on the use of microsatellite markers.

In my opinion, these objectives were well met. The schedule was arranged to mix laboratory, lecture, computer and other activity types, and the participants seemed engaged and

interested. Frequent informal question and answer sessions were held and questions asked by the participants were thoughtful and reflective. 83% of the responses to the question pertaining to this on the evaluation were in the 2 highest categories (with no responses in the lowest 2 categories), so it appears the participants also felt objectives were well met. In addition, responses to questions regarding relevance of course content, degree of interest generated by the course, and overall effectiveness of the course, were also very positive. Upon returning from the workshop, I received several very positive emails of thanks from participants.

An additional objective was to encourage the use of the facilities available in Pretoria as a “hub” for those in the region who may not have such facilities available to them. Whether this objective was met can only be determined over time, but to this end the participants were shown many of the facilities available both within FABI and University of Pretoria and also nearby at Inqaba Biotec. In addition, Jane Morris presented opportunities for collaboration with the newest GCP member, the African Centre for Gene Technologies.

Course announcements were sent to the GCP-coordinated list of African NARS, as well as contacts of both Dr. Fulton and Dr. Kunert, and people who requested applications after seeing the announcement on the GCP website. A GCP-sponsored needs assessment workshop held last year identified that top priorities for researchers in the southern African region include improving the links between plant breeding and molecular biology and biotechnology; thus some priority was given to participants applying as a team that include a researcher that emphasized plant breeding in his/her work, and one concentrating on molecular biology or germplasm resource management.

As 65 applications were received by the deadline, with others arriving after, it seems that the methods of announcing the workshop were effective. The applicants represented 17 countries in Africa, so clearly the announcement was widely received. Of the 65, about half applied as a team. Several applicants identified themselves as germplasm managers, however, most applicants identified themselves as plant breeders.

The final pool of participants included 13 outside FABI and 3 internal participants who were also laboratory demonstrators or assisted with logistics. Six were members of “team applications” (3 teams). The participants represented 8 African countries, including: Ethiopia, Nigeria, Kenya, Tanzania, Eritrea, Sudan, S. Africa, and Namibia. Most of the participants fulfilled my expectations, as far seeming motivated to learn, interactive and engaged, and optimistic about being able to incorporate what they had learned into their work in the future. Three or four of the participants seemed to have a much harder time understanding the basic genetic concepts. Their applications did not seem to have any common elements that might point to a reason for this.

For reasons unclear, only 12 of the applicants were female (18%), and most did not fit the criteria very well so only two women were among the selected applicants. Furthermore (and discouragingly from a personal point of view), the two women attendees were not among the most interactive or committed of the attendees. Therefore in the future I believe more attention should be given to specifically encouraging African women scientists to apply, to produce a larger pool of females to select from. This might be done by sending the announcement to networks of women scientists (if such exist), or stating in the announcement that women are especially encouraged to apply.

Also attending the lecture and discussion aspects of the course were several graduate students and laboratory assistants currently working at FABI, and several people from nearby seed companies such as Sakata and Pannar. These external people were particularly helpful in discussions as they brought a different viewpoint, that of commercial uses and problems of markers in the private sector.

One issue that arose that may be particular to Africa is that the average age of people in the process of acquiring a Ph.D. is much older than is typical for other regions of the world. As the cost of higher education can be inhibitory, many people in this region have full-time jobs before and usually during the time they are working on a Ph.D. So while the ages of the participants ranged from early 20s to early 50s, the oldest one was in fact just now a Ph.D. candidate. So the distinction between “young” and “early career” scientist is different than for most regions. This is an important and somewhat challenging consideration in application selection for the region, as the general GCP guidelines have been to prioritize “early career” scientists, that is, just finishing or recently finishing a Ph.D. While one would certainly like to encourage continuing education, especially the cases of plant breeders acquiring complementary molecular biology skills, potential future impact and funding can become an issue, for example, IFS will only fund scientists that can be thought to have a career of 15 years ahead of them, which realistically means that they do not fund people over 45.

The workshop agenda was structured with two main goals, 1) to have a logical conceptual order such that basic, key concepts were covered first and that these concepts were linked to the timing of the related laboratory activities, and 2) to have a mix of lecture, laboratory, computer lab, and field trip activities to help keep the participants engaged and interested. From an educational perspective, these goals are both key to producing a deeper level of understanding and also to reach different types of learners.

The resource people had been chosen to have complementary skills, so there was an (in my opinion) extremely good mix of presentations, as can be seen from the agenda. Presentations were either taken from the Fulton and de Vicente CD or powerpoint presentations developed by the resource people. The room where the lectures were held was a perfect size and set up, whereas participants had a writing table and yet the room was small enough for an informal feel which encouraged discussion and participation. Time for questions was included in the agenda, but in general it seemed the participants felt comfortable to ask questions at any point, as very often excellent discussions arose in the middle of presentations.

In addition, several assignments were given to the participants, including group presentations on selected journal papers and summaries of key points (ie. types of markers).

Each weekday included a morning and afternoon break with coffee, juice, and snacks in the outside courtyard. Lunch was at Adler’s on the university campus, just a building or 2 away from FABI, and was a buffet which varied daily and included meat, fish or vegetarian options. The participants were given a per diem for dinner, which could be taken either at the guesthouse where they lodged or at one of the many nearby restaurants.

Several field trips were included. A trip to Pannar Seed Company was taken on Saturday, with all the participants. Pannar is one of the largest seed companies in South Africa, and also sent a resource person to the course who gave a very useful and practical seminar on how molecular markers are used in commercial practice. Their current focus is on maize, soybean, sunflowers and wheat, with some work on vegetables. Another field trip, one late afternoon, was to Inqaba Biotec, a facility offering sequencing and other services to the Africa region.

An optional excursion to the Sterkfontein caves (“the cradle of humankind”) was taken on a Sunday, as well as one to a shopping mall, as requested by the participants.

While we wanted to emphasize the practical aspects of molecular markers, things that could be realistically applied by the participants in their careers or in collaborations, we did also include topics that are more “on the edge” or less currently applicable, so that the participants would also have an overview of what else is going on in scientific research and what might be available to them in the future. These included things like genomics, proteomics, microarrays, etc. On the other hand we also included discussions of the real-life current problems affecting the participants’ research (one example given was water quality and reliable electricity), and a hands-on computer lab of how to use publicly available databases for real questions in their own research such as what sorts of genes are involved in drought tolerance and finding microsatellites for their crop species.

Materials given to the participants included: a folder with the agenda, participant list, pen and notepad, GCP information, and a nametag (in the beginning of the course); and a copy of the Fulton and de Vicente markers CD and a CD (see enclosed) containing all the presentations, agenda, participant list, laboratory protocols, and literature discussed (at the end of the course).

I am quite pleased with the evaluations filled out by the participants (see Summary table and graph below, and original evaluations sent via mail). Overall most scores for all the questions were 4s and 5s, with the exception of the financial-logistical problems (see next section), and the quality and quantity of information received prior to the course. Other than the agenda and list of topics to be covered, I am not sure what else we could have given the participants in advance, as I do feel its important to begin the course with the very basics and without preconceived notions of what might (and might not) be learned. None of the participants that gave this question a low score indicated specifically what might have been helpful.

An interesting point to note is that the majority of participants said they would like more laboratory practicals. In previous discussions prior to organizing the course, there had been some discussion from other course organizers about minimizing or even eliminating hands-on laboratory work. At least for the region of Africa, practical lab experience seems to be a much desired and important part of this type of course. While I do not personally agree that *more* time was needed, as this would most likely have been redundant, it does seem important to keep this minimum amount included in future courses. Two evaluations mentioned the wish for more discussion of the laboratory results. I did feel that we spent some time looking at the gel pictures and discussing what the bands meant, but it seems participants would have liked still more of this.

Strengths noted included knowledgeable instructors and the interactions with participants, high quality of the facilities and laboratory equipment, the amount of information provided and the meeting of the original objectives, balance between topics. I was especially pleased to see the very favorable scores for the quality of the resource people and well-prepared presentation as, indeed, I myself felt that this was a key strength of this particular course. The resource people included those with very diverse, specific expertise to the topics they presented (see agenda). The week before the course Dr. Kunert and Dr. Cullis, in particular, ran through all the lab protocols and generated backup samples in case something didn’t work with the participants (as it turned out, most of the participants did have good results). The weekend before the course began all the resource people available met to go over the details of each person’s presentations and which topics they were responsible for.

A few other comments included wanting to have more bioinformatics and statistics included, preferring to have a weekend day completely free (we gave most of a Sunday off, with just a brief Sunday afternoon meeting), preferring more participants (I disagree with this—more than 16 in a laboratory at one time would be too much I think), and wanting more discussion of conventional breeding techniques (I did not feel that this was part of our objectives, and indeed the majority of the participants had degrees in plant breeding already).

A few people felt that the per diem was too low. We gave USD \$25 per day. As breakfast and lunch was already included in the program, really the participants only needed to buy their evening meal. My colleague and I ate at several different restaurants during the course and never spent even half of this amount on dinner, so I must disagree that a higher per diem was necessary. Only one evaluation was quite a bit more critical than the others, noting that there was too much information presented too fast, too complicated, and lacking clear explanations. Since this opinion did not seem to be a consensus, it may be that this was from one of the people that seemed to have somewhat less of a scientific background than the others (see above).

A few logistical difficulties were encountered that were specifically due to the political turmoil that is ongoing in the Africa region. One was in arranging flights for the participants. In many cases direct flights from one country to another are not available, and in some participants from particular countries were required to go via one country rather than another (for example, from Eritrea). Flights from some countries are vastly more expensive than others due to the political problems of large numbers of refugees coming from that country (eg. Nigeria). Therefore it was absolutely invaluable to have a co-organizer from the region who knew this in advance so we could budget for it.

Another problem was in reimbursing participants and giving per diems given that, due to the apparently high amount of fraud, changing currencies and cashing cheques are difficult to impossible, and participants either did not have bank accounts for money to be transferred into or it was not possible to do so. I do not expect this problem to be solved in Africa anytime soon, but it is something I will be more prepared for in the future.

Finally, I would like to note that the venue was an extremely convenient and well-suited location for this type of course. Large, clean and well-equipped laboratories and computer facilities were available. Guesthouses for the participants were comfortable but simple and inexpensive. University apartments were available for the resource people, including full kitchens and laundry facilities. Both were located within easy walking distance of FABI, and also within walking distance of restaurants and stores. FABI is located within 30-45 minutes drive from the Johannesburg International Airport, making it easy to have participants picked up upon arrival. Overall, FABI proved to be a very good venue to hold training courses.

CB11. ICRISAT Plant Genetic Diversity Analysis and Marker-Assisted Breeding

Principal Investigator:

Subhash Chandra, ICRISAT

Collaborating Scientists:

Morakot Tanticharoen, BIOTEC, Thailand

Theerayut Toojinda, BIOTEC, Thailand

Julapark Chunwongse, BIOTEC, Thailand

MID-YEAR REPORT

Preparations for the course started in late Jan 2005, with Dr de Vicente (Leader SP5) approaching Dr Morakot, Director BIOTEC Thailand, to host the course in Thailand. Dr Morakot kindly agreed and suggested Kasetsart University, Kamphaeng Saen (KPS) Campus, as the venue of the course, with Dr Julapark as Co-Organizer of the course. Dates of the course were finalized in early Feb in consultation with Dr Julapark.

Course contents (**Appendix 1**) were finalized in March based on on-line discussions among Drs Fulton (Cornell), Warburton (CIMMYT) and Chandra (ICRISAT) with advice from Dr de Vicente. We suggest that, in future, the local Co-Organizer should also be consulted in finalizing the course contents.

Course announcement was sent to Asia Region NARS by Mr Carlos (Program Assistant to Dr de Vicente) through email on 12 April as well as posted on the GCP, BIOTEC and Kasetsart websites. Revised announcements were emailed and posted on web on 18 April, 27 April and 04 May. Application deadline was set as 07 June. In future, such revised announcements should be avoided as they could create confusion. Also, the contents of the current application form needs to be appropriately revised to facilitate selection of appropriate participants.

A total of 86 applications were received, 46 from males and 40 from females. The two Co-Organizers met in June at Kasetsart University to identify course participants in accordance with the GCP guidelines. In this meeting, all matters related to local arrangements/logistics were also discussed and the required budget finalized. The identified list of participants was sent to Dr de Vicente. The list was revised and finalized (**Appendix 2**) as per suggestions from Dr de Vicente. The selected participants were intimated of their selection by email and courier on 01 July. Ms Siriporn and Ms Pawanee from BIOTEC arranged Letters of Invitation to be sent to selected participants (and resource persons) to facilitate obtaining Thai visa. Resource persons (**Appendix 3**) were identified by the two Co-Organizers as per GCP guidelines.

The course was attended by 20 participants (**Appendix 4**) from 10 Asian countries. One participant from India, one from Korea and one from PNG could not come. Excellent arrangements were in place to receive the participants and resource persons at the Bangkok International Airport, their transit stay at Bangkok, local travel to and from Kasetsart University (KPS Campus), and their stay at KPS Campus during the entire period of course.

Participants and Resource Persons all stayed at the Saen Palm Training Home at KPS Campus. Saen Palm is fully equipped with all necessary facilities for a comfortable stay, including laundry, food, international telephone facility, and Internet connection. The lecture and lab facilities used for the course were excellent.

The course was delivered by each Resource Person in the most effective way, taking maximally into account the expectations of the participants, their level of understanding and background. The presentations of the keynote speakers in the evening sessions were both informative and insightful in to the subjects related to the course.

The participants evaluated each *Unit of Learning*, immediately after its completion, using an evaluation form (**Appendix 5**) prepared as per GCP guidelines. They also provided an overall

evaluation of the course as a whole at the end of the course using evaluation proforma at **Appendix 6**. The results of the evaluations are given in **Appendix 7**. On a scale of 1-5, the participants adjudged the average performance of different units as follows:

Unit	Average Score
Unit 1	4.14
Unit 2	3.73
Unit 3	3.91
Unit 4	3.81
Unit 5	4.33
Unit 6	4.17
Average	4.02

In the **overall evaluation** of the course, the participants adjudged the **performance of the course as a whole** as **4.37** on a scale of 1-5. The course, without doubt, was a very successful mission. This could become possible only due to the excellent efforts made locally to provide the participants a conducive and comfortable learning environment, coupled of course with the effective delivery of the course by the each resource person.

The participants agreed to form an Asia Region Network, called **AgBioAsia**, to foster *collaborations among participating scientists*, the CGIAR, and other GCP Consortium members, including mechanisms for *technical backstopping, re-training and problem solving* in the Asia region. This was created as a Yahoo Group www.yahogroups.com/groups/AgBioAsia. Most of the participants have enrolled themselves as members of this network.

The participants were provided with a CD containing the course material, including software used in the course.

The institutions of each participant will be provided with a copy of the NTSys software, used in Unit 2, to facilitate diversity analyses. The software used in other Units are all freely available.

The participants have been encouraged to use AgBioAsia network to remain in touch with each other, to share information and experiences, to build partnerships, and to seek help and advice on professional matters.

Participants have also been encouraged to submit proposals for possible IFS funding. Four participants were identified to submit these proposals to course Co-Organizers by 30 Sept. These, after proper scrutiny and revision, will be submitted to Dr de Vicente for onward submission to IFS.

Our special thanks go to BIOTEC staff (Ms Siriporn, Ms Tanapon, Ms Pawanee) to facilitate the process of making excellent arrangements for visa, airport pickup and comfortable transit stay of participants and resource persons in Bangkok. This report would be incomplete if we do not place on record our heartfelt appreciation of the relentless effort that Ms Sukhumarn made before and during the entire period of the course to make everyone feel at home. Well done Sukhumarn. We also heartily thank Dr M Carmen de Vicente, Leader GCP-SP5, for providing us the opportunity to organize this important Asia Region training course.

CB12. Writing Quality Project Proposals that Connect Agricultural Scientists, Stakeholders, and Donors

Principal Investigator:
Karine Malgrand, IPGRI

MID-YEAR REPORT

Objectives:

To increase the capacity of the GCP African, Latin American and Asian partners and potential partners in proposal development that leads to more effective distribution of research outputs and results and to an increased fund raising ability.

Outputs

- Training material on planning and writing concept notes developed (all presentations will be available online after the 3rd workshop) and CDs were distributed to participants.
- 75 participants more confident and efficient in planning for delivery of outputs to users, budgeting and writing successful proposals (2 training workshops have taken place, one is early September)
- Non tangible one: ‘Training style’ using coaching and facilitation methods to empower participants experimented and successful

Gains:

- 75 scientists from NARS and CGIAR, partners and potential partners to the GCP are trained in planning and writing quality project proposals with an emphasis on delivery of outputs to users. Participants are competent in planning and writing proposals for a potential call for proposals on delivery of GCP products; they can also design and participate in research projects with the objective of creating self-reliance in developing country partners.
- Training of trainer’s effect: participants were motivated and confident to deliver the course to their colleagues
- GCP and other donors funding criteria (dictated by the need for social impact and sustainability) are known, respected and included in designing concept notes/project proposals
- The 75 “needs assessment forms (available upon request)” listing the strengths and capacities of the participants and their organizations to contribute to the GCP will allow the GCP consortium members to further develop partnerships to create self-reliance in developing country partners; you can contact the following participants/potential partners:

In Asia :

R.Chandra Babu	Professor Tamil Nadu Agricultural University - Department of Plant Molecular Biology and Biotechnology- Coimbatore-INDIA	chandrar2000@yahoo.com
Kioumars Ghamkhar	Research Associate Centre for Legumes in Mediterranean Agriculture (CLIMA) University of Western Australia (UWA) – Australia	kioumars@cyllene.uwa.edu.au

Feng Dongxin	Deputy Division Chief Department of International Cooperation, Chinese Academy of Agricultural Sciences - Beijing, P. R. China	fengdx@mail.caas.net.cn
Shiv Kumar Agrawal	Senior Scientist Indian Institute of Pulse Research- Kalyanpur, KANPUR India	shiv_iipr@email.com
Rofina Yasmin Othman	Associate Professor University of Malaya- Institute of Biological Sciences, Faculty of Science- University of Malaya	yasmin@um.edu.my
Rongfeng Huang	Professor Biotech Research Institute, China-Beijing	rongfeng@public3.bta.net.cn
Muhamad Yunus	Head of Molecular Biology Laboratory Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development- Bogor – Indonesia	dirayati@indo.net.id
VS Saiprasad Gandra	Scientist Senior Scale ICAR, India- Division of Biotechnology, IHR, Hessaraghatta Lake Post, Bangalore- India	gvssaiprasad@yahoo.co.in
Robin Sabariappan	Associate Professor Centre for Plant Breeding & Genetics - Tamil Nadu Agricultural University- Coimbatore, India	srobin@eth.net
Nguyen Thi Quynh	Vice Director Institute of Tropical Biology - Vietnamese academy of science and technology (VAST)	qtnguyen@hcmc.netnam.vn
Prakash Mrutyunjay Salimath	Professor and Head- Department of Genetics and Plant Breeding University of Agricultural Sciences, Dharwad-Karnataka, INDIA	pmsalimath@hotmail.com
Mohammad Taher Hallajian	Nuclear Research Center for Agriculture and Medicine of IRAN	hallaji_2003@yahoo.com
Hwang Hae-Sung	Research Scientist in Genetic Resources Division National Institution Agriculture Biotechnology- Rep. of Korea	hae0323@rda.go.kr
Raman Babu	Scientist (Plant Breeding) Vivekananda Institute of Hill Agricultural Research, Indian Council of Agricultural Research- VPKAS (ICAR), ALMORA – INDIA	rbabu_icar@rediffmail.com
Wickneswari Ratnam	Associate Professor Univ Kebangsaan Malaysia School of Environmental and Natural Resource Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia	wicki@pkriscc.ukm.my
Sahilah Adb. Mutalib	Research Officer MARDI Malaysia- STRATEGIC RESOURCE RESEARCH CENTRE	sahilah@mardi.my
Mohd Norowi Hamid	Senior Research Officer MARDI Malaysia	norowi@mardi.my
Rashid Anwar	Chief Scientific Officer/ Deputy Director General Institute of Agri. Biotechnology and Genetic Resources - PAKISTAN	rashid1946@yahoo.com

Amarathunga Perera	Senior Geneticist / Plant Breeder and Head - Coconut Research Institute Genetics and Plant Breeding Division, Coconut Research Institute, Lunuwila, Sri Lanka	kanthaperera@yahoo.com
Akinnola Nathaniel Akintunde	International Crop Information System and International Nurseries Scientist - ICARDA Syria	a.n.akintunde@cgiar.org
Orarat Mongkolporn	Assistant Professor Kasetsart University - Department of Horticulture, Kasetsart University	orarat.m@ku.ac.th
Dang Ngoc Ha	Head Science and International Cooperation Department National Maize Research Institute of Vietnam	nmri@hn.vnn.vn
Ngoc Trinh Luu	Director of Plant Genetic Resources Center Vietnam Agricultural Science Institute	trinh2003@hn.vnn.vn

In Africa:

Paul Kayode BAIYERI	Senior lecturer – Dept of crop science, University of Nigeria Nsukka	paulkayodebaiyeri@yahoo.com
Ousmane BOUKAR	Cowpea Breeder & Chief of Scientific Animation Service - IRAD Maroua - Cameroon	ouboukar@yahoo.fr
Rachel CHIKWAMBA	Adjunct Senior Lecturer - FABI, University of Pretoria- South Africa	rchikwamba@csir.co.za
Daniel FONCEKA	Molecular geneticist - CERAAS Thies,Senegal	danielfoncka@yahoo.fr
Alexandre DANSI ANAGONOU	Head of the Plant Genetic Resources Unit Faculty of Science and Technology, University of Abomey-Calavi (UAC) Cotonou, Benin	adansi2001@yahoo.fr
Chabane KAMEL	Senior scientist at Genetic Resources Unit - ICARDA Aleppo Syria	C.Kamel@cgiar.org
Eklou ATTIOGBEVI-SOMADO	Post – Doctoral Fellow, Genetic Resources Unit WARDA (Africa Rice Center) Cotonou, BENIN	e.somado@cgiar.org
Marie-Noelle NDJIONDJOP	Molecular Biologist WARDA (Africa Rice Center) Cotonou, BENIN	m.ndjiondjop@cgiar.org
Hana CHAIR	Researcher in charge of yam genetic diversity and phylogeny in West Africa CIRAD-UAC Cotonou-Benin	hana.chair@cirad.fr
Oscar EYOG-MATIG	PGR senior scientist - IPGRI Cotonou, Benin	o.eyog-matig@cgiar.org
Toussaint MIKPON	General Director’s Assistant - INRAB – Cotonou - Benin	owoyori@yahoo.fr
Frederick Oduor ATIENO	Associate Scientist - IPGRI Nairobi	f.atieno@cgiar.org
James Gichuru GETHI	Head, Maize Program, KARI-Katumani , Kenya	jgethi@wananchi.com
Koffi Simplicie KOUASSI	Program Leader - Projet Bourses du Vivrier (PROBOVI) Abidjan -Cote d’Ivoire	sahyallet@yahoo.fr
Mabel Jendeka MAHASI	Principal Research Scientist - KARI Njoro Kenya.	jendekamabel@yahoo.com
Jonathan MKUMBIRA	Cassava breeder/Geneticist - IITA Ibadan, Nigeria	jmkumbira@cgiar.org

Emmrold MNENEY	Head of Biotechnology Department - Mikocheni Agricultural Research Institute (MARI) Dar Es Salaam- Tanzania	emneney@yahoo.com
Elias Otieno OBUDHO	Lecturer - University of Nairobi - Kenya	eobudho@uonbi.ac.ke
Elizabeth OKAI	Research Officer (In-charge of a research station) - Crops Research Institute - Accra & Kumasi - Ghana	elizabeth_okai@yahoo.co.uk
James Aketch OKENO	Senior Lecturer Department of Horticulture. Teaching Plant Breeding & Genetics - Moi University- Eldoret , Kenya	okenojames@yahoo.com
Emmanuel OKOGBENIN	NRCRI/CIAT Cassava molecular geneticist/breeder - National Root Crops Research Institute(NRCRI) Umuahia, Nigeria.	Eokogbenin@yahoo.com
Altus VILJOEN	Senior lecturer: Dept. Plant Pathology and Microbiology - FABI, University of Pretoria - South Africa	altus.viljoen@fabi.up.ac.za

In Latin America - The list of participants will be communicated after the training workshop.

Duration, Milestones and Location of Research Activities:

Duration: March- October 2005	March-May 2005	Development of curriculum and logistic organization
	22-27 May 2005	1st workshop in Cotonou, Benin
	19-24 June 2005	2 nd workshop in Kuala Lumpur, Malaysia
	5-9 September 2005	3 rd workshop in Quito, Ecuador
	October 2005	Report writing

Users (Beneficiaries):

The primary users are the participants to the training workshops; the majority of participants said they would transfer the skills to their colleagues in national centres. The GCP consortium members and the more general Agriculture Research for Development community will benefit from better planned and written project proposals that lead to more effective distribution of research outputs and results and to an increased fund raising ability.

Collaborators (not exhaustive):

- INIAP, Instituto Nacional Autónomo de Investigaciones Agropecuarias - Ecuador
- César Tapia & Alvaro Monteros
- MARDI, the [Malaysian Agricultural Research and Development Institute](#)
- Abd.Shukor Abd. Rahman
- INRAB, l'Institut National de Recherches Agricoles du Bénin - Olorouto Delphin KOUDANDE
- CAS, the Central Advisory Service on Intellectual Property – Victoria Henson-Apollonio
- GCP, the Generation Challenge Program – Jenny Nelson
- WARDA, the African Rice Centre – Justin Kouka
- IITA the International Institute of Tropical Agriculture - Mahalakshmi Visvanathan
- IPGRI WCA - Raymond Vodouhe, Grace Gouda Mayrama, Hermene Avohou
- IPGRI APO- Ramanatha Rao, Yeang Nyet Poi, Choo Kwong Yan, Anita Anthonysamy
- IPGRI Americas - Adriana Sánchez, Carlos Tovar

Resource persons

In Kuala Lumpur, Malaysia	Zhi-Kang Li	Chief Scientist National Key Facility for Crop Gene Resources and Genetic Improvement - CAAS	lizhk@caas.net.cn
	Carmen de Vicente	Head, sub program 5 – GCP	c.devicente@cgiar.org
	Karine Malgrand	Project coordinator – Organizer and main trainer - IPGRI	karine@malgrand.org
	Jenny Nelson	Communications Coordinator - GCP	j.nelson@cgiar.org
	Victoria Henson-Apollonio	Senior Scientist, Project Manager - CAS	v.henson-apollo@cgiar.org

In Cotonou, Benin	Karine Malgrand	Project coordinator – Organizer and main trainer - IPGRI	karine@malgrand.org
	Jennifer Nelson	Communications Coordinator - GCP	j.nelson@cgiar.org
	Mahalakshmi Visvanathan	Genebank curator - IITA Ibadan	V.Mahalakshmi@cgiar.org
	Justin Kouka	Assistant Director - WARDA	p.kouka@cgiar.org
	Theo van Hintum	Interim Director of the GCP - Wageningen	theo.vanhintum@wur.nl

CB13. The Institute for Genomic Diversity's Interactive Resource Centre

Principal Investigator:

Theresa Fulton, Cornell University

MID-YEAR REPORT

Objectives:

To support plant scientists working in developing countries, firstly those working with the Generation Challenge Program and the CGIAR client network, but extending to all global scientists.

- To provide online resources for day-to-day research in agriculture and crop improvement
- To answer questions and provide assistance to scientists as needed

Outputs

Available at <http://irc.igd.cornell.edu>

- Online resources ranging from protocols to lists of plant databases etc
- Helpdesk for scientists to ask questions via email, fax, mail or phone

Gains:

Support for global scientists, furthering their work on crop improvement

- Help support them after returning home from training programs; helping them implement what they have learned
- Give them a place to ask questions

Duration:

2005 – 2009?

Milestones

2005

- Site online
- Helpdesk available
- Beginning resources posted, including guidelines on writing research proposals & publications, DNA extraction protocols, and links to freely available plant databases
- The Molecular Markers in plant genetic diversity CD, both Vol 1 and Vol 2, are available as downloadable pdf chapters and the CD can be requested (this has been the largest request from the site so far).

Location of Research Activities:

Institute for Genomic Diversity, Cornell University, Ithaca, New York, USA

Users (Beneficiaries):

- Primary intended recipients: global scientists, particularly in developing countries
- Beneficiaries: scientists in developing countries, but also eventually the farmers benefiting from the research, and eventually the communities as well
- Target groups: Firstly scientists that are clients of the Generation Challenge Program and the CGIAR system but extending to others as able and needed

Collaborators (not exhaustive):

Staff at the Institute for Genomic Diversity

The Computational Biology Support Unit at Cornell University

Niels Louwaar et al for Policy support content

CB14. Regional PGR Courses

Principal Investigator:

Marja Thijssen, Wageningen UR

MID-YEAR REPORT

Objectives:

To create a curriculum for regional courses on institutional genetic resource policies in order to create awareness, extend relevant knowledge and sharing experiences among scientists and science managers, which will allow them to develop or strengthen institutional policies and tools for handling Freedom to Operate on IPR and ABS in partner institutions of the CP.

Outputs:

A genetic resource policy course of 2 weeks was organised in Wageningen (using mainly other funds) with the assistance of IPGRI-CAS and Embrapa. This course yields materials for a curriculum for regional courses under the GCP.

Gains:

Materials have been developed ; just need to be arranged in a draft curriculum for GCP. Opportunities are discussed to jointly organise a first regional course at ICARDA in 2006 with joint Wageningen UR – GCP funding.

Duration:

2005 (curriculum development) – 2006 (implementation)

Milestones

2005

‘test-course’ organised ; developments discussed among partners

Location of Research Activities:

Wageningen (in 2006: to be discussed)

Users (Beneficiaries):

The plan will eventually identify 2 different target groups:

1. GCP institutions and partner institutions (in research and downstream work), aimed at awareness and institutional policies
2. GCP scientists, aimed at practical Freedom-to-Operate issues

Collaborators (not exhaustive):

Wageningen UR, Embrapa, IPGRI-CAS

CB15. Distant Policies: A Distance Learning Module for Scientists on Genetic Resource Policies and Their Implications for Freedom-to-Operate

Principal Investigator:

Niels Louwaars, Wageningen UR

MID-YEAR REPORT

CB16. IP Matters: An Intellectual property and Access & Benefit-Sharing Helpdesk and On-line Resource for the GCP Community, Partners and Stakeholders

Principal Investigator:

Victoria Henson-Apollonio, IPGRI

MID-YEAR REPORT

Work in this project is proceeding with an aim to have a webpage on the Internet by the time of the GCP-ARM at the end of September.

- We have secured funding for the hiring of a legal specialist, who will spend 30-50% (based on demand) of their time on this project. An appropriate person has been identified and we are in the process of hiring that individual.
- A consultant is helping us with the Web design to ensure that we will have a web page that is seamlessly integrated into the GCP website.
- We have begun to assemble materials that we and others, already have developed that will be appropriate for this site.

Tangible outputs delivered:

Expert group assembled; Initial list of materials to be put onto site.

Deviations from the work plan:

Hiring of legal specialist has been delayed, but is now underway. Will not cause a deviation from the work plan.

CB17. Reporting for Product Distribution: An Asset Inventory System for the Generation Challenge Programme

Principal Investigator:

Victoria Henson-Apollonio, IPGRI

MID-YEAR REPORT

Work in this project is proceeding with an aim to have prototype Internet-ready forms by the time of the GCP-ARM at the end of September.

- A network of experts has been assembled that are willing to field questions in a variety of areas.
- A consultant has been hired to prepare the templates for the inventory system.
- We have adapted the GCP-Consortium Agreement Schedule 4, with the help of Helen Cordell and Shawn Sullivan, (the 2 attorneys that have had the greatest input into the GCP-Consortium Agreement).
- We will be presenting these materials at the GCP-ARM, in order to get feedback from the scientists that will need to use these forms.
- We expect that the forms will be on-line before grantees are asked to file their year-end reports at the end of 2005.

Tangible outputs delivered: Draft templates

Deviations from the work plan: We are on target